The \LaTeX3 Sources

The \LaTeX Project*

Released 2021-11-22

Abstract

This is the reference documentation for the expl3 programming environment. The expl3 modules set up an experimental naming scheme for \LaTeX commands, which allow the \LaTeX programmer to systematically name functions and variables, and specify the argument types of functions.

The \TeX and \eTeX primitives are all given a new name according to these conventions. However, in the main direct use of the primitives is not required or encouraged: the expl3 modules define an independent low-level \LaTeX3 programming language.

The expl3 modules are designed to be loaded on top of \LaTeX2. With an up-to-date \LaTeX2 kernel, this material is loaded as part of the format. The fundamental programming code can also be loaded with other \TeX formats, subject to restrictions on the full range of functionality.

*E-mail: latex-team@latex-project.org
15.5.3 Items and ranges in token lists ........................................ 112
15.5.4 Sorting token lists ..................................................... 115
15.6 Manipulating tokens in token lists ..................................... 115
15.6.1 Replacing tokens ...................................................... 115
15.6.2 Reassigning category codes ......................................... 116
15.7 Constant token lists ..................................................... 117
15.8 Scratch token lists ...................................................... 117
16 The \texttt{l3str} package: Strings ........................................... 119
16.1 Creating and initialising string variables ............................ 120
16.2 Adding data to string variables ........................................ 121
16.3 String conditionals ....................................................... 121
16.4 Mapping over strings .................................................... 123
16.5 Working with the content of strings .................................. 125
16.6 Modifying string variables ............................................. 128
16.7 String manipulation ..................................................... 129
16.8 Viewing strings .......................................................... 130
16.9 Constant strings ........................................................ 131
16.10 Scratch strings ......................................................... 131
17 The \texttt{l3str-convert} package: string encoding conversions ........... 132
17.1 Encoding and escaping schemes ....................................... 132
17.2 Conversion functions .................................................... 134
17.3 Conversion by expansion (for PDF contexts) ......................... 134
17.4 Possibilities, and things to do ......................................... 134
18 The \texttt{l3quark} package: Quarks ........................................ 136
18.1 Quarks ........................................................................ 136
18.2 Defining quarks .......................................................... 137
18.3 Quark tests ................................................................. 137
18.4 Recursion ..................................................................... 138
18.4.1 An example of recursion with quarks .............................. 139
18.5 Scan marks ................................................................... 140
19 The \texttt{l3seq} package: Sequences and stacks ......................... 141
19.1 Creating and initialising sequences .................................... 141
19.2 Appending data to sequences .......................................... 143
19.3 Recovering items from sequences ...................................... 143
19.4 Recovering values from sequences with branching ............... 144
19.5 Modifying sequences ..................................................... 146
19.6 Sequence conditionals .................................................... 146
19.7 Mapping over sequences ................................................ 147
19.8 Using the content of sequences directly .............................. 149
19.9 Sequences as stacks ....................................................... 150
19.10 Sequences as sets ......................................................... 151
19.11 Constant and scratch sequences ..................................... 152
19.12 Viewing sequences ...................................................... 153
20 The \texttt{\texttt{l3int}} package: Integers \hfill \pageref{sec:13int}

20.1 Integer expressions \hfill \pageref{sec:13int:20.1}

20.2 Creating and initialising integers \hfill \pageref{sec:13int:20.2}

20.3 Setting and incrementing integers \hfill \pageref{sec:13int:20.3}

20.4 Using integers \hfill \pageref{sec:13int:20.4}

20.5 Integer expression conditionals \hfill \pageref{sec:13int:20.5}

20.6 Integer expression loops \hfill \pageref{sec:13int:20.6}

20.7 Integer step functions \hfill \pageref{sec:13int:20.7}

20.8 Formatting integers \hfill \pageref{sec:13int:20.8}

20.9 Converting from other formats to integers \hfill \pageref{sec:13int:20.9}

20.10 Random integers \hfill \pageref{sec:13int:20.10}

20.11 Viewing integers \hfill \pageref{sec:13int:20.11}

20.12 Constant integers \hfill \pageref{sec:13int:20.12}

20.13 Scratch integers \hfill \pageref{sec:13int:20.13}

20.14 Direct number expansion \hfill \pageref{sec:13int:20.14}

20.15 Primitive conditionals \hfill \pageref{sec:13int:20.15}

21 The \texttt{\texttt{l3flag}} package: Expandable flags \hfill \pageref{sec:13flag}

21.1 Setting up flags \hfill \pageref{sec:13flag:21.1}

21.2 Expandable flag commands \hfill \pageref{sec:13flag:21.2}

22 The \texttt{\texttt{l3clist}} package: Comma separated lists \hfill \pageref{sec:13clist}

22.1 Creating and initialising comma lists \hfill \pageref{sec:13clist:22.1}

22.2 Adding data to comma lists \hfill \pageref{sec:13clist:22.2}

22.3 Modifying comma lists \hfill \pageref{sec:13clist:22.3}

22.4 Comma list conditionals \hfill \pageref{sec:13clist:22.4}

22.5 Mapping over comma lists \hfill \pageref{sec:13clist:22.5}

22.6 Using the content of comma lists directly \hfill \pageref{sec:13clist:22.6}

22.7 Comma lists as stacks \hfill \pageref{sec:13clist:22.7}

22.8 Using a single item \hfill \pageref{sec:13clist:22.8}

22.9 Viewing comma lists \hfill \pageref{sec:13clist:22.9}

22.10 Constant and scratch comma lists \hfill \pageref{sec:13clist:22.10}

23 The \texttt{\texttt{l3token}} package: Token manipulation \hfill \pageref{sec:13token}

23.1 Creating character tokens \hfill \pageref{sec:13token:23.1}

23.2 Manipulating and interrogating character tokens \hfill \pageref{sec:13token:23.2}

23.3 Generic tokens \hfill \pageref{sec:13token:23.3}

23.4 Converting tokens \hfill \pageref{sec:13token:23.4}

23.5 Token conditionals \hfill \pageref{sec:13token:23.5}

23.6 Peeking ahead at the next token \hfill \pageref{sec:13token:23.6}

23.7 Description of all possible tokens \hfill \pageref{sec:13token:23.7}
## 24 The \texttt{l3prop} package: Property lists

24.1 Creating and initialising property lists ........................................ 198
24.2 Adding and updating property list entries .................................... 199
24.3 Recovering values from property lists ......................................... 200
24.4 Modifying property lists ............................................................ 201
24.5 Property list conditionals ............................................................ 201
24.6 Recovering values from property lists with branching ..................... 202
24.7 Mapping over property lists ....................................................... 202
24.8 Viewing property lists ............................................................... 204
24.9 Scratch property lists ............................................................... 204
24.10 Constants .................................................................................. 205

## 25 The \texttt{l3skip} package: Dimensions and skips

25.1 Creating and initialising \texttt{dim} variables .................................... 206
25.2 Setting \texttt{dim} variables ........................................................... 207
25.3 Utilities for dimension calculations ............................................... 207
25.4 Dimension expression conditionals .............................................. 208
25.5 Dimension expression loops ......................................................... 210
25.6 Dimension step functions ........................................................... 211
25.7 Using \texttt{dim} expressions and variables ..................................... 212
25.8 Viewing \texttt{dim} variables ........................................................... 213
25.9 Constant dimensions ................................................................ 214
25.10 Scratch dimensions .................................................................. 214
25.11 Creating and initialising \texttt{skip} variables .................................... 214
25.12 Setting \texttt{skip} variables ........................................................... 215
25.13 Skip expression conditionals ......................................................... 216
25.14 Using \texttt{skip} expressions and variables .................................... 216
25.15 Viewing \texttt{skip} variables ........................................................... 216
25.16 Constant skips ....................................................................... 217
25.17 Scratch skips ........................................................................... 217
25.18 Inserting skips into the output ...................................................... 217
25.19 Creating and initialising \texttt{muskip} variables ................................. 218
25.20 Setting \texttt{muskip} variables ....................................................... 218
25.21 Using \texttt{muskip} expressions and variables ................................. 219
25.22 Viewing \texttt{muskip} variables ....................................................... 219
25.23 Constant muskips ................................................................... 220
25.24 Scratch muskips ..................................................................... 220
25.25 Primitive conditional ................................................................. 220

## 26 The \texttt{l3keys} package: Key–value interfaces

26.1 Creating keys ........................................................................... 221
26.2 Sub-dividing keys ...................................................................... 222
26.3 Choice and multiple choice keys ................................................. 227
26.4 Setting keys ............................................................................. 227
26.5 Handling of unknown keys .......................................................... 229
26.6 Selective key setting .................................................................. 230
26.7 Utility functions for keys ............................................................. 231
26.8 Low-level interface for parsing key–val lists ................................ 232
### 33 The \texttt{l3box} package: Boxes

- 33.1 Creating and initialising boxes ........................................... 269
- 33.2 Using boxes ................................................................. 270
- 33.3 Measuring and setting box dimensions .................................... 271
- 33.4 Box conditionals ............................................................. 272
- 33.5 The last box inserted ......................................................... 272
- 33.6 Constant boxes ............................................................... 272
- 33.7 Scratch boxes ................................................................. 272
- 33.8 Viewing box contents ........................................................ 273
- 33.9 Boxes and color .............................................................. 273
- 33.10 Horizontal mode boxes ..................................................... 273
- 33.11 Vertical mode boxes ........................................................ 274
- 33.12 Using boxes efficiently ..................................................... 276
- 33.13 Affine transformations ..................................................... 277
- 33.14 Primitive box conditionals ............................................... 280

### 34 The \texttt{l3coffins} package: Coffin code layer

- 34.1 Creating and initialising coffins ......................................... 281
- 34.2 Setting coffin content and poles ......................................... 282
- 34.3 Coffin affine transformations ............................................. 283
- 34.4 Joining and using coffins .................................................. 283
- 34.5 Measuring coffins ............................................................ 284
- 34.6 Coffin diagnostics ........................................................... 284
- 34.7 Constants and variables .................................................... 285

### 35 The \texttt{l3color} package: Color support

- 35.1 Color in boxes ............................................................... 287
- 35.2 Color models ................................................................. 287
- 35.3 Color expressions ........................................................... 288
- 35.4 Named colors ............................................................... 289
- 35.5 Selecting colors ............................................................. 290
- 35.6 Colors for fills and strokes ............................................... 290
- 35.7 Multiple color models ..................................................... 291
- 35.8 Exporting color specifications .......................................... 291
- 35.9 Creating new color models .............................................. 292
- 35.9.1 Color profiles ........................................................... 293

### 36 The \texttt{l3pdf} package: Core PDF support

- 36.1 Objects .......................................................... 294
- 36.2 Version ............................................................... 294
- 36.3 Compression ............................................................ 295
- 36.4 Destinations ............................................................... 296

VII Additions and removals

298
37 The \texttt{l3candidates} package: Experimental additions to \texttt{l3kernel} \hspace{1cm} 299
37.1 Important notice \hspace{1cm} 299
37.2 Additions to \texttt{l3box} \hspace{1cm} 300
37.3 Additions to \texttt{l3expand} \hspace{1cm} 300
37.4 Additions to \texttt{l3fp} \hspace{1cm} 300
37.5 Additions to \texttt{l3file} \hspace{1cm} 301
37.6 Additions to \texttt{l3flag} \hspace{1cm} 301
37.7 Additions to \texttt{l3intarray} \hspace{1cm} 301
37.8 Additions to \texttt{l3msg} \hspace{1cm} 302
37.9 Additions to \texttt{l3prg} \hspace{1cm} 302
37.10 Additions to \texttt{l3prop} \hspace{1cm} 303
37.11 Additions to \texttt{l3seq} \hspace{1cm} 304
37.12 Additions to \texttt{l3sys} \hspace{1cm} 305
37.13 Additions to \texttt{l3tl} \hspace{1cm} 306
37.14 Additions to \texttt{l3token} \hspace{1cm} 307

VIII Implementation \hspace{1cm} 309

38 \texttt{l3bootstrap} implementation \hspace{1cm} 310
38.1 Lua\TeX{}-specific code \hspace{1cm} 310
38.2 The \texttt{\textbackslash pdfstrcmp} primitive in Xe\TeX{} \hspace{1cm} 310
38.3 Loading support Lua code \hspace{1cm} 311
38.4 Engine requirements \hspace{1cm} 311
38.5 Extending allocators \hspace{1cm} 313
38.6 The \texttt{\textbackslash LATEX}3 code environment \hspace{1cm} 313

39 \texttt{l3names} implementation \hspace{1cm} 316

40 \texttt{l3kernel-functions}: kernel-reserved functions \hspace{1cm} 342
40.1 Internal kernel functions \hspace{1cm} 342
40.2 Kernel backend functions \hspace{1cm} 348

41 \texttt{l3basics} implementation \hspace{1cm} 350
41.1 Renaming some \TeX{} primitives (again) \hspace{1cm} 350
41.2 Defining some constants \hspace{1cm} 352
41.3 Defining functions \hspace{1cm} 352
41.4 Selecting tokens \hspace{1cm} 353
41.5 Gobbling tokens from input \hspace{1cm} 355
41.6 Debugging and patching later definitions \hspace{1cm} 355
41.7 Conditional processing and definitions \hspace{1cm} 356
41.8 Dissecting a control sequence \hspace{1cm} 362
41.9 Exist or free \hspace{1cm} 364
41.10 Preliminaries for new functions \hspace{1cm} 366
41.11 Defining new functions \hspace{1cm} 367
41.12 Copying definitions \hspace{1cm} 369
41.13 Undefining functions \hspace{1cm} 369
41.14 Generating parameter text from argument count \hspace{1cm} 370
41.15 Defining functions from a given number of arguments \hspace{1cm} 371
41.16 Using the signature to define functions \hspace{1cm} 372
45.3.4 Framework .................................................. 469
45.3.5 Quantifiers ................................................. 472
45.3.6 Raw characters .............................................. 474
45.3.7 Character properties ........................................ 476
45.3.8 Anchoring and simple assertions ......................... 477
45.3.9 Character classes ........................................... 478
45.3.10 Groups and alternations .................................. 481
45.3.11 Catcodes and csnames .................................... 484
45.3.12 Raw token lists with \u .................................. 488
45.3.13 Other ....................................................... 491
45.3.14 Showing regexes ........................................... 491
45.4 Building ....................................................... 498
45.4.1 Variables used while building ............................. 498
45.4.2 Framework ................................................ 498
45.4.3 Helpers for building an NFA ............................... 500
45.4.4 Building classes ............................................ 501
45.4.5 Building groups ............................................. 503
45.4.6 Others ...................................................... 508
45.5 Matching ....................................................... 509
45.5.1 Variables used when matching ............................. 510
45.5.2 Matching: framework ....................................... 512
45.5.3 Using states of the NFA ...................................... 515
45.5.4 Actions when matching ..................................... 516
45.6 Replacement .................................................. 519
45.6.1 Variables and helpers used in replacement ............... 519
45.6.2 Query and brace balance ................................... 520
45.6.3 Framework ............................................... 521
45.6.4 Submatches ............................................... 524
45.6.5 Csnames in replacement ................................... 526
45.6.6 Characters in replacement ................................. 527
45.6.7 An error .................................................... 531
45.7 User functions ............................................... 531
45.7.1 Variables and helpers for user functions ................ 533
45.7.2 Matching .................................................. 535
45.7.3 Extracting submatches ..................................... 535
45.7.4 Replacement .............................................. 540
45.7.5 Peeking ahead ............................................ 543
45.8 Messages ...................................................... 549
45.9 Code for tracing .............................................. 555

46 l3prg implementation ........................................ 556
46.1 Primitive conditionals ........................................ 556
46.2 Defining a set of conditional functions .................... 556
46.3 The boolean data type ....................................... 556
46.4 Internal auxiliaries .......................................... 558
46.5 Boolean expressions ......................................... 559
46.6 Logical loops ................................................ 564
46.7 Producing multiple copies ................................... 565
46.8 Detecting \TeX’s mode ....................................... 567
46.9 Internal programming functions ............................ 567
47 l3sys implementation

47.1 Kernel code ........................................ 569
  47.1.1 Detecting the engine ......................... 569
  47.1.2 Randomness .................................. 571
  47.1.3 Platform ..................................... 571
  47.1.4 Configurations ................................ 571
  47.1.5 Access to the shell ......................... 573

47.2 Dynamic (every job) code .......................... 576
  47.2.1 The name of the job ............................ 576
  47.2.2 Time and date ................................ 576
  47.2.3 Random numbers ............................... 577
  47.2.4 Access to the shell ......................... 579
  47.2.5 Held over from l3file ....................... 579

47.3 Last-minute code .................................. 580
  47.3.1 Detecting the output ........................... 580
  47.3.2 Configurations ................................ 580

48 l3msg implementation ............................... 582

48.1 Internal auxiliaries ............................... 582
48.2 Creating messages .................................. 582
48.3 Messages: support functions and text .............. 584
48.4 Showing messages: low level mechanism .......... 585
48.5 Displaying messages ................................ 587
48.6 Kernel-specific functions ......................... 596
48.7 Internal messages .................................. 597
48.8 Expandable errors .................................. 603
48.9 Message formatting ................................ 604

49 l3file implementation ............................... 605

49.1 Input operations .................................. 605
  49.1.1 Variables and constants ...................... 605
  49.1.2 Stream management ............................. 606
  49.1.3 Reading input ................................ 609

49.2 Output operations ................................ 612
  49.2.1 Variables and constants ...................... 612
  49.2.2 Internal auxiliaries ......................... 613

49.3 Stream management ................................. 613
  49.3.1 Deferred writing ............................... 615
  49.3.2 Immediate writing ................................ 615
  49.3.3 Special characters for writing ............... 616
  49.3.4 Hard-wrapping lines to a character count ...... 617

49.4 File operations .................................. 626
  49.4.1 Internal auxiliaries ......................... 628

49.5 GetIdInfo ......................................... 643
49.6 Checking the version of kernel dependencies ....... 644
49.7 Messages ......................................... 646
49.8 Functions delayed from earlier modules .......... 647
50 \texttt{l3luatex implementation} 649
50.1 Breaking out to Lua 649
50.2 Messages 650
50.3 Lua functions for internal use 650
50.4 Preserving \texttt{initex} Lua data for runs 657

51 \texttt{l3legacy} Implementation 659

52 \texttt{l3tl} implementation 661
52.1 Functions 661
52.2 Constant token lists 663
52.3 Adding to token list variables 663
52.4 Internal quarks and quark-query functions 666
52.5 Reassigning token list category codes 666
52.6 Modifying token list variables 669
52.7 Token list conditionals 673
52.8 Mapping over token lists 678
52.9 Using token lists 680
52.10 Working with the contents of token lists 681
52.11 The first token from a token list 684
52.12 Token by token changes 689
52.13 Using a single item 691
52.14 Viewing token lists 694
52.15 Internal scan marks 696
52.16 Scratch token lists 696

53 \texttt{l3str} implementation 697
53.1 Internal auxiliaries 697
53.2 Creating and setting string variables 698
53.3 Modifying string variables 699
53.4 String comparisons 700
53.5 Mapping over strings 703
53.6 Accessing specific characters in a string 705
53.7 Counting characters 710
53.8 The first character in a string 711
53.9 String manipulation 712
53.10 Viewing strings 714

54 \texttt{l3str-convert} implementation 715
54.1 Helpers 715
54.1.1 Variables and constants 715
54.2 String conditionals 717
54.3 Conversions 718
54.3.1 Producing one byte or character 718
54.3.2 Mapping functions for conversions 719
54.3.3 Error-reporting during conversion 720
54.3.4 Framework for conversions 721
54.3.5 Byte unescape and escape 725
54.3.6 Native strings 726
54.3.7 \texttt{clist} 727
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>54.3.8</td>
<td>8-bit encodings</td>
<td>727</td>
</tr>
<tr>
<td>54.4</td>
<td>Messages</td>
<td>730</td>
</tr>
<tr>
<td>54.5</td>
<td>Escaping definitions</td>
<td>731</td>
</tr>
<tr>
<td>54.5.1</td>
<td>Unescape methods</td>
<td>732</td>
</tr>
<tr>
<td>54.5.2</td>
<td>Escape methods</td>
<td>736</td>
</tr>
<tr>
<td>54.6</td>
<td>Encoding definitions</td>
<td>738</td>
</tr>
<tr>
<td>54.6.1</td>
<td>UTF-8 support</td>
<td>738</td>
</tr>
<tr>
<td>54.6.2</td>
<td>UTF-16 support</td>
<td>743</td>
</tr>
<tr>
<td>54.6.3</td>
<td>UTF-32 support</td>
<td>748</td>
</tr>
<tr>
<td>54.7</td>
<td>PDF names and strings by expansion</td>
<td>751</td>
</tr>
<tr>
<td>54.7.1</td>
<td>ISO 8859 support</td>
<td>752</td>
</tr>
<tr>
<td>55</td>
<td>l3quark implementation</td>
<td>769</td>
</tr>
<tr>
<td>55.1</td>
<td>Quarks</td>
<td>769</td>
</tr>
<tr>
<td>55.2</td>
<td>Scan marks</td>
<td>777</td>
</tr>
<tr>
<td>56</td>
<td>l3seq implementation</td>
<td>779</td>
</tr>
<tr>
<td>56.1</td>
<td>Allocation and initialisation</td>
<td>780</td>
</tr>
<tr>
<td>56.2</td>
<td>Appending data to either end</td>
<td>783</td>
</tr>
<tr>
<td>56.3</td>
<td>Modifying sequences</td>
<td>784</td>
</tr>
<tr>
<td>56.4</td>
<td>Sequence conditionals</td>
<td>786</td>
</tr>
<tr>
<td>56.5</td>
<td>Recovering data from sequences</td>
<td>788</td>
</tr>
<tr>
<td>56.6</td>
<td>Mapping over sequences</td>
<td>792</td>
</tr>
<tr>
<td>56.7</td>
<td>Using sequences</td>
<td>796</td>
</tr>
<tr>
<td>56.8</td>
<td>Sequence stacks</td>
<td>797</td>
</tr>
<tr>
<td>56.9</td>
<td>Viewing sequences</td>
<td>798</td>
</tr>
<tr>
<td>56.10</td>
<td>Scratch sequences</td>
<td>799</td>
</tr>
<tr>
<td>57</td>
<td>l3int implementation</td>
<td>800</td>
</tr>
<tr>
<td>57.1</td>
<td>Integer expressions</td>
<td>801</td>
</tr>
<tr>
<td>57.2</td>
<td>Creating and initialising integers</td>
<td>803</td>
</tr>
<tr>
<td>57.3</td>
<td>Setting and incrementing integers</td>
<td>805</td>
</tr>
<tr>
<td>57.4</td>
<td>Using integers</td>
<td>806</td>
</tr>
<tr>
<td>57.5</td>
<td>Integer expression conditionals</td>
<td>806</td>
</tr>
<tr>
<td>57.6</td>
<td>Integer expression loops</td>
<td>810</td>
</tr>
<tr>
<td>57.7</td>
<td>Integer step functions</td>
<td>811</td>
</tr>
<tr>
<td>57.8</td>
<td>Formatting integers</td>
<td>813</td>
</tr>
<tr>
<td>57.9</td>
<td>Converting from other formats to integers</td>
<td>818</td>
</tr>
<tr>
<td>57.10</td>
<td>Viewing integer</td>
<td>821</td>
</tr>
<tr>
<td>57.11</td>
<td>Random integers</td>
<td>822</td>
</tr>
<tr>
<td>57.12</td>
<td>Constant integers</td>
<td>822</td>
</tr>
<tr>
<td>57.13</td>
<td>Scratch integers</td>
<td>822</td>
</tr>
<tr>
<td>57.14</td>
<td>Integers for earlier modules</td>
<td>823</td>
</tr>
<tr>
<td>58</td>
<td>l3flag implementation</td>
<td>824</td>
</tr>
<tr>
<td>58.1</td>
<td>Non-expandable flag commands</td>
<td>824</td>
</tr>
<tr>
<td>58.2</td>
<td>Expandable flag commands</td>
<td>825</td>
</tr>
</tbody>
</table>
59 l3clist implementation
  59.1 Removing spaces around items ........................................ 828
  59.2 Allocation and initialisation ........................................ 829
  59.3 Adding data to comma lists ......................................... 831
  59.4 Comma lists as stacks .............................................. 832
  59.5 Modifying comma lists ............................................. 834
  59.6 Comma list conditionals ........................................... 837
  59.7 Mapping over comma lists ......................................... 838
  59.8 Using comma lists .................................................. 842
  59.9 Using a single item ................................................ 844
  59.10 Viewing comma lists .............................................. 846
  59.11 Scratch comma lists ............................................... 847

60 l3token implementation
  60.1 Internal auxiliaries .............................................. 848
  60.2 Manipulating and interrogating character tokens ................. 848
  60.3 Creating character tokens ....................................... 851
  60.4 Generic tokens .................................................... 859
  60.5 Token conditionals .............................................. 860
  60.6 Peeking ahead at the next token .................................. 870

61 l3prop implementation
  61.1 Internal auxiliaries .............................................. 878
  61.2 Allocation and initialisation .................................... 879
  61.3 Accessing data in property lists .................................. 881
  61.4 Property list conditionals ....................................... 885
  61.5 Recovering values from property lists with branching ......... 886
  61.6 Mapping over property lists ..................................... 886
  61.7 Viewing property lists .......................................... 888

62 l3skip implementation
  62.1 Length primitives renamed .................................... 890
  62.2 Internal auxiliaries .............................................. 890
  62.3 Creating and initialising dim variables ......................... 890
  62.4 Setting dim variables ........................................... 891
  62.5 Utilities for dimension calculations ........................... 892
  62.6 Dimension expression conditionals .............................. 893
  62.7 Dimension expression loops ..................................... 895
  62.8 Dimension step functions ....................................... 896
  62.9 Using dim expressions and variables ............................ 898
  62.10 Viewing dim variables ........................................... 900
  62.11 Constant dimensions ............................................ 900
  62.12 Scratch dimensions ............................................. 900
  62.13 Creating and initialising skip variables ....................... 900
  62.14 Setting skip variables ......................................... 902
  62.15 Skip expression conditionals .................................... 902
  62.16 Using skip expressions and variables ......................... 903
  62.17 Inserting skips into the output ................................ 903
  62.18 Viewing skip variables ......................................... 904
  62.19 Constant skips .................................................. 904
67 l3fp-traps Implementation

67.1 Flags .................................................. 977
67.2 Traps .................................................. 977
67.3 Errors ............................................... 981
67.4 Messages ............................................. 981

68 l3fp-round implementation

68.1 Rounding tools ...................................... 983
68.2 The round function .................................. 987

69 l3fp-parse implementation

69.1 Work plan .............................................. 992
  69.1.1 Storing results .................................. 993
  69.1.2 Precedence and infix operators ................. 994
  69.1.3 Prefix operators, parentheses, and functions 997
  69.1.4 Numbers and reading tokens one by one ....... 998
69.2 Main auxiliary functions ............................ 1000
69.3 Helpers ............................................... 1001
69.4 Parsing one number ................................ 1002
  69.4.1 Numbers: trimming leading zeros ............... 1008
  69.4.2 Number: small significand ...................... 1009
  69.4.3 Number: large significand ...................... 1011
  69.4.4 Number: beyond 16 digits, rounding .......... 1013
  69.4.5 Number: finding the exponent ................. 1016
69.5 Constants, functions and prefix operators ....... 1019
  69.5.1 Prefix operators ................................ 1019
  69.5.2 Constants ......................................... 1022
  69.5.3 Functions ......................................... 1023
69.6 Main functions .................................... 1024
69.7 Infix operators .................................... 1026
  69.7.1 Closing parentheses and commas ................. 1027
  69.7.2 Usual infix operators ........................... 1029
  69.7.3 Juxtaposition ................................... 1030
  69.7.4 Multi-character cases ........................... 1030
  69.7.5 Ternary operator ................................ 1031
  69.7.6 Comparisons ..................................... 1031
69.8 Tools for functions ................................ 1033
69.9 Messages ............................................. 1036

70 l3fp-assign implementation .......................... 1037

70.1 Assigning values ................................... 1037
70.2 Updating values ..................................... 1038
70.3 Showing values ...................................... 1038
70.4 Some useful constants and scratch variables .... 1039
71 l3fp-logic Implementation
  71.1 Syntax of internal functions ....................................... 1040
  71.2 Tests ........................................................................ 1040
  71.3 Comparison .................................................................. 1041
  71.4 Floating point expression loops .................................... 1044
  71.5 Extrema ...................................................................... 1047
  71.6 Boolean operations ...................................................... 1049
  71.7 Ternary operator .......................................................... 1050

72 l3fp-basics Implementation
  72.1 Addition and subtraction ................................................ 1052
    72.1.1 Sign, exponent, and special numbers ......................... 1053
    72.1.2 Absolute addition .................................................. 1055
    72.1.3 Absolute subtraction .............................................. 1057
  72.2 Multiplication ............................................................. 1061
    72.2.1 Signs, and special numbers .................................... 1061
    72.2.2 Absolute multiplication ......................................... 1063
  72.3 Division ................................................................. 1065
    72.3.1 Signs, and special numbers .................................... 1065
    72.3.2 Work plan ............................................................ 1066
    72.3.3 Implementing the significand division ....................... 1069
  72.4 Square root .............................................................. 1074
  72.5 About the sign and exponent ........................................ 1081
  72.6 Operations on tuples .................................................... 1082

73 l3fp-extended implementation
  73.1 Description of fixed point numbers .................................. 1084
  73.2 Helpers for numbers with extended precision .................... 1085
  73.3 Multiplying a fixed point number by a short one ................. 1086
  73.4 Dividing a fixed point number by a small integer ............... 1086
  73.5 Adding and subtracting fixed points ................................ 1087
  73.6 Multiplying fixed points ............................................... 1088
  73.7 Combining product and sum of fixed points ....................... 1089
  73.8 Extended-precision floating point numbers ....................... 1092
  73.9 Dividing extended-precision numbers ............................... 1094
  73.10 Inverse square root of extended precision numbers .......... 1098
  73.11 Converting from fixed point to floating point .................. 1100

74 l3fp-expo implementation
  74.1 Logarithm ............................................................... 1102
    74.1.1 Work plan .......................................................... 1102
    74.1.2 Some constants ................................................... 1103
    74.1.3 Sign, exponent, and special numbers ......................... 1103
    74.1.4 Absolute Ln ......................................................... 1103
  74.2 Exponential ............................................................ 1110
    74.2.1 Sign, exponent, and special numbers ......................... 1110
  74.3 Power ................................................................. 1115
  74.4 Factorial ............................................................... 1121
75 13fp-trig Implementation

75.1 Direct trigonometric functions ........................................... 1125
  75.1.1 Filtering special cases ........................................... 1125
  75.1.2 Distinguishing small and large arguments ....................... 1128
  75.1.3 Small arguments ........................................... 1129
  75.1.4 Argument reduction in degrees .................................... 1129
  75.1.5 Argument reduction in radians .................................... 1130
  75.1.6 Computing the power series ..................................... 1138

75.2 Inverse trigonometric functions ........................................ 1140
  75.2.1 Arctangent and arccotangent ..................................... 1141
  75.2.2 Arcsine and arccosine ........................................... 1146
  75.2.3 Arcsecant and arccosecant ...................................... 1148

76 13fp-convert implementation ............................................. 1150

76.1 Dealing with tuples ................................................... 1150
76.2 Trimming trailing zeros ................................................ 1150
76.3 Scientific notation .................................................... 1151
76.4 Decimal representation .............................................. 1152
76.5 Token list representation ............................................ 1154
76.6 Formatting ............................................................ 1155
76.7 Convert to dimension or integer ...................................... 1155
76.8 Convert from a dimension ............................................. 1156
76.9 Use and eval ......................................................... 1157
76.10 Convert an array of floating points to a comma list ............... 1158

77 13fp-random Implementation ............................................. 1160

77.1 Engine support ....................................................... 1160
77.2 Random floating point ............................................... 1164
77.3 Random integer ..................................................... 1164

78 13fparray implementation ............................................... 1170

78.1 Allocating arrays .................................................... 1170
78.2 Array items .......................................................... 1171

79 13ctab implementation ................................................... 1175

79.1 Variables ............................................................ 1175
79.2 Allocating category code tables ..................................... 1176
79.3 Saving category code tables ......................................... 1177
79.4 Using category code tables .......................................... 1178
79.5 Category code table conditionals .................................... 1183
79.6 Constant category code tables ...................................... 1184
79.7 Messages ............................................................ 1186

80 13unicode implementation .............................................. 1188

81 13text implementation ................................................... 1192

81.1 Internal auxiliaries .................................................. 1192
81.2 Utilities .............................................................. 1193
81.3 Configuration variables ............................................. 1196
81.4 Expansion to formatted text ......................................... 1197
82 \texttt{\textit{\texttt{l3text-case}} implementation}  
82.1 Case changing \hfill 1205  
82.2 Case changing data for 8-bit engines \hfill 1226  

83 \texttt{\textit{\texttt{l3text-purify}} implementation}  
83.1 Purifying text \hfill 1238  
83.2 Accent and letter-like data for purifying text \hfill 1244  

84 \texttt{\textit{\texttt{l3box}} implementation}  
84.1 Support code \hfill 1252  
84.2 Creating and initialising boxes \hfill 1252  
84.3 Measuring and setting box dimensions \hfill 1253  
84.4 Using boxes \hfill 1254  
84.5 Box conditionals \hfill 1255  
84.6 The last box inserted \hfill 1255  
84.7 Constant boxes \hfill 1255  
84.8 Scratch boxes \hfill 1256  
84.9 Viewing box contents \hfill 1256  
84.10 Horizontal mode boxes \hfill 1257  
84.11 Vertical mode boxes \hfill 1259  
84.12 Affine transformations \hfill 1262  

85 \texttt{\textit{\texttt{l3coffins}} Implementation}  
85.1 Coffins: data structures and general variables \hfill 1272  
85.2 Basic coffin functions \hfill 1273  
85.3 Measuring coffins \hfill 1279  
85.4 Coffins: handle and pole management \hfill 1279  
85.5 Coffins: calculation of pole intersections \hfill 1283  
85.6 Affine transformations \hfill 1285  
85.7 Aligning and typesetting of coffins \hfill 1293  
85.8 Coffin diagnostics \hfill 1298  
85.9 Messages \hfill 1304  

86 \texttt{\textit{\texttt{l3color}} Implementation}  
86.1 Basics \hfill 1305  
86.2 Predefined color names \hfill 1306  
86.3 Setup \hfill 1307  
86.4 Utility functions \hfill 1307  
86.5 Model conversion \hfill 1307  
86.6 Color expressions \hfill 1308  
86.7 Selecting colors (and color models) \hfill 1316  
86.8 Fill and stroke color \hfill 1317  
86.9 Defining named colors \hfill 1318  
86.10 Exporting colors \hfill 1321  
86.11 Additional color models \hfill 1323  
86.12 Applying profiles \hfill 1337  
86.13 Diagnostics \hfill 1338  
86.14 Messages \hfill 1339
87 \texttt{l3pdf} implementation \hspace{1cm} 1342
  87.1 Compression \hspace{1cm} 1342
  87.2 Objects \hspace{1cm} 1343
  87.3 Version \hspace{1cm} 1343
  87.4 Destinations \hspace{1cm} 1345

88 \texttt{l3candidates} Implementation \hspace{1cm} 1346
  88.1 Additions to \texttt{l3box} \hspace{1cm} 1346
    88.1.1 Viewing part of a box \hspace{1cm} 1346
  88.2 Additions to \texttt{l3flag} \hspace{1cm} 1348
  88.3 Additions to \texttt{l3msg} \hspace{1cm} 1349
  88.4 Additions to \texttt{l3prg} \hspace{1cm} 1350
  88.5 Additions to \texttt{l3prop} \hspace{1cm} 1351
  88.6 Additions to \texttt{l3seq} \hspace{1cm} 1351
  88.7 Additions to \texttt{l3sys} \hspace{1cm} 1356
  88.8 Additions to \texttt{l3file} \hspace{1cm} 1357
  88.9 Additions to \texttt{l3tl} \hspace{1cm} 1357
    88.9.1 Building a token list \hspace{1cm} 1357
    88.9.2 Other additions to \texttt{l3tl} \hspace{1cm} 1361
  88.10 Additions to \texttt{l3token} \hspace{1cm} 1361

89 \texttt{l3deprecation} implementation \hspace{1cm} 1364
  89.1 Helpers and variables \hspace{1cm} 1364
  89.2 Patching definitions to deprecate \hspace{1cm} 1365
  89.3 Removed functions \hspace{1cm} 1368
  89.4 Loading the patches \hspace{1cm} 1372
  89.5 Deprecated \texttt{l3str} functions \hspace{1cm} 1372
  89.6 Deprecated \texttt{l3seq} functions \hspace{1cm} 1373
  89.7 Deprecated \texttt{l3tl} functions \hspace{1cm} 1373
  89.8 Deprecated \texttt{l3token} functions \hspace{1cm} 1374

Index \hspace{1cm} 1375
Part I
Introduction
Chapter 1

Introduction to expl3 and this document

This document is intended to act as a comprehensive reference manual for the expl3 language. A general guide to the LATEX3 programming language is found in expl3.pdf.

1.1 Naming functions and variables

LATEX3 does not use @ as a “letter” for defining internal macros. Instead, the symbols _, and : are used in internal macro names to provide structure. The name of each function is divided into logical units using _, while : separates the name of the function from the argument specifier (“arg-spec”). This describes the arguments expected by the function. In most cases, each argument is represented by a single letter. The complete list of arg-spec letters for a function is referred to as the signature of the function.

Each function name starts with the module to which it belongs. Thus apart from a small number of very basic functions, all expl3 function names contain at least one underscore to divide the module name from the descriptive name of the function. For example, all functions concerned with comma lists are in module cist and begin \clist_.

Every function must include an argument specifier. For functions which take no arguments, this will be blank and the function name will end :. Most functions take one or more arguments, and use the following argument specifiers:

N and n These mean no manipulation, of a single token for N and of a set of tokens given in braces for n. Both pass the argument through exactly as given. Usually, if you use a single token for an n argument, all will be well.

c This means csname, and indicates that the argument will be turned into a csname before being used. So \foo:c {ArgumentOne} will act in the same way as \foo:N \ArgumentOne.

V and v These mean value of variable. The V and v specifiers are used to get the content of a variable without needing to worry about the underlying LATEX structure containing the data. A V argument will be a single token (similar to N), for example \foo:V \MyVariable; on the other hand, using v a csname is constructed first, and then the value is recovered, for example \foo:v {MyVariable}.
This means *expansion once*. In general, the \texttt{V} and \texttt{v} specifiers are favoured over \texttt{o} for recovering stored information. However, \texttt{o} is useful for correctly processing information with delimited arguments.

The \texttt{x} specifier stands for *exhaustive expansion*: every token in the argument is fully expanded until only unexpandable ones remain. The \TeX\ \texttt{edef} primitive carries out this type of expansion. Functions which feature an \texttt{x}-type argument are not expandable.

The \texttt{e} specifier is in many respects identical to \texttt{x}, but with a very different implementation. Functions which feature an \texttt{e}-type argument may be expandable. The drawback is that \texttt{e} is extremely slow (often more than 200 times slower) in older engines, more precisely in non-Lua\TeX\ engines older than 2019.

The \texttt{f} specifier stands for *full expansion*, and in contrast to \texttt{x} stops at the first non-expandable token (reading the argument from left to right) without trying to expand it. If this token is a ⟨space token⟩, it is gobbled, and thus won’t be part of the resulting argument. For example, when setting a token list variable (a macro used for storage), the sequence

\begin{verbatim}
\tl_set:Nn \l_mya_tl { A }
\tl_set:Nn \l_myb_tl { B }
\tl_set:Nf \l_mya_tl { \l_mya_tl \l_myb_tl }
\end{verbatim}

will leave \texttt{\l_mya_tl} with the content \texttt{A\l_myb_tl}, as \texttt{A} cannot be expanded and so terminates expansion before \texttt{\l_myb_tl} is considered.

For logic tests, there are the branch specifiers \texttt{T} (true) and \texttt{F} (false). Both specifiers treat the input in the same way as \texttt{n} (no change), but make the logic much easier to see.

The letter \texttt{p} indicates \TeX\ parameters. Normally this will be used for delimited functions as expl3 provides better methods for creating simple sequential arguments.

Finally, there is the \texttt{w} specifier for *weird* arguments. This covers everything else, but mainly applies to delimited values (where the argument must be terminated by some specified string).

The \texttt{D} stands for *Do not use*. All of the \TeX\ primitives are initially \texttt{\let} to a \texttt{D} name, and some are then assigned a second name. These functions have no standardized syntax, they are engine dependent and their name can change without warning, thus their use is *strongly discouraged* in package code: programmers should instead use the interfaces documented in interface3.pdf\footnote{If a primitive offers a functionality not yet in the kernel, programmers and users are encouraged to write to the \LaTeX-L mailing list (mailto:LATEX-L@listserv.uni-heidelberg.de) describing their use-case and intended behaviour, so that a possible interface can be discussed. Temporarily, while an interface is not provided, programmers may use the procedure described in the \LaTeXstyleguide.pdf.}.

Notice that the argument specifier describes how the argument is processed prior to being passed to the underlying function. For example, \texttt{\foo:c} will take its argument, convert it to a control sequence and pass it to \texttt{\foo:N}.

Variables are named in a similar manner to functions, but begin with a single letter to define the type of variable:
c Constant: global parameters whose value should not be changed.

g Parameters whose value should only be set globally.

l Parameters whose value should only be set locally.

Each variable name is then build up in a similar way to that of a function, typically starting with the module\(^2\) name and then a descriptive part. Variables end with a short identifier to show the variable type:

clist Comma separated list.

dim “Rigid” lengths.

fp Floating-point values;

int Integer-valued count register.

muskip “Rubber” lengths for use in mathematics.

seq “Sequence”: a data-type used to implement lists (with access at both ends) and stacks.

skip “Rubber” lengths.

str String variables: contain character data.

tl Token list variables: placeholder for a token list.

Applying \(V\)-type or \(v\)-type expansion to variables of one of the above types is supported, while it is not supported for the following variable types:

bool Either true or false.

box Box register.

coffin A “box with handles” — a higher-level data type for carrying out box alignment operations.

flag Integer that can be incremented expandably.

fparray Fixed-size array of floating point values.

intarray Fixed-size array of integers.

ior/iow An input or output stream, for reading from or writing to, respectively.

prop Property list: analogue of dictionary or associative arrays in other languages.

regex Regular expression.

\(^{2}\)The module names are not used in case of generic scratch registers defined in the data type modules, e.g., the \texttt{int} module contains some scratch variables called \texttt{\_tmpa\_int}, \texttt{\_tmpb\_int}, and so on. In such a case adding the module name up front to denote the module and in the back to indicate the type, as in \texttt{\_int\_tmpa\_int} would be very unreadable.
1.1.1 Scratch variables

Modules focussed on variable usage typically provide four scratch variables, two local and two global, with names of the form \( \langle \text{scope} \rangle_{\text{tmpa}} \langle \text{type} \rangle / \langle \text{scope} \rangle_{\text{tmpb}} \langle \text{type} \rangle \). These are never used by the core code. The nature of \TeX{} grouping means that as with any other scratch variable, these should only be set and used with no intervening third-party code.

1.1.2 Terminological inexactitude

A word of warning. In this document, and others referring to the expl3 programming modules, we often refer to “variables” and “functions” as if they were actual constructs from a real programming language. In truth, \TeX{} is a macro processor, and functions are simply macros that may or may not take arguments and expand to their replacement text. Many of the common variables are also macros, and if placed into the input stream will simply expand to their definition as well — a “function” with no arguments and a “token list variable” are almost the same. On the other hand, some “variables” are actually registers that must be initialised and their values set and retrieved with specific functions.

The conventions of the expl3 code are designed to clearly separate the ideas of “macros that contain data” and “macros that contain code”, and a consistent wrapper is applied to all forms of “data” whether they be macros or actually registers. This means that sometimes we will use phrases like “the function returns a value”, when actually we just mean “the macro expands to something”. Similarly, the term “execute” might be used in place of “expand” or it might refer to the more specific case of “processing in \TeX{}’s stomach” (if you are familiar with the \TeX{}book parlance).

If in doubt, please ask; chances are we’ve been hasty in writing certain definitions and need to be told to tighten up our terminology.

1.2 Documentation conventions

This document is typeset with the experimental \l3doc class; several conventions are used to help describe the features of the code. A number of conventions are used here to make the documentation clearer.

Each group of related functions is given in a box. For a function with a “user” name, this might read:

\begin{verbatim}
\ExplSyntaxOn
\ExplSyntaxOff
\end{verbatim}

The textual description of how the function works would appear here. The syntax of the function is shown in mono-spaced text to the right of the box. In this example, the function takes no arguments and so the name of the function is simply reprinted.

For programming functions, which use _ and : in their name there are a few additional conventions: If two related functions are given with identical names but different argument specifiers, these are termed variants of each other, and the latter functions are printed in grey to show this more clearly. They will carry out the same function but will take different types of argument:

\footnote{\TeX{}nically, functions with no arguments are \texttt{\long} while token list variables are not.}
When a number of variants are described, the arguments are usually illustrated only for
the base function. Here, \texttt{⟨sequence⟩} indicates that \texttt{\seq_new:N} expects the name of a
sequence. From the argument specifier, \texttt{\seq_new:c} also expects a sequence name, but
as a name rather than as a control sequence. Each argument given in the illustration
should be described in the following text.

**Fully expandable functions** Some functions are fully expandable, which allows them
to be used within an $x$-type or $e$-type argument (in plain \TeX{} terms, inside an \texttt{edef} or
\texttt{\expanded}), as well as within an $f$-type argument. These fully expandable functions are
indicated in the documentation by a star:

\texttt{\cs_to_str:N ★} \texttt{\cs_to_str:N ⟨cs⟩}

As with other functions, some text should follow which explains how the function works.
Usually, only the star will indicate that the function is expandable. In this case, the
function expects a \texttt{⟨cs⟩}, shorthand for a \texttt{⟨control sequence⟩}.

**Restricted expandable functions** A few functions are fully expandable but cannot
be fully expanded within an $f$-type argument. In this case a hollow star is used to indicate
this:

\texttt{\seq_map_function:NN ★} \texttt{\seq_map_function:NN ⟨seq⟩ ⟨function⟩}

**Conditional functions** Conditional \texttt{(if)} functions are normally defined in three vari-
ants, with \texttt{T}, \texttt{F} and \texttt{TF} argument specifiers. This allows them to be used for different
“true”/“false” branches, depending on which outcome the conditional is being used to
test. To indicate this without repetition, this information is given in a shortened form:

\texttt{\sys_if_engine_xetex:TF ★} \texttt{\sys_if_engine_xetex:TF ⟨true code⟩ ⟨false code⟩}

The underlining and italic of \texttt{TF} indicates that three functions are available:

- \texttt{\sys_if_engine_xetex:T}
- \texttt{\sys_if_engine_xetex:F}
- \texttt{\sys_if_engine_xetex:TF}

Usually, the illustration will use the \texttt{TF} variant, and so both \texttt{⟨true code⟩} and \texttt{⟨false code⟩}
will be shown. The two variant forms \texttt{T} and \texttt{F} take only \texttt{⟨true code⟩} and \texttt{⟨false code⟩},
respectively. Here, the star also shows that this function is expandable. With some minor
exceptions, all conditional functions in the \texttt{exp3} modules should be defined in this way.

Variables, constants and so on are described in a similar manner:

\texttt{\l_tmpa_tl}

A short piece of text will describe the variable: there is no syntax illustration in this case.

In some cases, the function is similar to one in \texttt{\LaTeX{}} or plain \TeX{}. In these cases,
the text will include an extra \texttt{“\TeX{}hackers note”} section:
The normal description text.

**TeXhackers note**: Detail for the experienced TeX or \LaTeX\ 2ε programmer. In this case, it would point out that this function is the TeX primitive `\string`.

**Changes to behaviour** When new functions are added to expl3, the date of first inclusion is given in the documentation. Where the documented behaviour of a function changes after it is first introduced, the date of the update will also be given. This means that the programmer can be sure that any release of expl3 after the date given will contain the function of interest with expected behaviour as described. Note that changes to code internals, including bug fixes, are not recorded in this way unless they impact on the expected behaviour.

### 1.3 Formal language conventions which apply generally

As this is a formal reference guide for \LaTeX\ 3 programming, the descriptions of functions are intended to be reasonably “complete”. However, there is also a need to avoid repetition. Formal ideas which apply to general classes of function are therefore summarised here.

For tests which have a \texttt{TF} argument specification, the test if evaluated to give a logically \texttt{TRUE} or \texttt{FALSE} result. Depending on this result, either the \texttt{true code} or the \texttt{false code} will be left in the input stream. In the case where the test is expandable, and a predicate (_p) variant is available, the logical value determined by the test is left in the input stream: this will typically be part of a larger logical construct.

### 1.4 TeX concepts not supported by \LaTeX\ 3

The TeX concept of an “\outer” macro is not supported at all by \LaTeX\ 3. As such, the functions provided here may break when used on top of \LaTeX\ 2ε if \outer tokens are used in the arguments.
Part II

Bootstrapping
Chapter 2

The \texttt{l3bootstrap} package

Bootstrap code

2.1 Using the \texttt{\LaTeX3} modules

The modules documented in \texttt{source3} are designed to be used on top of \texttt{\LaTeX2\epsilon} and are loaded all as one with the usual \texttt{\usepackage{expl3}} or \texttt{\RequirePackage{expl3}} instructions.

As the modules use a coding syntax different from standard \texttt{\LaTeX2\epsilon} it provides a few functions for setting it up.

\begin{verbatim}
\ExplSyntaxOn
⟨code⟩ \ExplSyntaxOff
\end{verbatim}

The \texttt{\ExplSyntaxOn} function switches to a category code regime in which spaces and new lines are ignored, and in which the colon (:) and underscore (_\texttt{\_}) are treated as “letters”, thus allowing access to the names of code functions and variables. Within this environment, - \texttt{\textasciitilde} is used to input a space. The \texttt{\ExplSyntaxOff} reverts to the document category code regime.

\textbf{\LaTeX3hackers note:} Spaces introduced by \texttt{\textasciitilde} behave much in the same way as normal space characters in the standard category code regime: they are ignored after a control word or at the start of a line, and multiple consecutive \texttt{\textasciitilde} are equivalent to a single one. However, \texttt{\textasciitilde} is not ignored at the end of a line.

\begin{verbatim}
\ProvidesExplPackage \RequiresExplPackage
\ProvidesExplClass \RequiresExplClass
\ProvidesExplFile \RequiresExplFile
\end{verbatim}

\texttt{\ProvidesExplPackage} \texttt{\ProvidesExplClass} and \texttt{\ProvidesExplFile} provide a similar role to \texttt{\ProvidesPackage} and \texttt{\ProvidesClass}. These functions act broadly in the same way as the corresponding \texttt{\LaTeX2\epsilon} kernel functions \texttt{\ProvidesPackage} and \texttt{\ProvidesClass} and \texttt{\ProvidesFile}. However, they also implicitly switch \texttt{\ExplSyntaxOn} for the remainder of the code with the file. At the end of the file, \texttt{\ExplSyntaxOff} will be called to reverse this. (This is the same concept as \texttt{\LaTeX2\epsilon} provides in turning on \texttt{\makeatletter} within package and class code.) The \texttt{\langle date\rangle} should be given in the format \texttt{(year)/(month)/(day)} or in the ISO date format \texttt{(year)-(month)-(day)}. If the \texttt{\langle version\rangle} is given then it will be prefixed with \texttt{v} in the package identifier line.
Extracts all information from a SVN field. Spaces are not ignored in these fields. The information pieces are stored in separate control sequences with \ExplFileName for the part of the file name leading up to the period, \ExplFileDate for date, \ExplFileVersion for version and \ExplFileDescription for the description.

To summarize: Every single package using this syntax should identify itself using one of the above methods. Special care is taken so that every package or class file loaded with \RequirePackage or similar are loaded with usual \LaTeX2ε category codes and the \LaTeX3 category code scheme is reloaded when needed afterwards. See implementation for details. If you use the \GetIdInfo command you can use the information when loading a package with

\ProvidesExplPackage{\ExplFileName}
{\ExplFileDate}{\ExplFileVersion}{\ExplFileDescription}
Chapter 3

The l3names package
Namespace for primitives

3.1 Setting up the \LaTeX3 programming language

This module is at the core of the \LaTeX3 programming language. It performs the following tasks:

- defines new names for all \TeX primitives;
- emulate required primitives not provided by default in \LuaTeX;
- switches to the category code régime for programming;

This module is entirely dedicated to primitives (and emulations of these), which should not be used directly within \LaTeX3 code (outside of “kernel-level” code). As such, the primitives are not documented here: *The \TeXbook*, \TeX by Topic and the manuals for pdf\TeX, Xe\TeX, \Lua\TeX, p\TeX and up\TeX should be consulted for details of the primitives. These are named \texttt{\tex_⟨name⟩;D}, typically based on the primitive’s ⟨name⟩ in pdf\TeX and omitting a leading \texttt{pdf} when the primitive is not related to pdf output.
Part III
Programming Flow
Chapter 4

The \texttt{l3basics} package

Basic definitions

As the name suggest this package holds some basic definitions which are needed by most or all other packages in this set.

Here we describe those functions that are used all over the place. With that we mean functions dealing with the construction and testing of control sequences. Furthermore the basic parts of conditional processing are covered; conditional processing dealing with specific data types is described in the modules specific for the respective data types.

4.1 No operation functions

\prg_do_nothing: 
An expandable function which does nothing at all: leaves nothing in the input stream after a single expansion.

\scan_stop: 
A non-expandable function which does nothing. Does not vanish on expansion but produces no typeset output.

4.2 Grouping material

\group_begin: 
\group_end: 
These functions begin and end a group for definition purposes. Assignments are local to groups unless carried out in a global manner. (A small number of exceptions to this rule will be noted as necessary elsewhere in this document.) Each \texttt{\group_begin}: must be matched by a \texttt{\group_end}:, although this does not have to occur within the same function. Indeed, it is often necessary to start a group within one function and finish it within another, for example when seeking to use non-standard category codes.
\group_insert_after:N \group_insert_after:N \langle token \rangle

Adds \langle token \rangle to the list of \langle tokens \rangle to be inserted when the current group level ends. The list of \langle tokens \rangle to be inserted is empty at the beginning of a group: multiple applications of \group_insert_after:N may be used to build the inserted list one \langle token \rangle at a time. The current group level may be closed by a \group_end: function or by a token with category code 2 (close-group), namely a \} if standard category codes apply.

\group_show_list: \group_show_list:
\group_log_list: \group_log_list:

Display (to the terminal or log file) a list of the groups that are currently opened. This is intended for tracking down problems.

\TeX{}hackers note: This is a wrapper around the \showgroups primitive.

4.3 Control sequences and functions

As \TeX{} is a macro language, creating new functions means creating macros. At point of use, a function is replaced by the replacement text (“code”) in which each parameter in the code \(#1, \#2, \ldots\) is replaced the appropriate arguments absorbed by the function. In the following, \langle code \rangle is therefore used as a shorthand for “replacement text”.

Functions which are not “protected” are fully expanded inside an x expansion. In contrast, “protected” functions are not expanded within x expansions.

4.3.1 Defining functions

Functions can be created with no requirement that they are declared first (in contrast to variables, which must always be declared). Declaring a function before setting up the code means that the name chosen is checked and an error raised if it is already in use. The name of a function can be checked at the point of definition using the \cs_new\ldots functions: this is recommended for all functions which are defined for the first time.

There are three ways to define new functions. All classes define a function to expand to the substitution text. Within the substitution text the actual parameters are substituted for the formal parameters \(#1, \#2, \ldots\).

\new Create a new function with the \new scope, such as \cs_new:Npn. The definition is global and results in an error if it is already defined.

\set Create a new function with the \set scope, such as \cs_set:Npn. The definition is restricted to the current \TeX{} group and does not result in an error if the function is already defined.

\gset Create a new function with the \gset scope, such as \cs_gset:Npn. The definition is global and does not result in an error if the function is already defined.

Within each set of scope there are different ways to define a function. The differences depend on restrictions on the actual parameters and the expandability of the resulting function.

\nopar Create a new function with the \nopar restriction, such as \cs_set_nopar:Npn. The parameter may not contain \par tokens.
protected Create a new function with the protected restriction, such as `\cs_set_-_protected:Npn`. The parameter may contain `\par` tokens but the function will not expand within an x-type or e-type expansion.

Finally, the functions in Subsections 4.3.2 and 4.3.3 are primarily meant to define **base functions** only. Base functions can only have the following argument specifiers:

- **N** and **n** No manipulation.
- **T** and **F** Functionally equivalent to **n** (you are actually encouraged to use the family of `\prg_new_conditional` functions described in Section 9.1).
- **p** and **w** These are special cases.

The `\cs_new:` functions below (and friends) do not stop you from using other argument specifiers in your function names, but they do not handle expansion for you. You should define the base function and then use `\cs_generate_variant:Nn` to generate custom variants as described in Section 5.2.

### 4.3.2 Defining new functions using parameter text

```
\cs_new:Npn \cs_new_nopar:Npn \cs_new_protected:Npn \cs_new_protected_nopar:Npn
\cs_new:cpn \cs_new_nopar:cpn \cs_new_protected:cpn \cs_new_protected_nopar:cpn
\cs_new:Npx \cs_new_nopar:Npx \cs_new_protected:Npx \cs_new_protected_nopar:Npx
\cs_new:cpx \cs_new_nopar:cpx \cs_new_protected:cpx \cs_new_protected_nopar:cpx
```

**\cs_new:Npn** `\cs_new:Npn \cs_new_nopar:Npn \cs_new_protected:Npn \cs_new_protected_nopar:Npn` **(function) (parameters) \{\{code\}\}**

Creates **(function)** to expand to **\{\{code\}\}** as replacement text. Within the (**\{\{code\}\}**), the (**parameters**) (**#1**, **#2**, etc.) will be replaced by those absorbed by the function. The definition is global and an error results if the **(function)** is already defined.

**\cs_new_nopar:Npn** `\cs_new_nopar:Npn \cs_new_nopar:cpn \cs_new_nopar:Npx \cs_new_nopar:cpx` **(function) (parameters) \{\{code\}\}**

Creates **(function)** to expand to **\{\{code\}\}** as replacement text. Within the (**\{\{code\}\}**), the (**parameters**) (**#1**, **#2**, etc.) will be replaced by those absorbed by the function. When the **(function)** is used the (**parameters**) absorbed cannot contain `\par` tokens. The definition is global and an error results if the **(function)** is already defined.

**\cs_new_protected:Npn** `\cs_new_protected:Npn \cs_new_protected:cpn \cs_new_protected:Npx \cs_new_protected:cpx` **(function) (parameters) \{\{code\}\}**

Creates **(function)** to expand to **\{\{code\}\}** as replacement text. Within the (**\{\{code\}\}**), the (**parameters**) (**#1**, **#2**, etc.) will be replaced by those absorbed by the function. The **(function)** will not expand within an x-type argument. The definition is global and an error results if the **(function)** is already defined.


Creates **(function)** to expand to **\{\{code\}\}** as replacement text. Within the (**\{\{code\}\}**), the (**parameters**) (**#1**, **#2**, etc.) will be replaced by those absorbed by the function. When the **(function)** is used the (**parameters**) absorbed cannot contain `\par` tokens. The **(function)** will not expand within an x-type or e-type argument. The definition is global and an error results if the **(function)** is already defined.
\cs_set:Npn \cs_set:cpn \cs_set:Npx \cs_set:cpx

Sets \textit{function} to expand to \textit{code} as replacement text. Within the \textit{code}, the \textit{parameters} \((#1, #2, \text{etc})\) will be replaced by those absorbed by the function. The assignment of a meaning to the \textit{function} is restricted to the current \TeX{} group level.

\cs_set_nopar:Npn \cs_set_nopar:cpn \cs_set_nopar:Npx \cs_set_nopar:cpx

Sets \textit{function} to expand to \textit{code} as replacement text. Within the \textit{code}, the \textit{parameters} \((#1, #2, \text{etc})\) will be replaced by those absorbed by the function. When the \textit{function} is used the \textit{parameters} absorbed cannot contain \texttt{par} tokens. The assignment of a meaning to the \textit{function} is restricted to the current \TeX{} group level.

\cs_set_protected:Npn \cs_set_protected:cpn \cs_set_protected:Npx \cs_set_protected:cpx

Sets \textit{function} to expand to \textit{code} as replacement text. Within the \textit{code}, the \textit{parameters} \((#1, #2, \text{etc})\) will be replaced by those absorbed by the function. The assignment of a meaning to the \textit{function} is restricted to the current \TeX{} group level. The \textit{function} will not expand within an \texttt{x}-type or \texttt{e}-type argument.

\cs_set_protected_nopar:Npn \cs_set_protected_nopar:cpn \cs_set_protected_nopar:Npx \cs_set_protected_nopar:cpx

Sets \textit{function} to expand to \textit{code} as replacement text. Within the \textit{code}, the \textit{parameters} \((#1, #2, \text{etc})\) will be replaced by those absorbed by the function. When the \textit{function} is used the \textit{parameters} absorbed cannot contain \texttt{par} tokens. The assignment of a meaning to the \textit{function} is restricted to the current \TeX{} group level. The \textit{function} will not expand within an \texttt{x}-type or \texttt{e}-type argument.

\cs_gset:Npn \cs_gset:cpn \cs_gset:Npx \cs_gset:cpx

Globally sets \textit{function} to expand to \textit{code} as replacement text. Within the \textit{code}, the \textit{parameters} \((#1, #2, \text{etc})\) will be replaced by those absorbed by the function. The assignment of a meaning to the \textit{function} is not restricted to the current \TeX{} group level: the assignment is global.

\cs_gset_nopar:Npn \cs_gset_nopar:cpn \cs_gset_nopar:Npx \cs_gset_nopar:cpx

Globally sets \textit{function} to expand to \textit{code} as replacement text. Within the \textit{code}, the \textit{parameters} \((#1, #2, \text{etc})\) will be replaced by those absorbed by the function. When the \textit{function} is used the \textit{parameters} absorbed cannot contain \texttt{par} tokens. The assignment of a meaning to the \textit{function} is not restricted to the current \TeX{} group level: the assignment is global.

\cs_gset_protected:Npn \cs_gset_protected:cpn \cs_gset_protected:Npx \cs_gset_protected:cpx

Globally sets \textit{function} to expand to \textit{code} as replacement text. Within the \textit{code}, the \textit{parameters} \((#1, #2, \text{etc})\) will be replaced by those absorbed by the function. The assignment of a meaning to the \textit{function} is not restricted to the current \TeX{} group level: the assignment is global. The \textit{function} will not expand within an \texttt{x}-type or \texttt{e}-type argument.
<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\cs_gset_protected_nopar:Np</td>
<td>Globally sets \langle function\rangle to expand to \langle code\rangle as replacement text. Within the \langle code\rangle, the \langle parameters\rangle (#1, #2, etc.) will be replaced by those absorbed by the function. When the \langle function\rangle is used the \langle parameters\rangle absorbed cannot contain \par tokens. The assignment of a meaning to the \langle function\rangle is not restricted to the current \TeX{} group level: the assignment is global. The \langle function\rangle will not expand within an x-type argument.</td>
</tr>
</tbody>
</table>

### 4.3.3 Defining new functions using the signature

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\cs_new:Nn</td>
<td>Creates \langle function\rangle to expand to \langle code\rangle as replacement text. Within the \langle code\rangle, the number of \langle parameters\rangle is detected automatically from the function signature. These \langle parameters\rangle (#1, #2, etc.) will be replaced by those absorbed by the function. The definition is global and an error results if the \langle function\rangle is already defined.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\cs_new_nopar:Nn</td>
<td>Creates \langle function\rangle to expand to \langle code\rangle as replacement text. Within the \langle code\rangle, the number of \langle parameters\rangle is detected automatically from the function signature. These \langle parameters\rangle (#1, #2, etc.) will be replaced by those absorbed by the function. When the \langle function\rangle is used the \langle parameters\rangle absorbed cannot contain \par tokens. The definition is global and an error results if the \langle function\rangle is already defined.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\cs_new_protected:Nn</td>
<td>Creates \langle function\rangle to expand to \langle code\rangle as replacement text. Within the \langle code\rangle, the number of \langle parameters\rangle is detected automatically from the function signature. These \langle parameters\rangle (#1, #2, etc.) will be replaced by those absorbed by the function. The \langle function\rangle will not expand within an x-type argument. The definition is global and an error results if the \langle function\rangle is already defined.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\cs_new_protected_nopar:Nn</td>
<td>Creates \langle function\rangle to expand to \langle code\rangle as replacement text. Within the \langle code\rangle, the number of \langle parameters\rangle is detected automatically from the function signature. These \langle parameters\rangle (#1, #2, etc.) will be replaced by those absorbed by the function. When the \langle function\rangle is used the \langle parameters\rangle absorbed cannot contain \par tokens. The \langle function\rangle will not expand within an x-type or e-type argument. The definition is global and an error results if the \langle function\rangle is already defined.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\cs_set:Nn</td>
<td>Sets \langle function\rangle to expand to \langle code\rangle as replacement text. Within the \langle code\rangle, the number of \langle parameters\rangle is detected automatically from the function signature. These \langle parameters\rangle (#1, #2, etc.) will be replaced by those absorbed by the function. The assignment of a meaning to the \langle function\rangle is restricted to the current \TeX{} group level.</td>
</tr>
</tbody>
</table>
\cs_set_nopar:Nn \cs_set_nopar:(cn|Nx|cx) \cs_set_protected:Nn \cs_set_protected:(cn|Nx|cx) \cs_set_protected_nopar:Nn \cs_set_protected_nopar:(cn|Nx|cx) \cs_gset:Nn \cs_gset:(cn|Nx|cx) \cs_gset_nopar:Nn \cs_gset_nopar:(cn|Nx|cx) \cs_gset_protected:Nn \cs_gset_protected:(cn|Nx|cx) \cs_gset_protected_nopar:Nn \cs_gset_protected_nopar:(cn|Nx|cx)

Sets \texttt{(function)} to expand to \texttt{(code)} as replacement text. Within the \texttt{(code)}, the number of \texttt{parameters} is detected automatically from the function signature. These \texttt{parameters} (\#1, \#2, etc.) will be replaced by those absorbed by the function. When the \texttt{(function)} is used the \texttt{parameters} absorbed cannot contain \texttt{\par} tokens. The assignment of a meaning to the \texttt{(function)} is restricted to the current \TeX{} group level.

Sets \texttt{(function)} to expand to \texttt{(code)} as replacement text. Within the \texttt{(code)}, the number of \texttt{parameters} is detected automatically from the function signature. These \texttt{parameters} (\#1, \#2, etc.) will be replaced by those absorbed by the function. The \texttt{(function)} will not expand within an \texttt{x}-type argument. The assignment of a meaning to the \texttt{(function)} is restricted to the current \TeX{} group level.

Sets \texttt{(function)} to expand to \texttt{(code)} as replacement text. Within the \texttt{(code)}, the number of \texttt{parameters} is detected automatically from the function signature. These \texttt{parameters} (\#1, \#2, etc.) will be replaced by those absorbed by the function. The assignment of a meaning to the \texttt{(function)} is global.

Sets \texttt{(function)} to expand to \texttt{(code)} as replacement text. Within the \texttt{(code)}, the number of \texttt{parameters} is detected automatically from the function signature. These \texttt{parameters} (\#1, \#2, etc.) will be replaced by those absorbed by the function. The \texttt{(function)} will not expand within an \texttt{x}-type argument. The assignment of a meaning to the \texttt{(function)} is global.
Sets \( \text{function} \) to expand to \( \text{code} \) as replacement text. Within the \( \text{code} \), the number of \( \text{parameters} \) is detected automatically from the function signature. These \( \text{parameters} \) (#1, #2, etc.) will be replaced by those absorbed by the function. When the \( \text{function} \) is used the \( \text{parameters} \) absorbed cannot contain \( \text{par} \) tokens. The \( \text{function} \) will not expand within an \( \text{x-type} \) or \( \text{e-type} \) argument. The assignment of a meaning to the \( \text{function} \) is global.

Uses the \( \text{creator} \) function (which should have signature \( \text{Npn} \), for example \( \cs_new:Npn \)) to define a \( \text{function} \) which takes \( \text{number} \) arguments and has \( \text{code} \) as replacement text. The \( \text{number} \) of arguments is an integer expression, evaluated as detailed for \( \text{int_eval:n} \).

### 4.3.4 Copying control sequences

Control sequences (not just functions as defined above) can be set to have the same meaning using the functions described here. Making two control sequences equivalent means that the second control sequence is a *copy* of the first (rather than a pointer to it). Thus the old and new control sequence are not tied together: changes to one are not reflected in the other.

In the following text “cs” is used as an abbreviation for “control sequence”.

Globally creates \( \text{control sequence}_1 \) and sets it to have the same meaning as \( \text{control sequence}_2 \) or \( \text{token} \). The second control sequence may subsequently be altered without affecting the copy.

Sets \( \text{control sequence}_1 \) to have the same meaning as \( \text{control sequence}_2 \) (or \( \text{token} \)). The second control sequence may subsequently be altered without affecting the copy. The assignment of a meaning to the \( \text{control sequence}_1 \) is restricted to the current \( \text{TeX} \) group level.

Globally sets \( \text{control sequence}_1 \) to have the same meaning as \( \text{control sequence}_2 \) (or \( \text{token} \)). The second control sequence may subsequently be altered without affecting the copy. The assignment of a meaning to the \( \text{control sequence}_1 \) is *not* restricted to the current \( \text{TeX} \) group level: the assignment is global.
4.3.5 Deleting control sequences

There are occasions where control sequences need to be deleted. This is handled in a very simple manner.

\cs_undefine:N \langle control sequence \rangle

Sets \langle control sequence \rangle to be globally undefined.

4.3.6 Showing control sequences

\cs_meaning:N \langle control sequence \rangle

This function expands to the meaning of the \langle control sequence \rangle control sequence. For a macro, this includes the \langle replacement text \rangle.

\textbf{TeXhackers note:} This is \TeX 's \meaning primitive. For tokens that are not control sequences, it is more logical to use \token_to_meaning:N. The \texttt{c} variant correctly reports undefined arguments.

\cs_show:N \langle control sequence \rangle

Displays the definition of the \langle control sequence \rangle on the terminal.

\textbf{TeXhackers note:} This is similar to the \TeX primitive \show, wrapped to a fixed number of characters per line.

\cs_log:N \langle control sequence \rangle

Writes the definition of the \langle control sequence \rangle in the log file. See also \cs_show:N which displays the result in the terminal.

4.3.7 Converting to and from control sequences

\use:c { \langle control sequence name \rangle }

Expands the \langle control sequence name \rangle until only characters remain, and then converts this into a control sequence. This process requires two expansions. As in other \texttt{c}-type arguments the \langle control sequence name \rangle must, when fully expanded, consist of character tokens, typically a mixture of category code 10 (space), 11 (letter) and 12 (other).

\textbf{TeXhackers note:} Protected macros that appear in a \texttt{c}-type argument are expanded despite being protected; \exp_not:n also has no effect. An internal error occurs if non-characters or active characters remain after full expansion, as the conversion to a control sequence is not possible.

As an example of the \texttt{\use:c} function, both
\use:c { a b c }

and
\tl_new:N \l_my_tl
\tl_set:Nn \l_my_tl { a b c }
\use:c { \tl_use:N \l_my_tl }

would be equivalent to
\abc

after two expansions of \use:c.

\cs_if_exist_use:N \cs_if_exist_use:c \cs_if_exist_use:Ntf \cs_if_exist_use:Ctf

Tests whether the \textit{control sequence} is currently defined according to the conditional \cs_if_exist_use:Ntf (whether as a function or another control sequence type), and if it is inserts the \textit{control sequence} into the input stream followed by the \textit{true code}. Otherwise the \textit{false code} is used.

\cs:w \cs_end:

Converts the given \textit{control sequence name} into a single control sequence token. This process requires one expansion. The content for \textit{control sequence name} may be literal material or from other expandable functions. The \textit{control sequence name} must, when fully expanded, consist of character tokens which are not active: typically of category code 10 (space), 11 (letter) or 12 (other), or a mixture of these.

\TeX\hacksnote: These are the \TeX\ primitives \texttt{csname} and \texttt{endcsname}.

As an example of the \texttt{cs:w} and \texttt{cs_end:} functions, both
\cs:w a b c \cs_end:

and
\tl_new:N \l_my_tl
\tl_set:Nn \l_my_tl { a b c }
\cs:w \tl_use:N \l_my_tl \cs_end:

would be equivalent to
\abc

after one expansion of \cs:w.

\cs_to_str:N \cs_to_str:N (control sequence)

Converts the given \textit{control sequence} into a series of characters with category code 12 (other), except spaces, of category code 10. The result does not include the current escape token, contrarily to \texttt{token_to_str:N}. Full expansion of this function requires exactly 2 expansion steps, and so an \texttt{x}-type or \texttt{e}-type expansion, or two \texttt{o}-type expansions are required to convert the \textit{control sequence} to a sequence of characters in the input stream. In most cases, an \texttt{f}-expansion is correct as well, but this loses a space at the start of the result.
4.4 Analysing control sequences

\cs_split_function:N \langle function \rangle

Splits the \langle function \rangle into the \langle name \rangle (i.e. the part before the colon) and the \langle signature \rangle (i.e. after the colon). This information is then placed in the input stream in three parts: the \langle name \rangle, the \langle signature \rangle and a logic token indicating if a colon was found (to differentiate variables from function names). The \langle name \rangle does not include the escape character, and both the \langle name \rangle and \langle signature \rangle are made up of tokens with category code 12 (other).

The next three functions decompose \TeX macros into their constituent parts: if the \langle token \rangle passed is not a macro then no decomposition can occur. In the latter case, all three functions leave \texttt{\scan_stop;} in the input stream.

\cs_prefix_spec:N \langle token \rangle

If the \langle token \rangle is a macro, this function leaves the applicable \TeX prefixes in input stream as a string of tokens of category code 12 (with spaces having category code 10). Thus for example

\begin{verbatim}
\cs_set:Npn \next:nn \#1#2 \{ x \#1-y \#2 \}
\cs_prefix_spec:N \next:nn
\end{verbatim}

leaves \texttt{\long} in the input stream. If the \langle token \rangle is not a macro then \texttt{\scan_stop;} is left in the input stream.

\TeXhackers note: The prefix can be empty, \texttt{\long}, \texttt{\protected} or \texttt{\protected\long} with backslash replaced by the current escape character.

\cs_argument_spec:N \langle token \rangle

If the \langle token \rangle is a macro, this function leaves the primitive \TeX argument specification in input stream as a string of character tokens of category code 12 (with spaces having category code 10). Thus for example

\begin{verbatim}
\cs_set:Npn \next:nn \#1#2 \{ x \#1 y \#2 \}
\cs_argument_spec:N \next:nn
\end{verbatim}

leaves \texttt{\#1\#2} in the input stream. If the \langle token \rangle is not a macro then \texttt{\scan_stop;} is left in the input stream.

\TeXhackers note: If the argument specification contains the string ->, then the function produces incorrect results.
\texttt{\cs_replacement_spec:N} \texttt{⟨token⟩}

If the \texttt{(token)} is a macro, this function leaves the replacement text in input stream as a string of character tokens of category code 12 (with spaces having category code 10). Thus for example

\begin{verbatim}
\cs_set:Npn \next:nn \#1\#2 \{ x \#1-y \#2 \}
\cs_replacement_spec:N \next:nn
\end{verbatim}

leaves \texttt{x\#1 \#y\#2} in the input stream. If the \texttt{(token)} is not a macro then \texttt{\scan_stop:} is left in the input stream.

\textbf{\TeXhackers note:} If the argument specification contains the string \texttt{->}, then the function produces incorrect results.

\section*{4.5 Using or removing tokens and arguments}

Tokens in the input can be read and used or read and discarded. If one or more tokens are wrapped in braces then when absorbing them the outer set is removed. At the same time, the category code of each token is set when the token is read by a function (if it is read more than once, the category code is determined by the situation in force when first function absorbs the token).

\begin{verbatim}
\use:n \{⟨group₁⟩\}
\use:nn \{⟨group₁⟩\} \{⟨group₂⟩\}
\use:nnn \{⟨group₁⟩\} \{⟨group₂⟩\} \{⟨group₃⟩\}
\use:nnnn \{⟨group₁⟩\} \{⟨group₂⟩\} \{⟨group₃⟩\} \{⟨group₄⟩\}
\end{verbatim}

As illustrated, these functions absorb between one and four arguments, as indicated by the argument specifier. The braces surrounding each argument are removed and the remaining tokens are left in the input stream. The category code of these tokens is also fixed by this process (if it has not already been by some other absorption). All of these functions require only a single expansion to operate, so that one expansion of

\begin{verbatim}
\use:nn \{ abc \} \{ \{ def \} \}
\end{verbatim}

results in the input stream containing

\begin{verbatim}
abc \{ def \}
\end{verbatim}

\textit{i.e.} only the outer braces are removed.

\textbf{\TeXhackers note:} The \texttt{\use:n} function is equivalent to \LaTeX{}\texttt{2e}'s \texttt{\@firstofone}.

23
These functions absorb two arguments from the input stream. The function \usei:nn discards the second argument, and leaves the content of the first argument in the input stream. \useii:nn discards the first argument and leaves the content of the second argument in the input stream. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the functions to take effect.

\TeXhackers note: These are equivalent to \LaTeX's \@firstoftwo and \@secondoftwo.

These functions absorb three arguments from the input stream. The function \usei:nnn discards the second and third arguments, and leaves the content of the first argument in the input stream. \useii:nnn and \useiii:nnn work similarly, leaving the content of second or third arguments in the input stream, respectively. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the functions to take effect.

These functions absorb four arguments from the input stream. The function \usei:nnnn discards the second, third and fourth arguments, and leaves the content of the first argument in the input stream. \useii:nnnn, \useiii:nnnn and \useiv:nnnn work similarly, leaving the content of second, third or fourth arguments in the input stream, respectively. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the functions to take effect.

This function absorbs three arguments and leaves the content of the first and second in the input stream. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the function to take effect. An example:

\usei:nnn { abc } { \{ def \} } { ghi }

results in the input stream containing

abc { def }

i.e. the outer braces are removed and the third group is removed.

This function absorbs two arguments and leaves the content of the second and first in the input stream. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the function to take effect.
These functions absorb between one and nine groups from the input stream, leaving nothing on the resulting input stream. These functions work after a single expansion. One or more of the n arguments may be an unbraced single token (i.e. an N argument).

\textbf{\LaTeXhackers note:} These are equivalent to \LaTeXeX 2e's \texttt{@gobble}, \texttt{@gobbletwo}, etc.

\texttt{\use:e \langle expandable tokens \rangle}

Fully expands the \langle token list \rangle in an \texttt{x}-type manner, \textit{but} the function remains fully expandable, and parameter character (usually \#) need not be doubled.

\textbf{\LaTeXhackers note:} \texttt{\use:e} is a wrapper around the primitive \texttt{\expanded} where it is available: it requires two expansions to complete its action. When \texttt{\expanded} is not available this function is very slow.

\texttt{\use:x \langle expandable tokens \rangle}

Fully expands the \langle expandable tokens \rangle and inserts the result into the input stream at the current location. Any hash characters (\#) in the argument must be doubled.

### 4.5.1 Selecting tokens from delimited arguments

A different kind of function for selecting tokens from the token stream are those that use delimited arguments.

Absorb the \langle balanced text \rangle from the input stream delimited by the marker given in the function name, leaving nothing in the input stream.

Absorb the \langle balanced text \rangle from the input stream delimited by the marker given in the function name, leaving \langle inserted tokens \rangle in the input stream for further processing.

### 4.6 Predicates and conditionals

\LaTeXXeX has three concepts for conditional flow processing:
Branching conditionals Functions that carry out a test and then execute, depending on its result, either the code supplied as the \textit{(true code)} or the \textit{(false code)}. These arguments are denoted with \texttt{T} and \texttt{F}, respectively. An example would be

\begin{verbatim}
\cs_if_free:cTF {abc} {⟨true code⟩} {⟨false code⟩}
\end{verbatim}
a function that turns the first argument into a control sequence (since it’s marked as \texttt{c}) then checks whether this control sequence is still free and then depending on the result carries out the code in the second argument (true case) or in the third argument (false case).

These type of functions are known as “conditionals”: whenever a \texttt{TF} function is defined it is usually accompanied by \texttt{T} and \texttt{F} functions as well. These are provided for convenience when the branch only needs to go a single way. Package writers are free to choose which types to define but the kernel definitions always provide all three versions.

Important to note is that these branching conditionals with \textit{(true code)} and/or \textit{(false code)} are always defined in a way that the code of the chosen alternative can operate on following tokens in the input stream.

These conditional functions may or may not be fully expandable, but if they are expandable they are accompanied by a “predicate” for the same test as described below.

Predicates “Predicates” are functions that return a special type of boolean value which can be tested by the boolean expression parser. All functions of this type are expandable and have names that end with \texttt{p} in the description part. For example,

\begin{verbatim}
\cs_if_free_p:N
\end{verbatim}

would be a predicate function for the same type of test as the conditional described above. It would return “true” if its argument (a single token denoted by \texttt{N}) is still free for definition. It would be used in constructions like

\begin{verbatim}
\bool_if:nTF { \cs_if_free_p:N \l_tmpz_tl || \cs_if_free_p:N \g_tmpz_tl } {⟨true code⟩} {⟨false code⟩}
\end{verbatim}

For each predicate defined, a “branching conditional” also exists that behaves like a conditional described above.

Primitive conditionals There is a third variety of conditional, which is the original concept used in plain \LaTeX{} and \LaTeX{}2\epsilon. Their use is discouraged in expl3 (although still used in low-level definitions) because they are more fragile and in many cases require more expansion control (hence more code) than the two types of conditionals described above.

\begin{verbatim}
\c_true_bool \c_false_bool
\end{verbatim}

Constants that represent \texttt{true} and \texttt{false}, respectively. Used to implement predicates.
4.6.1 Tests on control sequences

\cs_if_eq_p:NN \cs_if_eq_p:NN \langle cs1 \rangle \langle cs2 \rangle
\cs_if_eq:NNTF \cs_if_eq:NNTF \langle cs1 \rangle \langle cs2 \rangle \{ \langle true code \rangle \} \{ \langle false code \rangle \}

Compares the definition of two \langle control sequences \rangle and is logically \texttt{true} if they are the same, \textit{i.e.} if they have exactly the same definition when examined with \texttt{\cs_show:N}.

\cs_if_exist_p:N \cs_if_exist_p:N \langle control sequence \rangle
\cs_if_exist:NTF \cs_if_exist:NTF \langle control sequence \rangle \{ \langle true code \rangle \} \{ \langle false code \rangle \}

Tests whether the \langle control sequence \rangle is currently defined (whether as a function or another control sequence type). Any definition of \langle control sequence \rangle other than \texttt{\relax} evaluates as true.

\cs_if_free_p:N \cs_if_free_p:N \langle control sequence \rangle
\cs_if_free:NTF \cs_if_free:NTF \langle control sequence \rangle \{ \langle true code \rangle \} \{ \langle false code \rangle \}

Tests whether the \langle control sequence \rangle is currently free to be defined. This test is \texttt{false} if the \langle control sequence \rangle currently exists (as defined by \texttt{\cs_if_exist:NTF}).

4.6.2 Primitive conditionals

The \TeX{} engine itself provides many different conditionals. Some expand whatever comes after them and others don’t. Hence the names for these underlying functions often contains a \texttt{w} part but higher level functions are often available. See for instance \texttt{\int_compare_p:nNn} which is a wrapper for \texttt{\if_int_compare:w}.

Certain conditionals deal with specific data types like boxes and fonts and are described there. The ones described below are either the universal conditionals or deal with control sequences. We prefix primitive conditionals with \texttt{\if_}.

\texttt{\if_true:} \langle true code \rangle \else: \langle false code \rangle \fi:
\texttt{\if_false:} \langle true code \rangle \else: \langle false code \rangle \fi:
\texttt{\else:} \langle true code \rangle \else: \langle false code \rangle \fi:
\texttt{\reverse_if:N} \langle primitive conditional \rangle
\texttt{\if:} \langle true code \rangle, while \texttt{\if_false:} always executes \langle false code \rangle. \texttt{\reverse_if:N} reverses any two-way primitive conditional. \texttt{\else:} and \texttt{\fi:} delimit the branches of the conditional. The function \texttt{\or:} is documented in \texttt{l3int} and used in case switches.

\TeX{}hackers note: These are equivalent to their corresponding \TeX{} primitive conditionals; \texttt{\reverse_if:N} is \TeX{}’s \texttt{unless}.

\texttt{\if_meaning:w} \langle arg1 \rangle \langle arg2 \rangle \langle true code \rangle \else: \langle false code \rangle \fi:
\texttt{\if_meaning:w} executes \langle true code \rangle when \langle arg1 \rangle and \langle arg2 \rangle are the same, otherwise it executes \langle false code \rangle. \langle arg1 \rangle and \langle arg2 \rangle could be functions, variables, tokens; in all cases the \texttt{unexpanded} definitions are compared.

\TeX{}hackers note: This is \TeX{}’s \texttt{\ifx}.
These conditionals expand any following tokens until two unexpandable tokens are left. If you wish to prevent this expansion, prefix the token in question with \exp_not:N. \if_catcode:w tests if the category codes of the two tokens are the same whereas \if:w tests if the character codes are identical. \if_charcode:w is an alternative name for \if:w.

\if_cs_exist:N \if_cs_exist:w \if_mode_horizontal: \if_mode_vertical: \if_mode_math: \if_mode_inner: \mode_leave_vertical: 

4.7 Starting a paragraph

\mode_leave_vertical: 

\TeXhacks note: This results in the contents of the \everypar token register being inserted, after \mode_leave_vertical: is complete. Notice that in contrast to the \TeXe \leavevmode approach, no box is used by the method implemented here.
4.8 Debugging support

\debug_on:n \debug_off:n

\debug_on:n \{ \text{comma-separated list} \} \debug_off:n \{ \text{comma-separated list} \}

Turn on and off within a group various debugging code, some of which is also available as expl3 load-time options. The items that can be used in the \langle list \rangle are

- \textbf{check-declarations} that checks all expl3 variables used were previously declared and that local/global variables (based on their name or on their first assignment) are only locally/globally assigned;
- \textbf{check-expressions} that checks integer, dimension, skip, and muskip expressions are not terminated prematurely;
- \textbf{deprecation} that makes soon-to-be-deprecated commands produce errors;
- \textbf{log-functions} that logs function definitions;
- \textbf{all} that does all of the above.

Providing these as switches rather than options allows testing code even if it relies on other packages: load all other packages, call \texttt{\debug_on:n}, and load the code that one is interested in testing. These functions can only be used in \LaTeX{}2ε package mode loaded with \texttt{enable-debug} or another option implying it.

\debug_suspend: ... \debug_resume:

Suppress (locally) errors and logging from \texttt{debug} commands, except for the \texttt{deprecation} errors or warnings. These pairs of commands can be nested. This can be used around pieces of code that are known to fail checks, if such failures should be ignored. See for instance \texttt{l3coffins}.
Chapter 5

The \texttt{l3expan} package

Argument expansion

This module provides generic methods for expanding \TeX\ arguments in a systematic manner. The functions in this module all have prefix \texttt{exp}.

Not all possible variations are implemented for every base function. Instead only those that are used within the \LaTeXX\ kernel or otherwise seem to be of general interest are implemented. Consult the module description to find out which functions are actually defined. The next section explains how to define missing variants.

5.1 Defining new variants

The definition of variant forms for base functions may be necessary when writing new functions or when applying a kernel function in a situation that we haven’t thought of before.

Internally preprocessing of arguments is done with functions of the form \texttt{\exp_- ...}. They all look alike, an example would be \texttt{\exp\_args:NNo}. This function has three arguments, the first and the second are a single tokens, while the third argument should be given in braces. Applying \texttt{\exp\_args:NNo} expands the content of third argument once before any expansion of the first and second arguments. If \texttt{\seq_gpush:No} was not defined it could be coded in the following way:

\begin{verbatim}
\exp_args:NNo \seq_gpush:Nn
\g_file_name_stack
{ \l_tmpa_tl }
\end{verbatim}

In other words, the first argument to \texttt{\exp\_args:NNo} is the base function and the other arguments are preprocessed and then passed to this base function. In the example the first argument to the base function should be a single token which is left unchanged while the second argument is expanded once. From this example we can also see how the variants are defined. They just expand into the appropriate \texttt{\exp\_} function followed by the desired base function, \textit{e.g.}

\begin{verbatim}
\cs_generate_variant:Nn \seq_gpush:Nn { No }
\end{verbatim}

results in the definition of \texttt{\seq_gpush:No}
Providing variants in this way in style files is safe as the `\cs_generate_variant:Nn` function will only create new definitions if there is not already one available. Therefore adding such definition to later releases of the kernel will not make such style files obsolete.

The steps above may be automated by using the function `\cs_generate_variant:Nn`, described next.

### 5.2 Methods for defining variants

We recall the set of available argument specifiers.

- `N` is used for single-token arguments while `c` constructs a control sequence from its name and passes it to a parent function as an `N`-type argument.

- Many argument types extract or expand some tokens and provide it as an `n`-type argument, namely a braced multiple-token argument: `V` extracts the value of a variable, `v` extracts the value from the name of a variable, `n` uses the argument as it is, `o` expands once, `f` expands fully the front of the token list, `e` and `x` expand fully all tokens (differences are explained later).

- A few odd argument types remain: `T` and `F` for conditional processing, otherwise identical to `n`-type arguments, `p` for the parameter text in definitions, `w` for arguments with a specific syntax, and `D` to denote primitives that should not be used directly.
This function is used to define argument-specifier variants of the \langle parent control sequence \rangle for LATEX3 code-level macros. The \langle parent control sequence \rangle is first separated into the \langle base name \rangle and \langle original argument specifier \rangle. The comma-separated list of \langle variant argument specifiers \rangle is then used to define variants of the \langle original argument specifier \rangle if these are not already defined. For each \langle variant \rangle given, a function is created that expands its arguments as detailed and passes them to the \langle parent control sequence \rangle. So for example

\begin{verbatim}
\cs_set:Npn \foo:Nn #1#2 { code here }
\cs_generate_variant:Nn \foo:Nn { c }
\end{verbatim}

creates a new function \foo:cn which expands its first argument into a control sequence name and passes the result to \foo:Nn. Similarly

\begin{verbatim}
\cs_generate_variant:Nn \foo:Nn { NV , cV }
\end{verbatim}

generates the functions \foo:NV and \foo:cV in the same way. The \cs_generate_variant:Nn function can only be applied if the \langle parent control sequence \rangle is already defined. If the \langle parent control sequence \rangle is protected or if the \langle variant \rangle involves any x argument, then the \langle variant control sequence \rangle is also protected. The \langle variant \rangle is created globally, as is any \exp_args:N(\variant) function needed to carry out the expansion.

Only n and N arguments can be changed to other types. The only allowed changes are

- c variant of an N parent;
- o, V, v, f, e, or x variant of an n parent;
- N, n, T, F, or p argument unchanged.

This means the \langle parent \rangle of a \langle variant \rangle form is always unambiguous, even in cases where both an n-type parent and an N-type parent exist, such as for \tl_count:n and \tl_count:N.

For backward compatibility it is currently possible to make n, o, V, v, f, e, or x-type variants of an N-type argument or N or c-type variants of an n-type argument. Both are deprecated. The first because passing more than one token to an N-type argument will typically break the parent function’s code. The second because programmers who use that most often want to access the value of a variable given its name, hence should use a V-type or v-type variant instead of c-type. In those cases, using the lower-level \exp_args:N or \exp_args:Nc functions explicitly is preferred to defining confusing variants.

5.3 Introducing the variants

The V type returns the value of a register, which can be one of tl, clist, int, skip, dim, muskip, or built-in \TeX registers. The v type is the same except it first creates a control sequence out of its argument before returning the value.

In general, the programmer should not need to be concerned with expansion control. When simply using the content of a variable, functions with a V specifier should be used. For those referred to by (cs)name, the v specifier is available for the same purpose. Only
when specific expansion steps are needed, such as when using delimited arguments, should
the lower-level functions with \texttt{\texttt{o}} specifiers be employed.

The \texttt{\texttt{e}} type expands all tokens fully, starting from the first. More precisely the
expansion is identical to that of \LaTeX{}'s \texttt{\texttt{message}} (in particular \texttt{\texttt{#}} needs not be doubled).
It was added in May 2018. In recent enough engines (starting around 2019) it relies
on the primitive \texttt{\texttt{expanded}} hence is fast. In older engines it is very much slower. As
a result it should only be used in performance critical code if typical users will have a
recent installation of the \LaTeX{} ecosystem.

The \texttt{x} type expands all tokens fully, starting from the first. In contrast to \texttt{\texttt{e}}, all macro
parameter characters \texttt{#} must be doubled, and omitting this leads to low-level errors. In
addition this type of expansion is not expandable, namely functions that have \texttt{x} in their
signature do not themselves expand when appearing inside \texttt{x} or \texttt{e} expansion.

The \texttt{f} type is so special that it deserves an example. It is typically used in contexts
where only expandable commands are allowed. Then \texttt{x}-expansion cannot be used, and \texttt{f}-
expansion provides an alternative that expands the front of the token list as much as can
be done in such contexts. For instance, say that we want to evaluate the integer expression
\texttt{3 + 4} and pass the result \texttt{7} as an argument to an expandable function \texttt{\texttt{example:n}}. For
this, one should define a variant using \texttt{\texttt{cs\_generate\_variant:Nn \texttt{\texttt{example:n \{ f \},}}}}
then do

\begin{verbatim}
\example:f { \int\_eval:n { 3 + 4 } }
\end{verbatim}

Note that \texttt{x}-expansion would also expand \texttt{\int\_eval:n} fully to its result \texttt{7}, but the
variant \texttt{\texttt{\texttt{example:x}}} cannot be expandable. Note also that \texttt{\texttt{o}}-expansion would not expand
\texttt{\int\_eval:n} fully to its result since that function requires several expansions. Besides
the fact that \texttt{x}-expansion is protected rather than expandable, another difference between
\texttt{f}-expansion and \texttt{x}-expansion is that \texttt{f}-expansion expands tokens from the beginning and
stops as soon as a non-expandable token is encountered, while \texttt{x}-expansion continues
expanding further tokens. Thus, for instance

\begin{verbatim}
\example:f { \int\_eval:n { 1 + 2 } , \int\_eval:n { 3 + 4 } }
\end{verbatim}

results in the call

\begin{verbatim}
\example:n { 3 , \int\_eval:n { 3 + 4 } }
\end{verbatim}

while using \texttt{\texttt{\texttt{x}}} or \texttt{\texttt{\texttt{e}}} instead results in

\begin{verbatim}
\example:n { 3 , 7 }
\end{verbatim}

at the cost of being protected (for \texttt{x} type) or very much slower in old engines (for \texttt{e} type).
If you use \texttt{f} type expansion in conditional processing then you should stick to using \texttt{TF}
type functions only as the expansion does not finish any \texttt{\texttt{\texttt{if}}... \texttt{\texttt{fi}}: itself!}

It is important to note that both \texttt{f}- and \texttt{o}-type expansion are concerned with the
expansion of tokens from left to right in their arguments. In particular, \texttt{o}-type expansion
applies to the first \texttt{token} in the argument it receives: it is conceptually similar to

\begin{verbatim}
\exp\_after:wN <base function> \exp\_after:wN { <argument> }
\end{verbatim}

At the same time, \texttt{f}-type expansion stops at the \texttt{first} non-expandable token. This means
for example that both

\begin{verbatim}
\tl\_set:No \l\_tmpa_tl { { \g\_tmpb_tl } }
\end{verbatim}

33
and
\tl_set:Nf \l_tmpa_tl { { \g_tmpb_tl } }
leave \g_tmpb_tl unchanged: \{ is the first token in the argument and is non-expandable.

It is usually best to keep the following in mind when using variant forms.

- Variants with x-type arguments (that are fully expanded before being passed to the n-type base function) are never expandable even when the base function is. Such variants cannot work correctly in arguments that are themselves subject to expansion. Consider using f or e expansion.

- In contrast, e expansion (full expansion, almost like x except for the treatment of #) does not prevent variants from being expandable (if the base function is). The drawback is that e expansion is very much slower in old engines (before 2019). Consider using f expansion if that type of expansion is sufficient to perform the required expansion, or x expansion if the variant will not itself need to be expandable.

- Finally f expansion only expands the front of the token list, stopping at the first non-expandable token. This may fail to fully expand the argument.

When speed is essential (for functions that do very little work and whose variants are used numerous times in a document) the following considerations apply because internal functions for argument expansion come in two flavours, some faster than others.

- Arguments that might need expansion should come first in the list of arguments.

- Arguments that should consist of single tokens N, c, V, or v should come first among these.

- Arguments that appear after the first multi-token argument n, f, e, or o require slightly slower special processing to be expanded. Therefore it is best to use the optimized functions, namely those that contain only N, c, V, and v, and, in the last position, o, f, e, with possible trailing N or n or T or F, which are not expanded. Any x-type argument causes slightly slower processing.

### 5.4 Manipulating the first argument

These functions are described in detail: expansion of multiple tokens follows the same rules but is described in a shorter fashion.

\begin{verbatim}
\exp_args:Nc \exp_args:Nc \exp_args:cc \exp_args:cc
\end{verbatim}

This function absorbs two arguments (the \textit{function} name and the \textit{tokens}). The \textit{tokens} are expanded until only characters remain, and are then turned into a control sequence. The result is inserted into the input stream after reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.

The \texttt{:cc} variant constructs the \textit{function} name in the same manner as described for the \textit{tokens}.

\textbf{\LaTeX{}hackers note:} Protected macros that appear in a c-type argument are expanded despite being protected; \texttt{\exp_not:n} also has no effect. An internal error occurs if non-characters or active characters remain after full expansion, as the conversion to a control sequence is not possible.
\exp_args:No \( \langle \) function \( \rangle \) \{ \langle \) tokens \( \rangle \} \ldots

This function absorbs two arguments (the \( \langle \) function \( \rangle \) name and the \( \langle \) tokens \( \rangle \)). The \( \langle \) tokens \( \rangle \) are expanded once, and the result is inserted in braces into the input stream after reinsertion of the \( \langle \) function \( \rangle \). Thus the \( \langle \) function \( \rangle \) may take more than one argument: all others are left unchanged.

\exp_args:NV \( \langle \) function \( \rangle \) \{ \langle \) variable \( \rangle \}

This function absorbs two arguments (the names of the \( \langle \) function \( \rangle \) and the \( \langle \) variable \( \rangle \)). The content of the \( \langle \) variable \( \rangle \) are recovered and placed inside braces into the input stream after reinsertion of the \( \langle \) function \( \rangle \). Thus the \( \langle \) function \( \rangle \) may take more than one argument: all others are left unchanged.

\exp_args:Nv \( \langle \) function \( \rangle \) \{ \langle \) tokens \( \rangle \}

This function absorbs two arguments (the \( \langle \) function \( \rangle \) name and the \( \langle \) tokens \( \rangle \)). The \( \langle \) tokens \( \rangle \) are expanded until only characters remain, and are then turned into a control sequence. This control sequence should be the name of a \( \langle \) variable \( \rangle \). The content of the \( \langle \) variable \( \rangle \) are recovered and placed inside braces into the input stream after reinsertion of the \( \langle \) function \( \rangle \). Thus the \( \langle \) function \( \rangle \) may take more than one argument: all others are left unchanged.

**\TeXhackers note:** Protected macros that appear in a \( v \)-type argument are expanded despite being protected; \exp_not:n also has no effect. An internal error occurs if non-characters or active characters remain after full expansion, as the conversion to a control sequence is not possible.

\exp_args:Ne \( \langle \) function \( \rangle \) \{ \langle \) tokens \( \rangle \}

This function absorbs two arguments (the \( \langle \) function \( \rangle \) name and the \( \langle \) tokens \( \rangle \)) and exhaustively expands the \( \langle \) tokens \( \rangle \). The result is inserted in braces into the input stream after reinsertion of the \( \langle \) function \( \rangle \). Thus the \( \langle \) function \( \rangle \) may take more than one argument: all others are left unchanged.

**\TeXhackers note:** This relies on the \expanded primitive when available (in \LaTeX{} and starting around 2019 in other engines). Otherwise it uses some fall-back code that is very much slower. As a result it should only be used in performance-critical code if typical users have a recent installation of the \TeX{} ecosystem.

\exp_args:Nf \( \langle \) function \( \rangle \) \{ \langle \) tokens \( \rangle \}

This function absorbs two arguments (the \( \langle \) function \( \rangle \) name and the \( \langle \) tokens \( \rangle \)). The \( \langle \) tokens \( \rangle \) are fully expanded until the first non-expandable token is found (if that is a space it is removed), and the result is inserted in braces into the input stream after reinsertion of the \( \langle \) function \( \rangle \). Thus the \( \langle \) function \( \rangle \) may take more than one argument: all others are left unchanged.

35
\texttt{\exp_args:Nx} \texttt{\exp_args:Nx\ (function) \{(tokens)\}}

This function absorbs two arguments (the \texttt{(function)} name and the \texttt{(tokens)}) and exhaustively expands the \texttt{(tokens)}. The result is inserted in braces into the input stream \textit{after} reinsertion of the \texttt{(function)}. Thus the \texttt{(function)} may take more than one argument: all others are left unchanged.

5.5 Manipulating two arguments

\texttt{\exp_args:Nnc} \texttt{\exp_args:Nnc\ (token\textsubscript{1})\ (token\textsubscript{2})\ \{(tokens)\}}

These optimized functions absorb three arguments and expand the second and third as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second and third arguments.

\texttt{\exp_args:Noo} \texttt{\exp_args:Noo\ (token)\ \{(tokens\textsubscript{1})\}\ \{(tokens\textsubscript{2})\}}

These functions absorb three arguments and expand the second and third as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second and third arguments. These functions need slower processing.

\texttt{\exp_args:Nxx} \texttt{\exp_args:Nxx\ (token\textsubscript{1})\ (token\textsubscript{2})\ \{(tokens)\}}

These functions absorb three arguments and expand the second and third as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second and third arguments. These functions are not expandable due to their x-type argument.
5.6 Manipulating three arguments

\exp_args:NNNo (token_1) (token_2) (token_3) \{}\{tokens\}\}

These optimized functions absorb four arguments and expand the second, third and fourth as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second argument, \textit{etc.}

\exp_args:NNcf (token_1) (token_2) \{}\{tokens\}\}

These functions absorb four arguments and expand the second, third and fourth as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second argument, \textit{etc.} These functions need slower processing.

\exp_args:NNx (token_1) (token_2) \{}\{tokens\}\}

These functions absorb four arguments and expand the second, third and fourth as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second argument, \textit{etc.}
5.7 Unbraced expansion

\exp_last_unbraced:Nno \langle \text{token} \rangle \{ \langle \text{tokens}_1 \rangle \} \{ \langle \text{tokens}_2 \rangle \}

These functions absorb the number of arguments given by their specification, carry out the expansion indicated and leave the results in the input stream, with the last argument not surrounded by the usual braces. Of these, the :Nno, :Noo, :Nfo and :NnNo variants need slower processing.

\textbf{\TeX{}hackers note:} As an optimization, the last argument is unbraced by some of those functions before expansion. This can cause problems if the argument is empty: for instance, \texttt{\exp_last_unbraced:Nf \foo_bar:w \{ \}} \texttt{\q_stop} leads to an infinite loop, as the quark is \texttt{f}-expanded.

\begin{itemize}
  \item \exp_last_unbraced:No
  \item \exp_last_unbraced:NV
  \item \exp_last_unbraced:Ne
  \item \exp_last_unbraced:Nf
  \item \exp_last_unbraced:NNo
  \item \exp_last_unbraced:NNV
  \item \exp_last_unbraced:Nf
  \item \exp_last_unbraced:Nco
  \item \exp_last_unbraced:NcV
  \item \exp_last_unbraced:Nno
  \item \exp_last_unbraced:Noo
  \item \exp_last_unbraced:Nfo
  \item \exp_last_unbraced:NNNo
  \item \exp_last_unbraced:NNNV
  \item \exp_last_unbraced:NNf
  \item \exp_last_unbraced:NnNo
  \item \exp_last_unbraced:NNNNo
  \item \exp_last_unbraced:NNNf
\end{itemize}

\textbf{Updated: 2018-05-15}

\exp_last_unbraced:Nx \langle \text{function} \rangle \{ \langle \text{tokens} \rangle \}

This function fully expands the \langle \text{tokens} \rangle and leaves the result in the input stream after reinsertion of the \langle \text{function} \rangle. This function is not expandable.

\exp_last_two_unbraced:Noo \langle \text{token} \rangle \{ \langle \text{tokens}_1 \rangle \} \{ \langle \text{tokens}_2 \rangle \}

This function absorbs three arguments and expands the second and third once. The first argument of the function is then the next item on the input stream, followed by the expansion of the second and third arguments, which are not wrapped in braces. This function needs special (slower) processing.

\exp_after:wN \langle \text{token}_1 \rangle \langle \text{token}_2 \rangle

Carries out a single expansion of \langle \text{token}_2 \rangle (which may consume arguments) prior to the expansion of \langle \text{token}_1 \rangle. If \langle \text{token}_2 \rangle has no expansion (for example, if it is a character) then it is left unchanged. It is important to notice that \langle \text{token}_1 \rangle may be any single token, including group-opening and -closing tokens \{ or \} assuming normal \TeX{} category codes. Unless specifically required this should be avoided: expansion should be carried out using an appropriate argument specifier variant or the appropriate \texttt{\exp_arg:N} function.

\textbf{\TeX{}hackers note:} This is the \TeX{} primitive \texttt{\expandafter} renamed.

5.8 Preventing expansion

Despite the fact that the following functions are all about preventing expansion, they’re designed to be used in an expandable context and hence are all marked as being ‘expand-
able’ since they themselves disappear after the expansion has completed.

\exp_not:N \exp_not:N \langle \text{token} \rangle

Prevents expansion of the \langle \text{token} \rangle in a context where it would otherwise be expanded, for example an x-type argument or the first token in an o or e or f argument.

\textbf{TEXhackers note:} This is the \TeX \texttt{noexpand} primitive. It only prevents expansion. At the beginning of an f-type argument, a space \langle \text{token} \rangle is removed even if it appears as \exp_not:N \c_space_token. In an x-expanding definition (\cs_new:Npx), a macro parameter introduces an argument even if it appears as \exp_not:N \# 1. This differs from \exp_not:n.

\exp_not:c \exp_not:c \{\langle \text{tokens} \rangle\}

Expands the \langle \text{tokens} \rangle until only characters remain, and then converts this into a control sequence. Further expansion of this control sequence is then inhibited using \exp_not:N.

\textbf{TEXhackers note:} Protected macros that appear in a c-type argument are expanded despite being protected; \exp_not:n also has no effect. An internal error occurs if non-characters or active characters remain after full expansion, as the conversion to a control sequence is not possible.

\exp_not:n \exp_not:n \{\langle \text{tokens} \rangle\}

Prevents expansion of the \langle \text{tokens} \rangle in an e or x-type argument. In all other cases the \langle \text{tokens} \rangle continue to be expanded, for example in the input stream or in other types of arguments such as c, f, v. The argument of \exp_not:n \textit{must} be surrounded by braces.

\textbf{TEXhackers note:} This is the \epsilon-\TeX \texttt{unexpanded} primitive. In an x-expanding definition (\cs_new:Npx), \exp_not:n \{\#1\} is equivalent to \#1 rather than to \#1, namely it inserts the two characters \# and 1. In an e-type argument \exp_not:n \{\#\} is equivalent to \#, namely it inserts the character \#.

\exp_not:o \exp_not:o \{\langle \text{tokens} \rangle\}

Expands the \langle \text{tokens} \rangle once, then prevents any further expansion in x-type or e-type arguments using \exp_not:n.

\exp_not:V \exp_not:V \langle \text{variable} \rangle

Recovers the content of the \langle \text{variable} \rangle, then prevents expansion of this material in x-type or e-type arguments using \exp_not:n.
\exp_not:v \{\langle tokens\rangle\}

Expands the \langle tokens\rangle until only characters remains, and then converts this into a control sequence which should be a \langle variable\rangle name. The content of the \langle variable\rangle is recovered, and further expansion in x-type or e-type arguments is prevented using \exp_not:n.

\TeXhacksnote: Protected macros that appear in a v-type argument are expanded despite being protected; \exp_not:n also has no effect. An internal error occurs if non-characters or active characters remain after full expansion, as the conversion to a control sequence is not possible.

\exp_not:e \{\langle tokens\rangle\}

Expands \langle tokens\rangle exhaustively, then protects the result of the expansion (including any tokens which were not expanded) from further expansion in e or x-type arguments using \exp_not:n. This is very rarely useful but is provided for consistency.

\exp_not:f \{\langle tokens\rangle\}

Expands \langle tokens\rangle fully until the first unexpandable token is found (if it is a space it is removed). Expansion then stops, and the result of the expansion (including any tokens which were not expanded) is protected from further expansion in x-type or e-type arguments using \exp_not:n.

\exp_stop_f: \{\langle tokens\rangle \exp_stop_f: \langle more tokens\rangle\}

This function terminates an f-type expansion. Thus if a function \foo_bar:f starts an f-type expansion and all of \langle tokens\rangle are expandable \exp_stop_f: terminates the expansion of tokens even if \langle more tokens\rangle are also expandable. The function itself is an implicit space token. Inside an x-type expansion, it retains its form, but when typeset it produces the underlying space (\␣).

5.9 Controlled expansion

The expl3 language makes all efforts to hide the complexity of \TeX expansion from the programmer by providing concepts that evaluate/expand arguments of functions prior to calling the “base” functions. Thus, instead of using many \expandafter calls and other trickery it is usually a matter of choosing the right variant of a function to achieve a desired result.

Of course, deep down \TeX is using expansion as always and there are cases where a programmer needs to control that expansion directly; typical situations are basic data manipulation tools. This section documents the functions for that level. These commands are used throughout the kernel code, but we hope that outside the kernel there will be little need to resort to them. Instead the argument manipulation methods document above should usually be sufficient.

While \exp_after:wN expands one token (out of order) it is sometimes necessary to expand several tokens in one go. The next set of commands provide this functionality. Be aware that it is absolutely required that the programmer has full control over the tokens to be expanded, i.e., it is not possible to use these functions to expand unknown input as part of \langle expandable-tokens\rangle as that will break badly if unexpandable tokens are encountered in that place!
\exp:w \langle expandable-tokens \rangle \exp_end:

Expands \langle expandable-tokens \rangle until reaching \exp_end: at which point expansion stops. The full expansion of \langle expandable tokens \rangle has to be empty. If any token in \langle expandable tokens \rangle or any token generated by expanding the tokens therein is not expandable the expansion will end prematurely and as a result \exp_end: will be misinterpreted later on.\footnote{Due to the implementation you might get the character in position 0 in the current font (typically \\
'\'' in the output without any error message!}

In typical use cases the \exp_end: is hidden somewhere in the replacement text of \langle expandable-tokens \rangle rather than being on the same expansion level than \exp:w, e.g., you may see code such as

\exp:w \@@_case:NnTF #1 {#2} { } { }

where somewhere during the expansion of \@@_case:NnTF the \exp_end: gets generated.

\TeXhackers note: The current implementation uses \romannumeral hence ignores space tokens and explicit signs + and - in the expansion of the \langle expandable tokens \rangle, but this should not be relied upon.

\exp:w \langle expandable-tokens \rangle \exp_end_continue_f:w \langle further-tokens \rangle

Expands \langle expandable-tokens \rangle until reaching \exp_end_continue_f:w at which point expansion continues as an \f-type expansion expanding \langle further-tokens \rangle until an unexpandable token is encountered (or the \f-type expansion is explicitly terminated by \exp_stop_f:). As with all \f-type expansions a space ending the expansion gets removed.

The full expansion of \langle expandable-tokens \rangle has to be empty. If any token in \langle expandable-tokens \rangle or any token generated by expanding the tokens therein is not expandable the expansion will end prematurely and as a result \exp_end_continue_f:w will be misinterpreted later on.\footnote{In this particular case you may get a character into the output as well as an error message.}

In typical use cases \langle expandable-tokens \rangle contains no tokens at all, e.g., you will see code such as

\exp_after:wN \{ \exp:w \exp_end_continue_f:w #2 \}

where the \exp_after:wN triggers an \f-expansion of the tokens in #2. For technical reasons this has to happen using two tokens (if they would be hidden inside another command \exp_after:wN would only expand the command but not trigger any additional \f-expansion).

You might wonder why there are two different approaches available, after all the effect of

\exp:w \langle expandable-tokens \rangle \exp_end:

can be alternatively achieved through an \f-type expansion by using \exp_stop_f:, i.e.

\exp:w \exp_end_continue_f:w \langle expandable-tokens \rangle \exp_stop_f:

The reason is simply that the first approach is slightly faster (one less token to parse and less expansion internally) so in places where such performance really matters and where we want to explicitly stop the expansion at a defined point the first form is preferable.
The difference to \exp_end_continue_f:w is that we first we pick up an argument which is then returned to the input stream. If \langle further-tokens \rangle starts with space tokens then these space tokens are removed while searching for the argument. If it starts with a brace group then the braces are removed. Thus such spaces or braces will not terminate the f-type expansion.

## 5.10 Internal functions

\begin{verbatim}
\cs_new:Npn \exp_args:Ncof { \::c \::o \::f \::: }
\cs_new:Npn \exp_last_unbraced:Nno { \::n \::o_unbraced \::: }
\end{verbatim}

Internal forms for the base expansion types. These names do not conform to the general \LaTeX{}3 approach as this makes them more readily visible in the log and so forth. They should not be used outside this module.

\begin{verbatim}
\cs_new:Npn \exp_last_unbraced:Nno { \::n \::o_unbraced \::: }
\end{verbatim}

Internal forms for the expansion types which leave the terminal argument unbraced. These names do not conform to the general \LaTeX{}3 approach as this makes them more readily visible in the log and so forth. They should not be used outside this module.
Chapter 6

The \texttt{l3sort} package

Sorting functions

6.1 Controlling sorting

\LaTeX{} comes with a facility to sort list variables (sequences, token lists, or comma-lists) according to some user-defined comparison. For instance,

\begin{verbatim}
\clist_set:Nn \l_foo_clist { 3 , 01 , -2 , 5 , +1 }
\clist_sort:Nn \l_foo_clist
    { \int_compare:nNnTF { #1 } > { #2 }
        { \sort_return_swapped: }
        { \sort_return_same: }
    }
\end{verbatim}

results in \texttt{\l_foo_clist} holding the values \{ -2 , 01 , +1 , 3 , 5 \} sorted in non-decreasing order.

The code defining the comparison should call \texttt{\sort_return_swapped}: if the two items given as \texttt{#1} and \texttt{#2} are not in the correct order, and otherwise it should call \texttt{\sort_return_same}: to indicate that the order of this pair of items should not be changed.

For instance, a \texttt{(comparison code)} consisting only of \texttt{\sort_return_same:} with no test yields a trivial sort: the final order is identical to the original order. Conversely, using a \texttt{(comparison code)} consisting only of \texttt{\sort_return_swapped:} reverses the list (in a fairly inefficient way).

\textbf{\LaTeX{}hackers note}: The current implementation is limited to sorting approximately 20000 items (40000 in \LaTeX{}), depending on what other packages are loaded.

Internally, the code from \texttt{l3sort} stores items in \texttt{\toks} registers allocated locally. Thus, the \texttt{(comparison code)} should not call \texttt{\newtoks} or other commands that allocate new \texttt{\toks} registers. On the other hand, altering the value of a previously allocated \texttt{\toks} register is not a problem.
\texttt{\textbackslash sort\_return\_same:} \texttt{\textbackslash seq\_sort:\textbackslash Nn (seq var)}
\texttt{\textbackslash sort\_return\_swapped:} \quad \{ \ldots \texttt{\textbackslash sort\_return\_same:} \texttt{or \textbackslash sort\_return\_swapped:} \ldots \}

Indicates whether to keep the order or swap the order of two items that are compared in the sorting code. Only one of the \texttt{\textbackslash sort\_return\_\ldots} functions should be used by the code, according to the results of some tests on the items \#1 and \#2 to be compared.
Chapter 7

The \texttt{l3tl-analysis} package:

Analyzing token lists

This module provides functions that are particularly useful in the \texttt{l3regex} module for mapping through a token list one \texttt{(token)} at a time (including begin-group/end-group tokens). For \texttt{\tl_analysis_map_inline:Nn} or \texttt{\tl_analysis_map_inline:nn}, the token list is given as an argument; the analogous function \texttt{\peek_analysis_map_inline:n} documented in \texttt{l3token} finds tokens in the input stream instead. In both cases the user provides \texttt{(inline code)} that receives three arguments for each \texttt{(token)}:

- \texttt{(tokens)}, which both \texttt{o-expand} and \texttt{x-expand} to the \texttt{(token)}. The detailed form of \texttt{(tokens)} may change in later releases.
- \texttt{(char code)}, a decimal representation of the character code of the \texttt{(token)}, \texttt{-1} if it is a control sequence.
- \texttt{(catcode)}, a capital hexadecimal digit which denotes the category code of the \texttt{(token)} (0: control sequence, 1: begin-group, 2: end-group, 3: math shift, 4: alignment tab, 6: parameter, 7: superscript, 8: subscript, A: space, B: letter, C: other, D: active). This can be converted to an integer by writing "(\texttt{catcode})".

In addition, there is a debugging function \texttt{\tl_analysis_show:n}, very similar to the \texttt{\ShowTokens} macro from the \texttt{ted} package.

\begin{verbatim}
\tl_analysis_show:N
\tl_analysis_show:n \{\texttt{(token list)}\}
\tl_analysis_log:N \{\texttt{(token list)}\}
\end{verbatim}

Displays to the terminal (or log) the detailed decomposition of the \texttt{(token list)} into tokens, showing the category code of each character token, the meaning of control sequences and active characters, and the value of registers.

\begin{verbatim}
\tl_analysis_map_inline:nn
\tl_analysis_map_inline:Nn
\end{verbatim}

\texttt{\tl_analysis_map_inline:nn} \{\texttt{(token list)}\} \{\texttt{(inline function)}\}

\texttt{\tl_analysis_map_inline:Nn} \{\texttt{(token list)}\} \{\texttt{(inline function)}\}

Applies the \texttt{(inline function)} to each individual \texttt{(token)} in the \texttt{(token list)}. The \texttt{(inline function)} receives three arguments as explained above. As all other mappings the mapping is done at the current group level, \textit{i.e.} any local assignments made by the \texttt{(inline function)} remain in effect after the loop.
Chapter 8

The l3regex package: Regular expressions in \TeX

The l3regex package provides regular expression testing, extraction of submatches, splitting, and replacement, all acting on token lists. The syntax of regular expressions is mostly a subset of the PCRE syntax (and very close to POSIX), with some additions due to the fact that \TeX manipulates tokens rather than characters. For performance reasons, only a limited set of features are implemented. Notably, back-references are not supported.

Let us give a few examples. After

\tl_set:Nn \l_my_tl { That-cat. }
\regex_replace_once:nnN { at } { is } \l_my_tl

the token list variable \l_my_tl holds the text “This cat.”, where the first occurrence of “at” was replaced by “is”. A more complicated example is a pattern to emphasize each word and add a comma after it:

\regex_replace_all:nnN { \w+ } { \c{emph}\cB\{ \0 \cE\} , } \l_my_tl

The \w sequence represents any “word” character, and + indicates that the \w sequence should be repeated as many times as possible (at least once), hence matching a word in the input token list. In the replacement text, \0 denotes the full match (here, a word). The command \emph is inserted using \c{emph}, and its argument \0 is put between braces \cB{ and \cE}.

If a regular expression is to be used several times, it can be compiled once, and stored in a regex variable using \regex_set:Nn. For example,

\regex_new:N \l_foo_regex
\regex_set:Nn \l_foo_regex { \c{begin} \cB. (\c[^BE].*) \cE. }

stores in \l_foo_regex a regular expression which matches the starting marker for an environment: \begin, followed by a begin-group token (\cB.), then any number of tokens which are neither begin-group nor end-group character tokens (\c[^BE].*), ending with an end-group token (\cE.). As explained in the next section, the parentheses “capture” the result of \c[^BE].*, giving us access to the name of the environment when doing replacements.
8.1 Syntax of regular expressions

8.1.1 Regex examples

We start with a few examples, and encourage the reader to apply \regex_show:n to these regular expressions.

- **Cat** matches the word “Cat” capitalized in this way, but also matches the beginning of the word “Cattle”: use \bCat\b to match a complete word only.

- **[abc]** matches one letter among “a”, “b”, “c”; the pattern (a|b|c) matches the same three possible letters (but see the discussion of submatches below).

- **[A-Za-z]** matches any number (due to the quantifier *) of Latin letters (not accented).

- \c{[A-Za-z]**} matches a control sequence made of Latin letters.

- \_\{\_\}*\_ matches an underscore, any number of characters other than underscore, and another underscore; it is equivalent to \_\.*\_ where . matches arbitrary characters and the lazy quantifier *? means to match as few characters as possible, thus avoiding matching underscores.

- **[\+-]\?\d+** matches an explicit integer with at most one sign.

- **[\+-]\?[a-z]*\d+[\+-]...** matches an explicit integer with any number of + and – signs, with spaces allowed except within the mantissa, and surrounded by spaces.

- **[\+-]\?[a-z]*\d*[\+-]...** matches an explicit integer or decimal number; using \[.,\] instead of \. would allow the comma as a decimal marker.

- \{[\+-]\?[a-z]*\d*[\+-]...\}** matches an explicit dimension with any unit that \TeX knows, where (?i) means to treat lowercase and uppercase letters identically.

- \{[\+-]\?[a-z]*\d*[\+-]...\}** matches an explicit floating point number or the special values nan and inf (with signs and spaces allowed).

- **[\+-]\?[a-z]*\d*[\+-]...\}** matches an explicit integer or control sequence (without checking whether it is an integer variable).

- **\G.*?\K** at the beginning of a regular expression matches and discards (due to \K) everything between the end of the previous match (\G) and what is matched by the rest of the regular expression; this is useful in \regex_replace_all:nnN when the goal is to extract matches or submatches in a finer way than with \regex_extract_all:nnN.

While it is impossible for a regular expression to match only integer expressions, **\{[\+-]\?[a-z]*\d*[\+-]...\}** matches among other things all valid integer expressions (made only with explicit integers). One should follow it with further testing.
8.1.2 Characters in regular expressions

Most characters match exactly themselves, with an arbitrary category code. Some characters are special and must be escaped with a backslash (e.g., \* matches a star character). Some escape sequences of the form backslash–letter also have a special meaning (for instance \d matches any digit). As a rule,

- every alphanumeric character (A–Z, a–z, 0–9) matches exactly itself, and should not be escaped, because \A, \B, . . . have special meanings;
- non-alphanumeric printable ascii characters can (and should) always be escaped: many of them have special meanings (e.g., use \(, \), ?, . . . , ^);
- spaces should always be escaped (even in character classes);
- any other character may be escaped or not, without any effect: both versions match exactly that character.

Note that these rules play nicely with the fact that many non-alphanumeric characters are difficult to input into TEX under normal category codes. For instance, \abc\% matches the characters \abc\% (with arbitrary category codes), but does not match the control sequence \abc followed by a percent character. Matching control sequences can be done using the \c{⟨regex⟩} syntax (see below).

Any special character which appears at a place where its special behaviour cannot apply matches itself instead (for instance, a quantifier appearing at the beginning of a string), after raising a warning.

Characters.

\x{hh. .} Character with hex code hh. .
\xhh Character with hex code hh.
\a Alarm (hex 07).
\e Escape (hex 1B).
\f Form-feed (hex 0C).
\n New line (hex 0A).
\r Carriage return (hex 0D).
\t Horizontal tab (hex 09).

8.1.3 Characters classes

Character types.

. A single period matches any token.
\d Any decimal digit.
\h Any horizontal space character, equivalent to [\ \¬I]: space and tab.
\s Any space character, equivalent to [\ ¬I¬J¬L¬M].
\v Any vertical space character, equivalent to [\^\v\^\k\^\l\^\m]. Note that \^\k is a vertical space, but not a space, for compatibility with Perl.

\w Any word character, \textit{i.e.}, alphanumerics and underscore, equivalent to the explicit class [A-Za-z0-9\_].

\d Any token not matched by \d.

\h Any token not matched by \h.

\n Any token other than the \n character (hex 0A).

\s Any token not matched by \s.

\v Any token not matched by \v.

\w Any token not matched by \w.

Of those, \, \^\d, \^\h, \^\n, \^\s, \^\v, and \^\w match arbitrary control sequences.

Character classes match exactly one token in the subject.

[... ] Positive character class. Matches any of the specified tokens.

[^...] Negative character class. Matches any token other than the specified characters.

\^x-y Within a character class, this denotes a range (can be used with escaped characters).

[:\{\textit{name}\}:] Within a character class (one more set of brackets), this denotes the POSIX character class \textit{name}, which can be alnum, alpha, ascii, blank, cntrl, digit, graph, lower, print, punct, space, upper, word, or xdigit.

[:^\{\textit{name}\}]:] Negative POSIX character class.

For instance, [a-\^p-z\^\cC.] matches any lowercase latin letter except p, as well as control sequences (see below for a description of \c).

In character classes, only [, \^, ], \ and spaces are special, and should be escaped. Other non-alphabetic characters can still be escaped without harm. Any escape sequence which matches a single character (\d, \^\d, \etc.) is supported in character classes. If the first character is ^, then the meaning of the character class is inverted; ^ appearing anywhere else in the range is not special. If the first character (possibly following a leading ^) is ] then it does not need to be escaped since ending the range there would make it empty. Ranges of characters can be expressed using ~, for instance, [\^\d 0-5] and [^6-9] are equivalent.

8.1.4 Structure: alternatives, groups, repetitions

Quantifiers (repetition).

? 0 or 1, greedy.

?? 0 or 1, lazy.

* 0 or more, greedy.

*? 0 or more, lazy.

+ 1 or more, greedy.
+? 1 or more, lazy.

{n} Exactly n.

{n,} n or more, greedy.

{n,}? n or more, lazy.

{n, m} At least n, no more than m, greedy.

{n, m}? At least n, no more than m, lazy.

For greedy quantifiers the regex code will first investigate matches that involve as many repetitions as possible, while for lazy quantifiers it investigates matches with as few repetitions as possible first.

Alternation and capturing groups.

A|B|C Either one of A, B, or C, investigating A first.

(... ) Capturing group.

(?:...) Non-capturing group.

(?1...) Non-capturing group which resets the group number for capturing groups in each alternative. The following group is numbered with the first unused group number.

Capturing groups are a means of extracting information about the match. Parenthesized groups are labelled in the order of their opening parenthesis, starting at 1. The contents of those groups corresponding to the “best” match (leftmost longest) can be extracted and stored in a sequence of token lists using for instance \regex_extract_-

once:nnNTF.

The \K escape sequence resets the beginning of the match to the current position in the token list. This only affects what is reported as the full match. For instance,

\regex_extract_all:nnN { a \K . } { a123aaxyz } \l_foo_seq
results in \l_foo_seq containing the items \{1\} and \{a\}: the true matches are \{a1\} and \{aa\}, but they are trimmed by the use of \K. The \K command does not affect capturing groups: for instance,

\regex_extract_once:nnN { (. \K c)+ \d } { acbc3 } \l_foo_seq
results in \l_foo_seq containing the items \{c3\} and \{bc\}: the true match is \{acbc3\}, with first submatch \{bc\}, but \K resets the beginning of the match to the last position where it appears.

8.1.5 Matching exact tokens

The \c escape sequence allows to test the category code of tokens, and match control sequences. Each character category is represented by a single uppercase letter:

- C for control sequences;
- B for begin-group tokens;
- E for end-group tokens;
• M for math shift;
• T for alignment tab tokens;
• P for macro parameter tokens;
• U for superscript tokens (up);
• D for subscript tokens (down);
• S for spaces;
• L for letters;
• 0 for others; and
• A for active characters.

The \c escape sequence is used as follows.

\c\{\langle regex\rangle\} A control sequence whose csname matches the \langle regex\rangle, anchored at the beginning and end, so that \c\{begin\} matches exactly \begin, and nothing else.

\cX Applies to the next object, which can be a character, escape character sequence such as \x{0A}, character class, or group, and forces this object to only match tokens with category X (any of CBEMTPUDSLOA). For instance, \cL[A-Z]\d matches uppercase letters and digits of category code letter, \cC. matches any control sequence, and \c0(abc) matches where each character has category other.\footnote{This last example also captures “abc” as a regex group; to avoid this use a non-capturing group \c0(?:abc).}

\c[XYZ] Applies to the next object, and forces it to only match tokens with category X, Y, or Z (each being any of CBEMTPUDSLOA). For instance, \c[LSO](..) matches two tokens of category letter, space, or other.

\c[^XYZ] Applies to the next object and prevents it from matching any token with category X, Y, or Z (each being any of CBEMTPUDSLOA). For instance, \c[^O]\d matches digits which have any category different from other.

The category code tests can be used inside classes; for instance, [\c0\d \c[L0][A-F]] matches what \exp_not:v \{\langle var name\rangle\} at the time the regular expression is compiled. Within a \c\{...\} control sequence matching, the \u escape sequence only expands its argument once, in effect performing \tl_to_str:v. Quantifiers are supported.

The \ur escape sequence allows to insert the contents of a token list directly into a regular expression or a replacement, avoiding the need to escape special characters. Namely, \ur{\langle var name\rangle} matches the exact contents (both character codes and category codes) of the variable \langle var name\rangle, which are obtained by applying \exp_not:v \{\langle var name\rangle\} at the time the regular expression is compiled. Within a \c\{...\} control sequence matching, the \u escape sequence only expands its argument once, in effect performing \tl_to_str:v. Quantifiers are supported.

The \ur escape sequence allows to insert the contents of a regex variable into a larger regular expression. For instance, A\ur{l_tmpa_regex}D matches the tokens A and
D separated by something that matches the regular expression \l_tmpa_regex. This behaves as if a non-capturing group were surrounding \l_tmpa_regex, and any group contained in \l_tmpa_regex is converted to a non-capturing group. Quantifiers are supported.

For instance, if \l_tmpa_regex has value B|C, then A\ur{l_tmpa_regex}D is equivalent to A(?:B|C)D (matching ABD or ACD) and not to AB|CD (matching AB or CD). To get the latter effect, it is simplest to use \TeX's expansion machinery directly: if \l_mymodule_BC_tl contains B|C then the following two lines show the same result:

\regex_show:n { A \ur{l_mymodule_BC_tl} D }
\regex_show:n { A B | C D }

8.1.6 Miscellaneous

Anchors and simple assertions.

\b Word boundary: either the previous token is matched by \w and the next by \W, or the opposite. For this purpose, the ends of the token list are considered as \W.

\B Not a word boundary: between two \w tokens or two \W tokens (including the boundary).

^ or \A Start of the subject token list.

$ or \z End of the subject token list.

\G Start of the current match. This is only different from ^ in the case of multiple matches: for instance \regex_count:nnN { \G a } { aaba } \l_tmpa_int yields 2, but replacing \G by ^ would result in \l_tmpa_int holding the value 1.

The option (?i) makes the match case insensitive (identifying A–Z with a–z; no Unicode support yet). This applies until the end of the group in which it appears, and can be reverted using (?-i). For instance, in (?i)(a(?-i)b|c)d, the letters a and d are affected by the i option. Characters within ranges and classes are affected individually: (?i)[Y-\] is equivalent to \[YZ\[\yz], and (?i)[^aeiou] matches any character which is not a vowel. Neither character properties, nor \c{...} nor \u{...} are affected by the i option.

8.2 Syntax of the replacement text

Most of the features described in regular expressions do not make sense within the replacement text. Backslash introduces various special constructions, described further below:

- \0 is the whole match;
- \1 is the submatch that was matched by the first (capturing) group (...); similarly for \2, ..., \9 and \g{<number>};
- \␣ inserts a space (spaces are ignored when not escaped);
- \a, \e, \f, \n, \r, \t, \xhh, \x{hhh} correspond to single characters as in regular expressions;
• \c{⟨cs name⟩} inserts a control sequence;
• \c{⟨category⟩⟨character⟩} (see below);
• \u{⟨tl var name⟩} inserts the contents of the ⟨tl var⟩ (see below).

Characters other than backslash and space are simply inserted in the result (but since the replacement text is first converted to a string, one should also escape characters that are special for \TeX, for instance use \#). Non-alphanumeric characters can always be safely escaped with a backslash.

For instance,
\begin{verbatim}
\tl_set:Nn \l_my_tl { Hello,\ -world! }
\regex_replace_all:nnN { ([er]?l|o) . } { (\0--\1) } \l_my_tl
\end{verbatim}
results in \l_my_tl holding H(ell--el)(o,--o) w(or--o)(1d--1)!

The submatches are numbered according to the order in which the opening parenthesis of capturing groups appear in the regular expression to match. The \(n\)-th submatch is empty if there are fewer than \(n\) capturing groups or for capturing groups that appear in alternatives that were not used for the match. In case a capturing group matches several times during a match (due to quantifiers) only the last match is used in the replacement text. Submatches always keep the same category codes as in the original token list.

By default, the category code of characters inserted by the replacement are determined by the prevailing category code regime at the time where the replacement is made, with two exceptions:
• space characters (with character code 32) inserted with \\ or \x20 or \x{20} have category code 10 regardless of the prevailing category code regime;
• if the category code would be 0 (escape), 5 (newline), 9 (ignore), 14 (comment) or 15 (invalid), it is replaced by 12 (other) instead.

The escape sequence \c allows to insert characters with arbitrary category codes, as well as control sequences.
\begin{verbatim}
\cX(...) Produces the characters “...” with category \(X\), which must be one of CBEMTPUDSLOA as in regular expressions. Parentheses are optional for a single character (which can be an escape sequence). When nested, the innermost category code applies, for instance \cL(Hello\cS\ world)! gives this text with standard category codes.
\c{⟨text⟩} Produces the control sequence with csname ⟨text⟩. The ⟨text⟩ may contain references to the submatches \0, \1, and so on, as in the example for \u below.
\end{verbatim}

The escape sequence \u{⟨var name⟩} allows to insert the contents of the variable with name ⟨var name⟩ directly into the replacement, giving an easier control of category codes. When nested in \c{...} and \u{...} constructions, the \u and \c escape sequences perform \tl_to_str:v, namely extract the value of the control sequence and turn it into a string. Matches can also be used within the arguments of \c and \u. For instance,
\begin{verbatim}
\tl_set:Nn \l_my_one_tl { first }
\tl_set:Nn \l_my_two_tl { \textit{second} }
\tl_set:Nn \l_my_tl { one , two , one , one }
\regex_replace_all:nnN { [-,]+ } { \u{\l_my_\0_tl} } \l_my_tl
\end{verbatim}

53
results in \l_my_tl holding first,\emph{second},first,first.

Regex replacement is also a convenient way to produce token lists with arbitrary category codes. For instance

\lclear:N \l_tmpa_tl
\regex_replace_all:nnN { } { \cU\% \cA\~ } \l_tmpa_tl

results in \l_tmpa_tl containing the percent character with category code 7 (superscript) and an active tilde character.

### 8.3 Pre-compiling regular expressions

If a regular expression is to be used several times, it is better to compile it once rather than doing it each time the regular expression is used. The compiled regular expression is stored in a variable. All of the \l3regex module’s functions can be given their regular expression argument either as an explicit string or as a compiled regular expression.

\begin{verbatim}
\regex_new:N \l_my_regex
\regex_set:Nn \l_my_regex { my\ (simple\ )? reg(ex|ular\ expression) }
\end{verbatim}

\begin{verbatim}
\regex_const:Nn \l_regex { \A X|Y }
\regex_log:n { \l_regex }
\end{verbatim}

indicating that the anchor \A only applies to the first branch: the second branch is not anchored to the beginning of the match.
8.4 Matching

All regular expression functions are available in both :n and :N variants. The former require a “standard” regular expression, while the later require a compiled expression as generated by \regex_set:Nn.

\regex_match:nnTF { ⟨regex⟩ } { ⟨token list⟩ } { ⟨true code⟩ } { ⟨false code⟩ }

Tests whether the ⟨regular expression⟩ matches any part of the ⟨token list⟩. For instance,

\regex_match:nnTF { b [cde]* } { abecdx } { TRUE } { FALSE }
\regex_match:nnTF { [b-dq-w] } { example } { TRUE } { FALSE }

leaves TRUE then FALSE in the input stream.

\regex_count:nnN { ⟨regex⟩ } { ⟨token list⟩ } { ⟨int var⟩ }

Sets ⟨int var⟩ within the current \TeX group level equal to the number of times ⟨regular expression⟩ appears in ⟨token list⟩. The search starts by finding the left-most longest match, respecting greedy and lazy (non-greedy) operators. Then the search starts again from the character following the last character of the previous match, until reaching the end of the token list. Infinite loops are prevented in the case where the regular expression can match an empty token list: then we count one match between each pair of characters. For instance,

\int_new:N \l_foo_int
\regex_count:nnN { (b+|c) } { abbababbb } \l_foo_int

results in \l_foo_int taking the value 5.

8.5 Submatch extraction

\regex_extract_once:nnN { ⟨regex⟩ } { ⟨token list⟩ } { ⟨seq var⟩ }
\regex_extract_once:nnNTF { ⟨regex⟩ } { ⟨token list⟩ } { ⟨seq var⟩ } { ⟨true code⟩ } { ⟨false code⟩ }

Finds the first match of the ⟨regular expression⟩ in the ⟨token list⟩. If it exists, the match is stored as the first item of the ⟨seq var⟩, and further items are the contents of capturing groups, in the order of their opening parenthesis. The ⟨seq var⟩ is assigned locally. If there is no match, the ⟨seq var⟩ is cleared. The testing versions insert the ⟨true code⟩ into the input stream if a match was found, and the ⟨false code⟩ otherwise.

For instance, assume that you type

\regex_extract_once:nnNTF { \A(La)?TeX(!*)\Z } { LaTeX!!! } \l_foo_seq
{ true } { false }

Then the regular expression (anchored at the start with \A and at the end with \Z) must match the whole token list. The first capturing group, (La)?, matches La, and the second capturing group, (!*), matches !!! . Thus, \l_foo_seq contains as a result the items {LaTeX!!!}, {La}, and {!!!}, and the true branch is left in the input stream. Note that the n-th item of \l_foo_seq , as obtained using \seq_item:Nn, correspond to the submatch numbered (n – 1) in functions such as \regex_replace_once:nnN.
\regex_extract_all:nnN \{ \regex \} \{ \token_list \} \{ \seq \} \{ \true \} \{ \false \}

Finds all matches of the \langle regular expression \rangle in the \langle token list \rangle, and stores all the sub-match information in a single sequence (concatenating the results of multiple \regex_extract_once:nnN calls). The \langle seq var \rangle is assigned locally. If there is no match, the \langle seq var \rangle is cleared. The testing versions insert the \langle true code \rangle into the input stream if a match was found, and the \langle false code \rangle otherwise. For instance, assume that you type

\regex_extract_all:nnNTF \{ \w+ \} \{ Hello,-world! \} \l_foo_seq
{ true } { false }

Then the regular expression matches twice, the resulting sequence contains the two items {Hello} and {world}, and the \true branch is left in the input stream.

\regex_split:nnN \{ \langle regular expression \rangle \} \{ \token_list \} \{ \seq \}
\regex_split:nnNTF \{ \langle regular expression \rangle \} \{ \token_list \} \{ \seq \} \{ \true \} \{ \false \}

Splits the \langle token list \rangle into a sequence of parts, delimited by matches of the \langle regular expression \rangle. If the \langle regular expression \rangle has capturing groups, then the token lists that they match are stored as items of the sequence as well. The assignment to \langle seq var \rangle is local. If no match is found the resulting \langle seq var \rangle has the \langle token list \rangle as its sole item. If the \langle regular expression \rangle matches the empty token list, then the \langle token list \rangle is split into single tokens. The testing versions insert the \langle true code \rangle into the input stream if a match was found, and the \langle false code \rangle otherwise. For example, after

\seq_new:N \l_path_seq
\regex_split:nnNTF \{ / \} \{ the/path/for/this/file.tex \} \l_path_seq
{ true } { false }

the sequence \l_path_seq contains the items \{the\}, \{path\}, \{for\}, \{this\}, and \{file.tex\}, and the \true branch is left in the input stream.

### 8.6 Replacement

\regex_replace_once:nnN \{ \langle regular expression \rangle \} \{ \langle replacement \rangle \} \{ \tl \} \{ \true \} \{ \false \}

Searches for the \langle regular expression \rangle in the contents of the \langle tl var \rangle and replaces the first match with the \langle replacement \rangle. In the \langle replacement \rangle, \0 represents the full match, \1 represent the contents of the first capturing group, \2 of the second, etc. The result is assigned locally to \langle tl var \rangle.
\regex_replace_all:nnN \regex_replace_all:nnNTF \regex_replace_all:NN \regex_replace_all:NnN
Replaces all occurrences of the \langle regular expression \rangle in the contents of the \langle tl var \rangle by the \langle replacement \rangle, where \0 represents the full match, \1 represents the contents of the first capturing group, \2 of the second, etc. Every match is treated independently, and matches cannot overlap. The result is assigned locally to \langle tl var \rangle.

8.7 Scratch regular expressions

\l_tmpa_regex \l_tmpb_regex
New: 2017-12-11
Scratch regex for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_regex \g_tmpb_regex
New: 2017-12-11
Scratch regex for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

8.8 Bugs, misfeatures, future work, and other possibilities

The following need to be done now.

- Rewrite the documentation in a more ordered way, perhaps add a BNF?
  Additional error-checking to come.
- Clean up the use of messages.
- Cleaner error reporting in the replacement phase.
- Add tracing information.
- Detect attempts to use back-references and other non-implemented syntax.
- Test for the maximum register \c{c_max_register_int}.
- Find out whether the fact that \W and friends match the end-marker leads to bugs. Possibly update \tex\_regex\_item\_reverse:n.
- The empty cs should be matched by \c{}, not by \c{csname.?endcsname\s?}.
  Code improvements to come.
- Shift arrays so that the useful information starts at position 1.
- Only build \c{...} once.
- Use arrays for the left and right state stacks when compiling a regex.
• Should \_regex_action_free_group:n only be used for greedy \{n,\} quantifier? (I think not.)

• Quantifiers for \u and assertions.

• When matching, keep track of an explicit stack of curr_state and curr_submatches.

• If possible, when a state is reused by the same thread, kill other subthreads.

• Use an array rather than \l__regex_balance_tl to build the function \__regex_replacement_balance_one_match:n.

• Reduce the number of epsilon-transitions in alternatives.

• Optimize simple strings: use less states (abcade should give two states, for abc and ade). [Does that really make sense?]

• Optimize groups with no alternative.

• Optimize states with a single \_regex_action_free:n.

• Optimize the use of \__regex_action_success: by inserting it in state 2 directly instead of having an extra transition.

• Optimize the use of \int_step... functions.

• Groups don’t capture within regexes for csnames; optimize and document.

• Better “show” for anchors, properties, and catcode tests.

• Does \K really need a new state for itself?

• When compiling, use a boolean in_cs and less magic numbers.

• Instead of checking whether the character is special or alphanumeric using its character code, check if it is special in regexes with \cs_if_exist tests.

The following features are likely to be implemented at some point in the future.

• General look-ahead/behind assertions.

• Regex matching on external files.

• Conditional subpatterns with look ahead/behind: “if what follows is [...], then [...].”

• (.*...) and (?...?) sequences to set some options.

• UTF-8 mode for pdfTeX.

• Newline conventions are not done. In particular, we should have an option for . not to match newlines. Also, \A should differ from ^, and \Z, \z and $ should differ.

• Unicode properties: \p{..} and \P{..}; \X which should match any “extended” Unicode sequence. This requires to manipulate a lot of data, probably using tree-boxes.
The following features of PCRE or Perl may or may not be implemented.

- Callout with \(?C...\) or other syntax: some internal code changes make that possible, and it can be useful for instance in the replacement code to stop a regex replacement when some marker has been found; this raises the question of a potential \texttt{\textbackslash regex\_break}: and then of playing well with \texttt{\textbackslash tl\_map\_break}: called from within the code in a regex. It also raises the question of nested calls to the regex machinery, which is a problem since \texttt{\textbackslash fontdimen} are global.

- Conditional subpatterns (other than with a look-ahead or look-behind condition): this is non-regular, isn’t it?

- Named subpatterns: T\TeX\ programmers have lived so far without any need for named macro parameters.

The following features of PCRE or Perl will definitely not be implemented.

- Back-references: non-regular feature, this requires backtracking, which is prohibitively slow.

- Recursion: this is a non-regular feature.

- Atomic grouping, possessive quantifiers: those tools, mostly meant to fix catastrophic backtracking, are unnecessary in a non-backtracking algorithm, and difficult to implement.

- Subroutine calls: this syntactic sugar is difficult to include in a non-backtracking algorithm, in particular because the corresponding group should be treated as atomic.

- Backtracking control verbs: intrinsically tied to backtracking.

- \texttt{\textbackslash ddd}, matching the character with octal code \texttt{ddd}: we already have \texttt{\textbackslash x\{...\}} and the syntax is confusingly close to what we could have used for backreferences (\texttt{\textbackslash 1, \textbackslash 2, ...}), making it harder to produce useful error message.

- \texttt{\textbackslash cx}, similar to T\TeX’s own \texttt{\textbackslash ^\textbackslash x}.

- Comments: T\TeX\ already has its own \texttt{\textbackslash ^\textbackslash x}.

- \texttt{\textbackslash Q...\E} escaping: this would require to read the argument verbatim, which is not in the scope of this module.

- \texttt{\textbackslash C} single byte in UTF-8 mode: X\texttt{e\textbackslash T\TeX} and Lua\texttt{e\textbackslash T\TeX} serve us characters directly, and splitting those into bytes is tricky, encoding dependent, and most likely not useful anyways.
Chapter 9

The \texttt{l3prg} package

Control structures

Conditional processing in \LaTeXe is defined as something that performs a series of tests, possibly involving assignments and calling other functions that do not read further ahead in the input stream. After processing the input, a \textit{state} is returned. The states returned are \langle \texttt{true} \rangle and \langle \texttt{false} \rangle.

\LaTeXe has two forms of conditional flow processing based on these states. The first form is predicate functions that turn the returned state into a boolean \langle true \rangle or \langle false \rangle. For example, the function \cs_if_free_p:N checks whether the control sequence given as its argument is free and then returns the boolean \langle true \rangle or \langle false \rangle values to be used in testing with \texttt{if predicate:w} or in functions to be described below. The second form is the kind of functions choosing a particular argument from the input stream based on the result of the testing as in \cs_if_free:NTF which also takes one argument (the N) and then executes either \texttt{true} or \texttt{false} depending on the result.

\textbf{\texttt{TpX}hackers note:} The arguments are executed after exiting the underlying \texttt{if...fi:} structure.

9.1 Defining a set of conditional functions

\begin{verbatim}
\prg_new_conditional:Npnn \prg_set_conditional:Npnn \prg_new_conditional:Nnn \prg_set_conditional:Nnn
\end{verbatim}

These functions create a family of conditionals using the same \langle \texttt{code} \rangle to perform the test created. Those conditionals are expandable if \texttt{code} is. The \texttt{new} versions check for existing definitions and perform assignments globally (\textit{cf.} \texttt{cs_new:Npn}) whereas the \texttt{set} versions do no check and perform assignments locally (\textit{cf.} \texttt{cs_set:Npn}). The conditionals created are dependent on the comma-separated list of \texttt{(conditions)}, which should be one or more of \texttt{p}, \texttt{T}, \texttt{F} and \texttt{TF}.
These functions create a family of protected conditionals using the same \{code\} to perform the test created. The \{code\} does not need to be expandable. The new version check for existing definitions and perform assignments globally (cf. \cs_new:Npn) whereas the set version do not (cf. \cs_set:Npn). The conditionals created are depended on the comma-separated list of \{conditions\}, which should be one or more of T, F and TF (not p).

The conditionals are defined by \prg_new_conditional:Npnn and friends as:

- \(\langle name\rangle\_p:\langle arg\ spec\rangle\) — a predicate function which will supply either a logical true or logical false. This function is intended for use in cases where one or more logical tests are combined to lead to a final outcome. This function cannot be defined for protected conditionals.

- \(\langle name\rangle:\langle arg\ spec\rangle\_T\) — a function with one more argument than the original \langle arg spec\rangle demands. The \{true branch\} code in this additional argument will be left on the input stream only if the test is true.

- \(\langle name\rangle:\langle arg\ spec\rangle\_F\) — a function with one more argument than the original \langle arg spec\rangle demands. The \{false branch\} code in this additional argument will be left on the input stream only if the test is false.

- \(\langle name\rangle:\langle arg\ spec\rangle\_TF\) — a function with two more argument than the original \langle arg spec\rangle demands. The \{true branch\} code in the first additional argument will be left on the input stream if the test is true, while the \{false branch\} code in the second argument will be left on the input stream if the test is false.

The \{code\} of the test may use \langle parameters\rangle as specified by the second argument to \prg_set_conditional:Npnn: this should match the \langle argument specification\rangle but this is not enforced. The Nnn versions infer the number of arguments from the argument specification given (cf. \cs_new:Nn, etc.). Within the \{code\}, the functions \prg_return_true: and \prg_return_false: are used to indicate the logical outcomes of the test.

An example can easily clarify matters here:

\begin{verbatim}
\prg_set_conditional:Npnn \foo_if_bar:NN \ #1\ #2 \{ \ p \ , \ T \ , \ TF \ } 
\{ 
  \if_meaning:w \ l_tmpa_tl \ #1
  \prg_return_true:
  \else:
  \if_meaning:w \ l_tmpa_tl \ #2
  \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \fi:
\}
\end{verbatim}
This defines the function \texttt{\textbackslash foo\_if\_bar\_p:NN}, \texttt{\textbackslash foo\_if\_bar:NNTF} and \texttt{\textbackslash foo\_if\_bar:NNT} but not \texttt{\textbackslash foo\_if\_bar:NNF} (because \texttt{F} is missing from the \texttt{\langle \text{conditions}\rangle} list). The return statements take care of resolving the remaining \texttt{\textbackslash else:} and \texttt{\textbackslash fi:} before returning the state. There must be a return statement for each branch; failing to do so will result in erroneous output if that branch is executed.

\begin{verbatim}
\prg_new_eq_conditional:NNn \langle name1 \rangle:⟨arg spec1⟩ \langle name2⟩:(arg spec2) \prg_set_eq_conditional:NNn \langle conditions⟩
\end{verbatim}

These functions copy a family of conditionals. The \texttt{new} version checks for existing definitions (cf. \texttt{\cs_new_eq:NN}) whereas the \texttt{set} version does not (cf. \texttt{\cs_set_eq:NN}). The conditionals copied are depended on the comma-separated list of \texttt{⟨conditions⟩}, which should be one or more of \texttt{p}, \texttt{T}, \texttt{F} and \texttt{TF}.

\begin{verbatim}
\prg_return_true: ⋆ \prg_return_false: ⋆
\end{verbatim}

These “return” functions define the logical state of a conditional statement. They appear within the code for a conditional function generated by \texttt{\prg_set_conditional:Npnn}, etc, to indicate when a true or false branch should be taken. While they may appear multiple times each within the code of such conditionals, the execution of the conditional must result in the expansion of one of these two functions \textit{exactly once}.

The return functions trigger what is internally an \texttt{f}-expansion process to complete the evaluation of the conditional. Therefore, after \texttt{\prg_return_true:} or \texttt{\prg_return_false:} there must be no non-expandable material in the input stream for the remainder of the expansion of the conditional code. This includes other instances of either of these functions.

\begin{verbatim}
\prg_generate_conditional_variant:Nnn \langle name \rangle:⟨arg spec⟩ \langle variant argument specifiers⟩ \langle condition specifiers⟩
\end{verbatim}

Defines argument-specifier variants of conditionals. This is equivalent to running \texttt{\cs_generate_variant:Nn \langle conditional \rangle \langle variant argument specifiers⟩} on each \texttt{⟨conditional⟩} described by the \texttt{⟨condition specifiers⟩}. These base-form \texttt{⟨conditionals⟩} are obtained from the \texttt{⟨name⟩} and \texttt{⟨arg spec⟩} as described for \texttt{\prg_new_conditional:Npnn}, and they should be defined.

### 9.2 The boolean data type

This section describes a boolean data type which is closely connected to conditional processing as sometimes you want to execute some code depending on the value of a switch (\textit{e.g.}, draft/final) and other times you perhaps want to use it as a predicate function in an \texttt{\if predicate:w} test. The problem of the primitive \texttt{\textbackslash if\_false:} and \texttt{\textbackslash if\_true:} tokens is that it is not always safe to pass them around as they may interfere with scanning for termination of primitive conditional processing. Therefore, we employ two canonical booleans: \texttt{\c_true_bool} or \texttt{\c_false_bool}. Besides preventing problems as described above, it also allows us to implement a simple boolean parser supporting the logical operations And, Or, Not, \textit{etc.} which can then be used on both the boolean type and predicate functions.
All conditional \bool functions except assignments are expandable and expect the input to also be fully expandable (which generally means being constructed from predicate functions and booleans, possibly nested).

\TeX hackers note: The bool data type is not implemented using the \iffalse/\iftrue primitives, in contrast to \newif, etc., in plain \TeX, \if\ifx2e and so on. Programmers should not base use of bool switches on any particular expectation of the implementation.

\noindent \begin{verbatim}
\bool_new:N \bool_new:c \bool_const:Nn \bool_const:cn \bool_set_false:N \bool_set_false:c \bool_gset_false:N \bool_gset_false:c \bool_set_true:N \bool_set_true:c \bool_gset_true:N \bool_gset_true:c \bool_set_eq:NN \bool_set_eq:N \bool_gset_eq:NN \bool_gset_eq:N \bool_set:Nn \bool_set:cn \bool_gset:Nn \bool_gset:cn \bool_if_p:N \bool_if_p:c \bool_if:N \bool_if:TF \bool_if:TF \bool_to_str:N \bool_to_str:c \bool_to_str:n \end{verbatim}

\noindent \verb|\bool_new:N| \verb|\bool_new:c| Creates a new \langle boolean\rangle or raises an error if the name is already taken. The declaration is global. The \langle boolean\rangle is initially false.

\noindent \verb|\bool_const:Nn| \verb|\bool_const:cn| Creates a new constant \langle boolean\rangle or raises an error if the name is already taken. The value of the \langle boolean\rangle is set globally to the result of evaluating the \langle boolexpr\rangle.

\noindent \verb|\bool_set_false:N| \verb|\bool_set_false:c| \verb|\bool_gset_false:N| \verb|\bool_gset_false:c| Sets \langle boolean\rangle logically false.

\noindent \verb|\bool_set_true:N| \verb|\bool_set_true:c| \verb|\bool_gset_true:N| \verb|\bool_gset_true:c| Sets \langle boolean\rangle logically true.

\noindent \verb|\bool_set_eq:NN| \verb|\bool_set_eq:N| \verb|\bool_gset_eq:NN| \verb|\bool_gset_eq:N| Sets \langle boolean\rangle to the current value of \langle boolean2\rangle.

\noindent \verb|\bool_set:Nn| \verb|\bool_set:cn| \verb|\bool_gset:Nn| \verb|\bool_gset:cn| Evaluates the \langle boolean expression\rangle as described for \bool_if:nTF, and sets the \langle boolean\rangle variable to the logical truth of this evaluation.

\noindent \verb|\bool_if_p:N| \verb|\bool_if_p:c| \verb|\bool_if:N| \verb|\bool_if:TF| \verb|\bool_if:TF| Tests the current truth of \langle boolean\rangle, and continues expansion based on this result.

\noindent \verb|\bool_to_str:N| \verb|\bool_to_str:c| \verb|\bool_to_str:n| Expands to the letters true or false depending on the logical truth of the \langle boolean\rangle or \langle boolean expression\rangle.

63
\bool_show:N \bool_show:N \langle boolean \rangle

Displays the logical truth of the \langle boolean \rangle on the terminal.

\bool_show:n \bool_show:n \langle boolean expression \rangle

Displays the logical truth of the \langle boolean expression \rangle on the terminal.

\bool_log:N \bool_log:N \langle boolean \rangle

Writes the logical truth of the \langle boolean \rangle in the log file.

\bool_log:n \bool_log:n \langle boolean expression \rangle

Writes the logical truth of the \langle boolean expression \rangle in the log file.

\bool_if_exist_p:N \bool_if_exist:NTF \langle boolean \rangle \{\langle true code \rangle \} \{\langle false code \rangle \}

Tests whether the \langle boolean \rangle is currently defined. This does not check that the \langle boolean \rangle really is a boolean variable.

\l_tmpa_bool \l_tmpb_bool

A scratch boolean for local assignment. It is never used by the kernel code, and so is safe for use with any \LaTeX3-defined function. However, it may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_bool \g_tmpb_bool

A scratch boolean for global assignment. It is never used by the kernel code, and so is safe for use with any \LaTeX3-defined function. However, it may be overwritten by other non-kernel code and so should only be used for short-term storage.

9.2.1 Scratch booleans

9.3 Boolean expressions

As we have a boolean datatype and predicate functions returning boolean \langle true \rangle or \langle false \rangle values, it seems only fitting that we also provide a parser for \langle boolean expressions \rangle.

A boolean expression is an expression which given input in the form of predicate functions and boolean variables, return boolean \langle true \rangle or \langle false \rangle. It supports the logical operations And, Or and Not as the well-known infix operators && and || and prefix ! with their usual precedences (namely, && binds more tightly than ||). In addition to this, parentheses can be used to isolate sub-expressions. For example,
\int_compare_p:n \{ 1 = 1 \} &&
  \int_compare_p:n \{ 2 = 3 \} ||
\int_compare_p:n \{ 4 <= 4 \} ||
\str_if_eq_p:nn \{ abc \} \{ def \} &&
! \int_compare_p:n \{ 2 = 4 \}

is a valid boolean expression.

Contrary to some other programming languages, the operators \&\& and \|\| evaluate both operands in all cases, even when the first operand is enough to determine the result. This “eager” evaluation should be contrasted with the “lazy” evaluation of \bool_lazy_- functions.

\TeXhackers note: The eager evaluation of boolean expressions is unfortunately necessary in \TeX. Indeed, a lazy parser can get confused if \&\& or \|\| or parentheses appear as (unbraced) arguments of some predicates. For instance, the innocuous-looking expression below would break (in a lazy parser) if \#1 were a closing parenthesis and \l_tmpa_bool were true.

( \l_tmpa_bool || \token_if_eq_meaning_p:NN X \#1 )

Minimal (lazy) evaluation can be obtained using the conditionals \bool_lazy_all:nTF, \bool_lazy_and:nnTF, \bool_lazy_any:nTF, or \bool_lazy_or:nnTF, which only evaluate their boolean expression arguments when they are needed to determine the resulting truth value. For example, when evaluating the boolean expression

\bool_lazy_and_p:nn
\{ \bool_lazy_any_p:n
  \{ \int_compare_p:n \{ 2 = 3 \} \}
  \{ \int_compare_p:n \{ 4 <= 4 \} \}
  \{ \int_compare_p:n \{ 1 = \error \} \} \% skipped
\}
\{ ! \int_compare_p:n \{ 2 = 4 \} \}

the line marked with \texttt{skipped} is not expanded because the result of \bool_lazy_any_p:n is known once the second boolean expression is found to be logically \texttt{true}. On the other hand, the last line is expanded because its logical value is needed to determine the result of \bool_lazy_and_p:nn.

\bool_if_p:n \{ \langle boolean expression \rangle \}
\bool_if:nTF \{ \langle boolean expression \rangle \} \{ \langle true code \rangle \} \{ \langle false code \rangle \}

Tests the current truth of \langle boolean expression \rangle, and continues expansion based on this result. The \langle boolean expression \rangle should consist of a series of predicates or boolean variables with the logical relationship between these defined using \&\& (“And”), \|\| (“Or”), \! (“Not”) and parentheses. The logical Not applies to the next predicate or group.
\bool_lazy_all_p:n * \bool_lazy_all:nTF *
\bool_lazy_all_p:n { {\expr} \expr \cdots \expr }
\bool_lazy_all:nTF { {\expr} \expr \cdots \expr } \true \false

Implements the “And” operation on the \textit{boolean expressions}, hence is \texttt{true} if all of
them are \texttt{true} and \texttt{false} if any of them is \texttt{false}. Contrarily to the infix operator \&\&, only the \textit{boolean expressions} which are needed to determine the result of \bool_lazy_all:nTF are evaluated. See also \bool_lazy_and:nTF when there are only two \textit{boolean expressions}.

\bool_lazy_and_p:nn * \bool_lazy_and:nTF *
\bool_lazy_and_p:nn { {\expr} \expr }
\bool_lazy_and:nTF { {\expr} \expr } \true \false

Implements the “And” operation between two \textit{boolean expressions}, hence is \texttt{true} if both are \texttt{true}. Contrarily to the infix operator \&\&, the \textit{boolean expressions} \bool_lazy_and:nTF is only evaluated if it is needed to determine the result of \bool_lazy_and:nTF. See also \bool_lazy_all:nTF when there are more than two \textit{boolean expressions}.

\bool_lazy_any_p:n * \bool_lazy_any:nTF *
\bool_lazy_any_p:n { {\expr} \expr \cdots \expr }
\bool_lazy_any:nTF { {\expr} \expr \cdots \expr } \true \false

Implements the “Or” operation on the \textit{boolean expressions}, hence is \texttt{true} if any of them is \texttt{true} and \texttt{false} if all of them are \texttt{false}. Contrarily to the infix operator ||, only the \textit{boolean expressions} which are needed to determine the result of \bool_lazy_any:nTF are evaluated. See also \bool_lazy_or:nTF when there are only two \textit{boolean expressions}.

\bool_lazy_or_p:nn * \bool_lazy_or:nTF *
\bool_lazy_or_p:nn { {\expr} \expr }
\bool_lazy_or:nTF { {\expr} \expr } \true \false

Implements the “Or” operation between two \textit{boolean expressions}, hence is \texttt{true} if either one is \texttt{true}. Contrarily to the infix operator ||, the \textit{boolean expressions} \bool_lazy_or:nTF is only evaluated if it is needed to determine the result of \bool_lazy_or:nTF. See also \bool_lazy_any:nTF when there are more than two \textit{boolean expressions}.

\bool_not_p:n \true \false
\bool_not_p:n { boolean expression}

Function version of !{boolean expression} within a boolean expression.

\bool_xor_p:nn * \bool_xor:nTF *
\bool_xor_p:nn { {\expr} \expr }
\bool_xor:nTF { {\expr} \expr } \true \false

Implements an “exclusive or” operation between two \textit{boolean expressions}. There is no
infix operation for this logical operation.

\section{Logical loops}

Loops using either boolean expressions or stored boolean values.

\bool_do_until:Nn \true \false
\bool_do_until:Nn \bool \{ \code \}

Places the \textit{code} in the input stream for \TeX{} to process, and then checks the logical value of the \textit{boolean}. If it is \texttt{false} then the \texttt{code} is inserted into the input stream again and the process loops until the \texttt{false} is \texttt{true}.

66
\bool_do_while:Nn \\bool_do_while:cn
\Updated: 2017-07-15

\bool_until_do:Nn \\bool_until_do:cn
\Updated: 2017-07-15

\bool_while_do:Nn \\bool_while_do:cn
\Updated: 2017-07-15

\bool_do_until:nn \\bool_do_until:cn
\Updated: 2017-07-15

\bool_do_while:Nn \langle boolean \rangle \{ \langle code \rangle \}
Places the \langle code \rangle in the input stream for \TeX{} to process, and then checks the logical value of the \langle boolean \rangle. If it is \texttt{true} then the \langle code \rangle is inserted into the input stream again and the process loops until the \langle boolean \rangle is \texttt{false}.

\bool_until_do:Nn \langle boolean \rangle \{ \langle code \rangle \}
This function firsts checks the logical value of the \langle boolean \rangle. If it is \texttt{false} the \langle code \rangle is placed in the input stream and expanded. After the completion of the \langle code \rangle the truth of the \langle boolean \rangle is re-evaluated. The process then loops until the \langle boolean \rangle is \texttt{true}.

\bool_while_do:Nn \langle boolean \rangle \{ \langle code \rangle \}
This function firsts checks the logical value of the \langle boolean \rangle. If it is \texttt{true} the \langle code \rangle is placed in the input stream and expanded. After the completion of the \langle code \rangle the truth of the \langle boolean \rangle is re-evaluated. The process then loops until the \langle boolean \rangle is \texttt{false}.

\bool_do_until:nn \langle boolean expression \rangle \{ \langle code \rangle \}
Places the \langle code \rangle in the input stream for \TeX{} to process, and then checks the logical value of the \langle boolean expression \rangle as described for \bool_if:nTF. If it is \texttt{false} then the \langle code \rangle is inserted into the input stream again and the process loops until the \langle boolean expression \rangle evaluates to \texttt{true}.

\bool_while_do:nn \langle boolean expression \rangle \{ \langle code \rangle \}
Places the \langle code \rangle in the input stream for \TeX{} to process, and then checks the logical value of the \langle boolean expression \rangle as described for \bool_if:nTF. If it is \texttt{true} then the \langle code \rangle is inserted into the input stream again and the process loops until the \langle boolean expression \rangle evaluates to \texttt{false}.

\bool_until_do:nn \langle boolean expression \rangle \{ \langle code \rangle \}
This function firsts checks the logical value of the \langle boolean expression \rangle (as described for \bool_if:nTF). If it is \texttt{false} the \langle code \rangle is placed in the input stream and expanded. After the completion of the \langle code \rangle the truth of the \langle boolean expression \rangle is re-evaluated. The process then loops until the \langle boolean expression \rangle is \texttt{true}.

\bool_while_do:nn \langle boolean expression \rangle \{ \langle code \rangle \}
This function firsts checks the logical value of the \langle boolean expression \rangle (as described for \bool_if:nTF). If it is \texttt{true} the \langle code \rangle is placed in the input stream and expanded. After the completion of the \langle code \rangle the truth of the \langle boolean expression \rangle is re-evaluated. The process then loops until the \langle boolean expression \rangle is \texttt{false}.

\9.5 Producing multiple copies
\prg_replicate:nn \langle integer expression \rangle \{ \langle tokens \rangle \}
Evaluates the \langle integer expression \rangle (which should be zero or positive) and creates the resulting number of copies of the \langle tokens \rangle. The function is both expandable and safe for nesting. It yields its result after two expansion steps.

67
9.6 Detecting \TeX’s mode

\mode_if_horizontal_p: * \mode_if_horizontal:p: \mode_if_horizontal:T\ false code}
Detects if \TeX{} is currently in horizontal mode.

\mode_if_inner_p: * \mode_if_inner:p: \mode_if_inner:T\ false code}
Detects if \TeX{} is currently in inner mode.

\mode_if_math_p: * \mode_if_math:p: \mode_if_math:T\ false code}
Detects if \TeX{} is currently in maths mode.

\mode_if_vertical_p: * \mode_if_vertical:p: \mode_if_vertical:T\ false code}
Detects if \TeX{} is currently in vertical mode.

9.7 Primitive conditionals

\if_predicate:w \if Predicate\ (true code) \else: (false code) \fi:
This function takes a predicate function and branches according to the result. (In practice this function would also accept a single boolean variable in place of the \texttt{\langle\ predicate\ \rangle} but to make the coding clearer this should be done through \texttt{\if_bool:N}.)

\if_bool:N \if Boolean\ (true code) \else: (false code) \fi:
This function takes a boolean variable and branches according to the result.

9.8 Nestable recursions and mappings

There are a number of places where recursion or mapping constructs are used in expl3. At a low-level, these typically require insertion of tokens at the end of the content to allow “clean up”. To support such mappings in a nestable form, the following functions are provided.

\prg_break_point:Nn \prg_break_point:Nn \langle\ type\ \rangle\_map_break: \langle\ code\rangle
Used to mark the end of a recursion or mapping: the functions \texttt{\langle\ type\ \rangle\_map_break:} and \texttt{\langle\ type\ \rangle\_map_break:n} use this to break out of the loop (see \texttt{\prg_map_break:Nn} for how to set these up). After the loop ends, the \texttt{\langle\ code\rangle} is inserted into the input stream. This occurs even if the break functions are \textit{not} applied: \texttt{\prg_break_point:Nn} is functionally-equivalent in these cases to \texttt{\use_ii:nn}.
Breaks a recursion in mapping contexts, inserting in the input stream the \(\langle\text{user code}\rangle\) after the \(\langle\text{ending code}\rangle\) for the loop. The function breaks loops, inserting their \(\langle\text{ending code}\rangle\), until reaching a loop with the same \(\langle\text{type}\rangle\) as its first argument. This \(\langle\text{type}\rangle\)-\map_break: argument must be defined; it is simply used as a recognizable marker for the \(\langle\text{type}\rangle\).

For types with mappings defined in the kernel, \(\langle\text{type}\rangle\)-\map_break: and \(\langle\text{type}\rangle\)-\map_break:n are defined as \map_break:Nn \(\langle\text{type}\rangle\)-\map_break: {} and the same with {} omitted.

9.8.1 Simple mappings

In addition to the more complex mappings above, non-nestable mappings are used in a number of locations and support is provided for these.

This copy of \do_nothing: is used to mark the end of a fast short-term recursion: the function \break:n uses this to break out of the loop.

Breaks a recursion which has no \(\langle\text{ending code}\rangle\) and which is not a user-breakable mapping (see for instance \prop_get:Nn), and inserts the \(\langle\text{code}\rangle\) in the input stream.

9.9 Internal programming functions

These functions are used to enclose material in a \TeX alignment environment within a specially-constructed group. This group is designed in such a way that it does not add brace groups to the output but does act as a group for the & token inside \halign. This is necessary to allow grabbing of tokens for testing purposes, as \TeX uses group level to determine the effect of alignment tokens. Without the special grouping, the use of a function such as \peek_after:Nw would result in a forbidden comparison of the internal \endtemplate token, yielding a fatal error. Each \group_align_safe_begin: must be matched by a \group_align_safe_end:, although this does not have to occur within the same function.
Chapter 10

The l3sys package:
System/runtime functions

10.1 The name of the job

\c_sys_jobname_str
New: 2015-09-19
Updated: 2019-10-27

Constant that gets the “job name” assigned when \TeX starts.

\TeXhackers note: This copies the contents of the primitive \jobname. For technical reasons, the string here is not of the same internal form as other, but may be manipulated using normal string functions.

10.2 Date and time

\c_sys_minute_int
\c_sys_hour_int
\c_sys_day_int
\c_sys_month_int
\c_sys_year_int
New: 2015-09-22

The date and time at which the current job was started: these are all reported as integers.

\TeXhackers note: Whilst the underlying primitives can be altered by the user, this interface to the time and date is intended to be the “real” values.
10.3 Engine

\sys_if_engine_luatex_p: *
\sys_if_engine_luatex:* TF
\sys_if_engine_pdftex_p: *
\sys_if_engine_pdftex:* TF
\sys_if_engine_ptex: *
\sys_if_engine_ptex:* TF
\sys_if_engine_uptex_p: *
\sys_if_engine_uptex:* TF
\sys_if_engine_xetex_p: *
\sys_if_engine_xetex:* TF

New: 2015-09-07

The current engine given as a lower case string: one of lualatex, pdftex, ptex, uptex or xetex.

\c_sys_engine_str
New: 2015-09-19

The name of the standard executable for the current TEX engine given as a lower case string: one of lualatex, luahbtex, pdftex, eptex, euptex or xetex.

\c_sys_engine_exec_str
New: 2020-08-20

The name of the preloaded format for the current TEX run given as a lower case string: one of lualatex, pdflatex, latex, platex, uplatex or xelatex for \LaTeX, similar names for plain TEX (except pdflatex in DVI mode yields etex), and cont-en for ConTEXt (i.e. the \fmtname).

\c_sys_engine_format_str
New: 2020-08-20

\sys_timer: *
New: 2020-09-24

Expands to the current value of the engine’s timer clock, a non-negative integer. This function is only defined for engines with timer support. This command measures not just CPU time but real time (including time waiting for user input). The unit are scaled seconds \((2^{-16} \text{ seconds})\).

10.4 Output format

\sys_if_output_dvi_p: *
\sys_if_output_dvi:* TF
\sys_if_output_pdf_p: *
\sys_if_output_pdf:* TF

New: 2015-09-19

Conditionals which give the current output mode the TEX run is operating in. This is always one of two outcomes, DVI mode or PDF mode. The two sets of conditionals are thus complementary and are both provided to allow the programmer to emphasise the most appropriate case.

\c_sys_output_str
New: 2015-09-19

The current output mode given as a lower case string: one of dvi or pdf.
### 10.5 Platform

<table>
<thead>
<tr>
<th>Conditional</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>\sys_if_platform_unix_p: *</td>
<td>Determines if the current platform is Unix-like.</td>
</tr>
<tr>
<td>\sys_if_platform_unix: TF</td>
<td>(true) if Unix, (false) otherwise.</td>
</tr>
<tr>
<td>\sys_if_platform_windows_p: *</td>
<td>Determines if the current platform is Windows.</td>
</tr>
<tr>
<td>\sys_if_platform_windows: TF</td>
<td>(true) if Windows, (false) otherwise.</td>
</tr>
</tbody>
</table>

Conditionals which allow platform-specific code to be used. The names follow the Lua `os.type()` function, i.e. all Unix-like systems are `unix` (including Linux and MacOS).

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\c_sys_platform_str</td>
<td>Returns the current platform as a lower case string: <code>unix</code>, <code>windows</code>, or <code>unknown</code>.</td>
</tr>
</tbody>
</table>

### 10.6 Random numbers

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\sys_rand_seed</td>
<td>Expands to the current value of the engine’s random seed, a non-negative integer. In engines without random number support this expands to 0.</td>
</tr>
<tr>
<td>\sys_gset_rand_seed:n</td>
<td>Globally sets the seed for the engine’s pseudo-random number generator to the (\langle\ integer \ expression \rangle). This random seed affects all ..._rand functions (such as \int_rand:nn or \clist_rand_item:n) as well as other packages relying on the engine’s random number generator. In engines without random number support this produces an error. <strong>TeXhackers note:</strong> While a 32-bit (signed) integer can be given as a seed, only the absolute value is used and any number beyond (2^{28}) is divided by an appropriate power of 2. We recommend using an integer in ([0, 2^{28} - 1]).</td>
</tr>
</tbody>
</table>

### 10.7 Access to the shell

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\sys_get_shell:nnN</td>
<td>Defines (tl\ var) to the text returned by the (shell\ command). The (shell\ command) is converted to a string using (\tl_to_str:n). Category codes may need to be set appropriately via the (setup) argument, which is run just before running the (shell\ command) (in a group). If shell escape is disabled, the (tl\ var) will be set to \q_no_value in the non-branching version. Note that quote characters (*) cannot be used inside the (shell\ command). The \sys_get_shell:nnN conditional returns true if the shell is available and no quote is detected, and false otherwise.</td>
</tr>
</tbody>
</table>

New: 2019-09-20
This variable exposes the internal triple of the shell escape status. The possible values are:

- 0 Shell escape is disabled
- 1 Unrestricted shell escape is enabled
- 2 Restricted shell escape is enabled

\sys_if_shell_p: \sys_if_shell:TF {\langle true code \rangle} {\langle false code \rangle}

Performs a check for whether shell escape is enabled. This returns true if either of restricted or unrestricted shell escape is enabled.

\sys_if_shell_unrestricted_p: \sys_if_shell_unrestricted:TF {\langle true code \rangle} {\langle false code \rangle}

Performs a check for whether unrestricted shell escape is enabled.

\sys_if_shell_restricted_p: \sys_if_shell_restricted:TF {\langle true code \rangle} {\langle false code \rangle}

Performs a check for whether restricted shell escape is enabled. This returns false if unrestricted shell escape is enabled. Unrestricted shell escape is not considered a superset of restricted shell escape in this case. To find whether any shell escape is enabled use \sys_if_shell:.

\sys_shell_now:n \sys_shell_now:x

Execute \langle tokens \rangle through shell escape immediately.

\sys_shell_shipout:n \sys_shell_shipout:x

Execute \langle tokens \rangle through shell escape at shipout.

### 10.8 Loading configuration data

\sys_load_backend:n \sys_load_backend:n \langle backend \rangle

Loads the additional configuration file needed for backend support. If the \langle backend \rangle is empty, the standard backend for the engine in use will be loaded. This command may only be used once.

\c_sys_backend_str

Set to the name of the backend in use by \sys_load_backend:n when issued.
Load the additional configuration files for debugging support and rolling back deprecations, respectively.

10.8.1 Final settings

Finalises all system-dependent functionality: required before loading a backend.
Chapter 11

The l3msg package

Messages need to be passed to the user by modules, either when errors occur or to indicate how the code is proceeding. The l3msg module provides a consistent method for doing this (as opposed to writing directly to the terminal or log).

The system used by l3msg to create messages divides the process into two distinct parts. Named messages are created in the first part of the process; at this stage, no decision is made about the type of output that the message will produce. The second part of the process is actually producing a message. At this stage a choice of message class has to be made, for example error, warning or info.

By separating out the creation and use of messages, several benefits are available. First, the messages can be altered later without needing details of where they are used in the code. This makes it possible to alter the language used, the detail level and so on. Secondly, the output which results from a given message can be altered. This can be done on a message class, module or message name basis. In this way, message behaviour can be altered and messages can be entirely suppressed.

11.1 Creating new messages

All messages have to be created before they can be used. The text of messages is automatically wrapped to the length available in the console. As a result, formatting is only needed where it helps to show meaning. In particular, \ may be used to force a new line and \ forces an explicit space. Additionally, \#, \%, \ and \~ can be used to produce the corresponding character.

Messages may be subdivided by one level using the / character. This is used within the message filtering system to allow for example the LaTeX kernel messages to belong to the module LaTeX while still being filterable at a more granular level. Thus for example

\msg_new:nnnn { mymodule } { submodule / message } ... will allow to filter out specifically messages from the submodule.
\msg_new:nnn \msg_new:nnn

\msg_new:nnn \msg_new:nnn

\msg_set:nnn \msg_set:nnn \msg_gset:nnn \msg_gset:nnn

\msg_if_exist_p:nn \msg_if_exist:nnTF

\msg_module_name:n \msg_module_type:n

\g_msg_module_name_prop \g_msg_module_type_prop

\msg_line_context:

11.2 Customizable information for message modules

\msg_module_name:n \msg_module_type:n

Expands to the public name of the \langle module\rangle as defined by \g_msg_module_name_prop (or otherwise leaves the \langle module\rangle unchanged).

Expands to the description which applies to the \langle module\rangle, for example a \texttt{Package} or \texttt{Class}. The information here is defined in \g_msg_module_type_prop, and will default to \texttt{Package} if an entry is not present.

Provides a mapping between the module name used for messages, and that for documentation. For example, \LaTeX{}3 core messages are stored in the reserved \LaTeX{} tree, but are printed as \LaTeX{}3.

Provides a mapping between the module name used for messages, and that type of module. For example, for \LaTeX{}3 core messages, an empty entry is set here meaning that they are not described using the standard \texttt{Package} text.

11.3 Contextual information for messages

\msg_line_context:

Prints the current line number when a message is given, and thus suitable for giving context to messages. The number itself is proceeded by the text \texttt{on line}.
\msg_line_number: \* \msg_line_number:
Prints the current line number when a message is given.

\msg_fatal_text:n \* \msg_fatal_text:n \{\langle\text{module}\rangle\}
Produces the standard text
Fatal Package \langle\text{module}\rangle\ Error
This function can be redefined to alter the language in which the message is given, using #1 as the name of the \langle\text{module}\rangle\ to be included.

\msg_critical_text:n \* \msg_critical_text:n \{\langle\text{module}\rangle\}
Produces the standard text
Critical Package \langle\text{module}\rangle\ Error
This function can be redefined to alter the language in which the message is given, using #1 as the name of the \langle\text{module}\rangle\ to be included.

\msg_error_text:n \* \msg_error_text:n \{\langle\text{module}\rangle\}
Produces the standard text
Package \langle\text{module}\rangle\ Error
This function can be redefined to alter the language in which the message is given, using #1 as the name of the \langle\text{module}\rangle\ to be included.

\msg_warning_text:n \* \msg_warning_text:n \{\langle\text{module}\rangle\}
Produces the standard text
Package \langle\text{module}\rangle\ Warning
This function can be redefined to alter the language in which the message is given, using #1 as the name of the \langle\text{module}\rangle\ to be included. The \langle\text{type}\rangle\ of \langle\text{module}\rangle\ may be adjusted: Package is the standard outcome: see \msg_module_type:n.

\msg_info_text:n \* \msg_info_text:n \{\langle\text{module}\rangle\}
Produces the standard text:
Package \langle\text{module}\rangle\ Info
This function can be redefined to alter the language in which the message is given, using #1 as the name of the \langle\text{module}\rangle\ to be included. The \langle\text{type}\rangle\ of \langle\text{module}\rangle\ may be adjusted: Package is the standard outcome: see \msg_module_type:n.
Produces the standard text

See the ⟨module⟩ documentation for further information.

This function can be redefined to alter the language in which the message is given, using #1 as the name of the ⟨module⟩ to be included. The name of the ⟨module⟩ is produced using \msg_module_name:n.

11.4 Issuing messages

Messages behave differently depending on the message class. In all cases, the message may be issued supplying 0 to 4 arguments. If the number of arguments supplied here does not match the number in the definition of the message, extra arguments are ignored, or empty arguments added (of course the sense of the message may be impaired). The four arguments are converted to strings before being added to the message text: the x-type variants should be used to expand material. Note that this expansion takes place with the standard definitions in effect, which means that shorthands such as \- or \\ are not available; instead one should use \iow_char:N \- and \iow_newline:, respectively. The following message classes exist:

- **fatal**, ending the TEX run;
- **critical**, ending the file being input;
- **error**, interrupting the TEX run without ending it;
- **warning**, written to terminal and log file, for important messages that may require corrections by the user;
- **note** (less common than info) for important information messages written to the terminal and log file;
- **info** for normal information messages written to the log file only;
- **term** and **log** for un-decorated messages written to the terminal and log file, or to the log file only;
- **none** for suppressed messages.

\msg_fatal:nnnnnn ⟨(module)⟩ ⟨(message)⟩ ⟨(arg one)⟩ ⟨(arg two)⟩ ⟨(arg three)⟩ ⟨(arg four)⟩

Issues ⟨module⟩ error ⟨message⟩, passing ⟨arg one⟩ to ⟨arg four⟩ to the text-creating functions. After issuing a fatal error the TEX run halts. No PDF file will be produced in this case (DVI mode runs may produce a truncated DVI file).


\textbf{Issues} \texttt{\langle module\rangle} error \texttt{\langle message\rangle}, passing \texttt{\langle arg\ one\rangle} to \texttt{\langle arg\ four\rangle} to the text-creating functions. After issuing a critical error, \TeX\ stops reading the current input file. This may halt the \TeX\ run (if the current file is the main file) or may abort reading a sub-file.

\textbf{\TeX\ hackers note:} The \TeX\ \texttt{\textbackslash endinput} primitive is used to exit the file. In particular, the rest of the current line remains in the input stream.

\textbf{\TeX\ hackers note:} The \TeX\ \texttt{\textbackslash endinput} primitive is used to exit the file. In particular, the rest of the current line remains in the input stream.

\textbf{Issues} \texttt{\langle module\rangle} error \texttt{\langle message\rangle}, passing \texttt{\langle arg\ one\rangle} to \texttt{\langle arg\ four\rangle} to the text-creating functions. The error interrupts processing and issues the text at the terminal. After user input, the run continues.

\textbf{\TeX\ hackers note:} The \TeX\ \texttt{\textbackslash endinput} primitive is used to exit the file. In particular, the rest of the current line remains in the input stream.

\textbf{Issues} \texttt{\langle module\rangle} warning \texttt{\langle message\rangle}, passing \texttt{\langle arg\ one\rangle} to \texttt{\langle arg\ four\rangle} to the text-creating functions. The warning text is added to the log file and the terminal, but the \TeX\ run is not interrupted.
Issues ⟨module⟩ information ⟨message⟩, passing ⟨arg one⟩ to ⟨arg four⟩ to the text-creating functions. For the more common \msg_info:nnnnnn, the information text is added to the log file only, while \msg_note:nnnnnn adds the info text to both the log file and the terminal. The \TeX run is not interrupted.

\msg_term:nnnnnn \msg_term:nnxxxx \msg_term:nnnnn \msg_term:nnxxx \msg_term:nnxx \msg_term:nn \msg_term:nnn \msg_term:nnx \msg_term:nnxx \msg_term:nnxxx \msg_term:nnxxxx \msg_term:nnxxxxx

Updated: 2012-08-11

\msg_log:nnnnnn \msg_log:nnxxxx \msg_log:nnnnn \msg_log:nnxxx \msg_log:nnxx \msg_log:nn \msg_log:nnn \msg_log:nnx \msg_log:nnxx \msg_log:nnxxx \msg_log:nnxxxx \msg_log:nnxxxxx

Updated: 2012-08-11

\msg_none:nnnnnn \msg_none:nnxxxx \msg_none:nnnnn \msg_none:nnxxx \msg_none:nnnn \msg_none:nnxx \msg_none:nnn \msg_none:nnx \msg_none:nn

Updated: 2012-08-11

\msg_note:nnnnnn \{module\} \{message\} \{⟨arg one⟩\} \{⟨arg two⟩\} \{⟨arg three⟩\} \{⟨arg four⟩\}

Does nothing: used as a message class to prevent any output at all (see the discussion of message redirection).
11.4.1 Messages for showing material

\msg_show:nnnnn \langle module \rangle \langle message \rangle \langle \text{arg one} \rangle \langle \text{arg two} \rangle \langle \text{arg three} \rangle \langle \text{arg four} \rangle

Issues \langle module \rangle \text{information} \langle message \rangle, passing \langle \text{arg one} \rangle \text{to} \langle \text{arg four} \rangle \text{to the text-creating functions. The information text is shown on the terminal and the \TeX run is interrupted in a manner similar to} \tl_show:n. \text{This is used in conjunction with} \msg_show_item:n \text{and similar functions to print complex variable contents completely. If the formatted text does not contain} \texttt{>~} \text{at the start of a line, an additional line} \texttt{>~.} \text{will be put at the end. In addition, a final period is added if not present.}

11.4.2 Expandable error messages

In very rare cases it may be necessary to produce errors in an expansion-only context. The functions in this section should only be used if there is no alternative approach using \msg_error:nnnnnn or other non-expandable commands from the previous section. Despite having a similar interface as non-expandable messages, expandable errors must be handled internally very differently from normal error messages, as none of the tools to print to the terminal or the log file are expandable. As a result, short-hands such as { or \ do not work, and messages must be very short (with default settings, they are truncated after approximately 50 characters). It is advisable to ensure that the message is understandable even when truncated, by putting the most important information up front. Another particularity of expandable messages is that they cannot be redirected or turned off by the user.

\msg_expandable_error:nnnnnn \star \msg_expandable_error:nnffff \star \msg_expandable_error:nnnnnn \star \msg_expandable_error:nnff \star \msg_expandable_error:nnf \star \msg_expandable_error:nn \star

Issues an “Undefined error” message from \TeX itself using the undefined control sequence \texttt{\::error} then prints “! \langle module \rangle: ”\langle error message \rangle, which should be short. With default settings, anything beyond approximately 50 characters long (or bytes in some engines) is cropped. A leading space might be removed as well.

11.5 Redirecting messages

Each message has a “name”, which can be used to alter the behaviour of the message when it is given. Thus we might have

\msg_new:nnnn \{ module \} \{ my-message \} \{ Some-text \} \{ Some-more-text \}
to define a message, with

\msg_error:nn { module } { my-message }

when it is used. With no filtering, this raises an error. However, we could alter the
behaviour with

\msg_redirect_class:nn { error } { warning }

to turn all errors into warnings, or with

\msg_redirect_module:nnn { module } { error } { warning }

to alter only messages from that module, or even

\msg_redirect_name:nnn { module } { my-message } { warning }

to target just one message. Redirection applies first to individual messages, then to
messages from one module and finally to messages of one class. Thus it is possible to
select out an individual message for special treatment even if the entire class is already
redirected.

Multiple redirections are possible. Redirections can be cancelled by providing an
empty argument for the target class. Redirection to a missing class raises an error
immediately. Infinite loops are prevented by eliminating the redirection starting from
the target of the redirection that caused the loop to appear. Namely, if redirections are
requested as $A \rightarrow B$, $B \rightarrow C$ and $C \rightarrow A$ in this order, then the $A \rightarrow B$ redirection is
cancelled.

Changes the behaviour of messages of \texttt{class one} so that they are processed using the
code for those of \texttt{class two}. Each \texttt{class} can be one of fatal, critical, error,
warning, note, info, term, log, none.

Redirects message of \texttt{class one} from \texttt{module} to act as though they were from \texttt{class two}. Messages of \texttt{class one} from sources other than \texttt{module} are not affected by this
redirection. This function can be used to make some messages “silent” by default. For
example, all of the warning messages of \texttt{module} could be turned off with:

\msg_redirect_module:nnn { module } { warning } { none }

Redirects a specific \texttt{message} from a specific \texttt{module} to act as a member of \texttt{class} of
messages. No further redirection is performed. This function can be used to make a
selected message “silent” without changing global parameters:

\msg_redirect_name:nnn { module } { annoying-message } { none }

Updated: 2012-04-27

Updated: 2012-04-27

Updated: 2012-04-27
Chapter 12

The l3file package
File and I/O operations

This module provides functions for working with external files. Some of these functions apply to an entire file, and have prefix \file_, while others are used to work with files on a line by line basis and have prefix \ior_ (reading) or \iow_ (writing).

It is important to remember that when reading external files \TeX{} attempts to locate them using both the operating system path and entries in the \TeX{} file database (most \TeX{} systems use such a database). Thus the “current path” for \TeX{} is somewhat broader than that for other programs.

For functions which expect a ⟨file name⟩ argument, this argument may contain both literal items and expandable content, which should on full expansion be the desired file name. Active characters (as declared in \l_char_active_seq) are not expanded, allowing the direct use of these in file names. Quote tokens (") are not permitted in file names as they are reserved for internal use by some \TeX{} primitives.

Spaces are trimmed at the beginning and end of the file name: this reflects the fact that some file systems do not allow or interact unpredictably with spaces in these positions. When no extension is given, this will trim spaces from the start of the name only.

12.1 Input–output stream management

As \TeX{} engines have a limited number of input and output streams, direct use of the streams by the programmer is not supported in \LaTeX3. Instead, an internal pool of streams is maintained, and these are allocated and deallocated as needed by other modules. As a result, the programmer should close streams when they are no longer needed, to release them for other processes.

Note that I/O operations are global: streams should all be declared with global names and treated accordingly.
Globally reserves the name of the \langle stream \rangle, either for reading or for writing as appropriate. The \langle stream \rangle is not opened until the appropriate \texttt{\_open:N} function is used. Attempting to use a \langle stream \rangle which has not been opened is an error, and the \langle stream \rangle will behave as the corresponding \texttt{\_term...}.

\ior_{\texttt{\_open:Nn}}\langle file name \rangle

Opens \langle file name \rangle for reading using \langle stream \rangle as the control sequence for file access. If the \langle stream \rangle was already open it is closed before the new operation begins. The \langle stream \rangle is available for access immediately and will remain allocated to \langle file name \rangle until a \texttt{\_close:N} instruction is given or the \TeX{} run ends. If the file is not found, an error is raised.

\ior_{\texttt{\_open:NnTF}} \langle file name \rangle \{\{true code\}\{false code\}\}

Opens \langle file name \rangle for reading using \langle stream \rangle as the control sequence for file access. If the \langle stream \rangle was already open it is closed before the new operation begins. The \langle stream \rangle is available for access immediately and will remain allocated to \langle file name \rangle until a \texttt{\_close:N} instruction is given or the \TeX{} run ends. The \langle true code \rangle is then inserted into the input stream. If the file is not found, no error is raised and the \langle false code \rangle is inserted into the input stream.

\ior_{\texttt{\_open:cn}}\langle stream \rangle \{\langle file name \rangle\}

Opens \langle file name \rangle for writing using \langle stream \rangle as the control sequence for file access. If the \langle stream \rangle was already open it is closed before the new operation begins. The \langle stream \rangle is available for access immediately and will remain allocated to \langle file name \rangle until a \texttt{\_close:N} instruction is given or the \TeX{} run ends. Opening a file for writing clears any existing content in the file (i.e. writing is not additive).

\ior_{\texttt{\_close:N}}\langle stream \rangle

Closes the \langle stream \rangle. Streams should always be closed when they are finished with as this ensures that they remain available to other programmers.

\ior_{\texttt{\_show:N}}\langle stream \rangle

Display (to the terminal or log file) the file name associated to the \langle read or write \rangle \langle stream \rangle.
Display (to the terminal or log file) a list of the file names associated with each open (read or write) stream. This is intended for tracking down problems.

12.1.1 Reading from files

Reading from files and reading from the terminal are separate processes in expl3. The functions \ior_get:NN and \ior_str_get:NN, and their branching equivalents, are designed to work with files.

\ior_get:NN (stream) (token list variable)  
\ior_get:NNTF (stream) (token list variable) (true code) (false code)

Function that reads one or more lines (until an equal number of left and right braces are found) from the file input (stream) and stores the result locally in the (token list) variable. The material read from the (stream) is tokenized by TEX according to the category codes and \endlinechar in force when the function is used. Assuming normal settings, any lines which do not end in a comment character % have the line ending converted to a space, so for example input

\begin{verbatim}
a b c
\end{verbatim}

results in a token list a b c \par. Any blank line is converted to the token \par. Therefore, blank lines can be skipped by using a test such as

\begin{verbatim}
\ior_get:NN l_my_stream \l_tmpa_tl
\tl_set:Nn \l_tmpb_tl \{ \par \}
\tl_if_eq:NNF \l_tmpa_tl \l_tmpb_tl \l_tmpa_tl \l_tmpb_tl
\end{verbatim}

Also notice that if multiple lines are read to match braces then the resulting token list can contain \par tokens. In the non-branching version, where the (stream) is not open the (tl var) is set to \q_no_value.

\TeXhackers note: This protected macro is a wrapper around the \TeXx primitive \read. Regardless of settings, \TeXx replaces trailing space and tab characters (character codes 32 and 9) in each line by an end-of-line character (character code \endlinechar, omitted if \endlinechar is negative or too large) before turning characters into tokens according to current category codes. With default settings, spaces appearing at the beginning of lines are also ignored.
The \ior_str_get:NN macro reads one line from the file input \langle stream \rangle and stores the result locally in the \langle token list \rangle variable. The material is read from the \langle stream \rangle as a series of tokens with category code 12 (other), with the exception of space characters which are given category code 10 (space). Multiple whitespace characters are retained by this process. It always only reads one line and any blank lines in the input result in the \langle token list variable \rangle being empty. Unlike \ior_get:NN, line ends do not receive any special treatment. Thus input
\begin{verbatim}
  a b c
\end{verbatim}
results in a token list \begin{verbatim}a b c\end{verbatim} with the letters a, b, and c having category code 12. In the non-branching version, where the \langle stream \rangle is not open the \langle tl var \rangle is set to \q_no_value.

\TeXhacksnote: This protected macro is a wrapper around the \e-\TeX\ primitive \texttt{\newline}. Regardless of settings, \TeX\ removes trailing space and tab characters (character codes 32 and 9). However, the end-line character normally added by this primitive is not included in the result of \ior_str_get:NN.

All mappings are done at the current group level, \ie any local assignments made by the \langle function \rangle or \langle code \rangle discussed below remain in effect after the loop.

\ior_map_inline:Nn \ior_map_inline:Nn \langle inline function \rangle

\TeXhacksnote: For each set of \langle lines \rangle obtained by calling \ior_get:NN until reaching the end of the file, \ror\ stores the \langle lines \rangle in the \langle tl var \rangle then applies the \langle code \rangle. The \langle code \rangle will usually make use of the \langle variable \rangle, but this is not enforced. The assignments to the \langle variable \rangle are local. Its value after the loop is the last set of \langle lines \rangle, or its original value if the \langle stream \rangle is empty. \TeX\ ignores any trailing new-line marker from the file it reads.

All mappings are done at the current group level, \ie any local assignments made by the \langle function \rangle or \langle code \rangle discussed below remain in effect after the loop.
\ior_str_map_variable:Nnn \ior_map_break: \ior_map_break:n

\ior_str_map_variable:Nnn \langle stream \rangle \langle variable \rangle \{ \langle code \rangle \}

For each \langle line \rangle in the \langle stream \rangle, stores the \langle line \rangle in the \langle variable \rangle then applies the \langle code \rangle. The material is read from the \langle stream \rangle as a series of tokens with category code 12 (other), with the exception of space characters which are given category code 10 (space). The \langle code \rangle will usually make use of the \langle variable \rangle, but this is not enforced. The assignments to the \langle variable \rangle are local. Its value after the loop is the last \langle line \rangle, or its original value if the \langle stream \rangle is empty. Note that \TeX removes trailing space and tab characters (character codes 32 and 9) from every line upon input. \TeX also ignores any trailing new-line marker from the file it reads. This function is typically faster than \texttt{\ior_str_map_inline:Nn}.

\ior_map_break:

\texttt{\ior_map_break:}

Used to terminate a \texttt{\ior_map...} function before all lines from the \langle stream \rangle have been processed. This normally takes place within a conditional statement, for example

\begin{verbatim}
\ior_map_inline:Nn \l_my_iore
 { \str_if_eq:nnTF { #1 } { bingo } { \ior_map_break: } { % Do something useful } }
\end{verbatim}

Use outside of a \texttt{\ior_map...} scenario leads to low level \TeX errors.

\TeXhackers note: When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.

\ior_map_break:n

\texttt{\ior_map_break:n \langle code \rangle}

Used to terminate a \texttt{\ior_map...} function before all lines in the \langle stream \rangle have been processed, inserting the \langle code \rangle after the mapping has ended. This normally takes place within a conditional statement, for example

\begin{verbatim}
\ior_map_inline:Nn \l_my_iore
 { \str_if_eq:nnTF { #1 } { bingo } { \ior_map_break:n \langle code \rangle } { % Do something useful } }
\end{verbatim}

Use outside of a \texttt{\ior_map...} scenario leads to low level \TeX errors.

\TeXhackers note: When the mapping is broken, additional tokens may be inserted before the \langle code \rangle is inserted into the input stream. This depends on the design of the mapping function.
Tests if the end of a file \textit{stream} has been reached during a reading operation. The test also returns a \texttt{true} value if the \textit{stream} is not open.

12.1.2 Writing to files

\texttt{\iow_now:Nn} \texttt{(stream) \{\langle tokens\rangle\}}

This function writes \textit{\langle tokens\rangle} to the specified \textit{stream} immediately (\textit{i.e.} the write operation is called on expansion of \texttt{\iow_now:Nn}).

\texttt{\iow_log:n} \texttt{\{\langle tokens\rangle\}}

This function writes the given \textit{\langle tokens\rangle} to the log (transcript) file immediately: it is a dedicated version of \texttt{\iow_now:Nn}.

\texttt{\iow_term:n} \texttt{\{\langle tokens\rangle\}}

This function writes the given \textit{\langle tokens\rangle} to the terminal file immediately: it is a dedicated version of \texttt{\iow_now:Nn}.

\texttt{\iow_shipout:Nn} \texttt{\{\langle tokens\rangle\}}

This function writes \textit{\langle tokens\rangle} to the specified \textit{stream} when the current page is finalised (\textit{i.e.} at shipout). The \texttt{x}-type variants expand the \textit{\langle tokens\rangle} at the point where the function is used but \texttt{not} when the resulting tokens are written to the \textit{stream} (cf. \texttt{\iow_shipout\_x:Nn}).

\textbf{TEX hackers note:} When using expl3 with a format other than \LaTeX, new line characters inserted using \texttt{\iow_newline:} or using the line-wrapping code \texttt{\iow_wrap:nnnN} are not recognized in the argument of \texttt{\iow_shipout:Nn}. This may lead to the insertion of additional unwanted line-breaks.

\texttt{\iow_shipout_x:Nn} \texttt{\{\langle tokens\rangle\}}

This function writes \textit{\langle tokens\rangle} to the specified \textit{stream} when the current page is finalised (\textit{i.e.} at shipout). The \textit{\langle tokens\rangle} are expanded at the time of writing in addition to any expansion when the function is used. This makes these functions suitable for including material finalised during the page building process (such as the page number integer).

\textbf{TEX hackers note:} This is a wrapper around the \TeX{} primitive \texttt{\write}. When using expl3 with a format other than \LaTeX, new line characters inserted using \texttt{\iow_newline:} or using the line-wrapping code \texttt{\iow_wrap:nnnN} are not recognized in the argument of \texttt{\iow_shipout:Nn}. This may lead to the insertion of additional unwanted line-breaks.
\iow_char:N \iow_char:N \langle \text{char} \rangle

Inserts \langle \text{char} \rangle into the output stream. Useful when trying to write difficult characters such as \%, \{, \}, etc. in messages, for example:

\iow_now:Nx \g_my_iow { \iow_char:N \{ text \iow_char:N \} }

The function has no effect if writing is taking place without expansion (e.g. in the second argument of \iow_now:Nn).

\iow_newline: \iow_newline:

Function to add a new line within the \langle \text{tokens} \rangle written to a file. The function has no effect if writing is taking place without expansion (e.g. in the second argument of \iow_now:Nn).

\TeX{}hackers note: When using expl3 with a format other than \LaTeX{}, the character inserted by \iow_newline: is not recognized by \TeX{}, which may lead to the insertion of additional unwanted line-breaks. This issue only affects \iow_shipout:Nn, \iow_shipout_x:Nn and direct uses of primitive operations.
12.1.3 Wrapping lines in output

\iow_wrap:nnnN \iow_wrap:nnnN \iow_wrap:nxnN

This function wraps the \langle text \rangle to a fixed number of characters per line. At the start of each line which is wrapped, the \langle run-on text \rangle is inserted. The line character count targeted is the value of \texttt{l_iow_line_count_int} minus the number of characters in the \langle run-on text \rangle for all lines except the first, for which the target number of characters is simply \texttt{l_iow_line_count_int} since there is no run-on text. The \langle text \rangle and \langle run-on text \rangle are exhaustively expanded by the function, with the following substitutions:

- \texttt{\textbackslash{} or \iow_newline:} may be used to force a new line,
- \texttt{\textbackslash{} may be used to represent a forced space (for example after a control sequence),}
- \texttt{\#, \%, \{, \}, \~{} may be used to represent the corresponding character,}
- \texttt{\iow_allow_break:} may be used to allow a line-break without inserting a space (this is experimental),
- \texttt{\iow_indent:n} may be used to indent a part of the \langle text \rangle (not the \langle run-on text \rangle).

Additional functions may be added to the wrapping by using the \langle set up \rangle, which is executed before the wrapping takes place: this may include overriding the substitutions listed.

Any expandable material in the \langle text \rangle which is not to be expanded on wrapping should be converted to a string using \texttt{\token_to_str:N}, \texttt{\tl_to_str:n}, \texttt{\tl_to_str:N}, etc.

The result of the wrapping operation is passed as a braced argument to the \langle function \rangle, which is typically a wrapper around a write operation. The output of \texttt{\iow_wrap:nnnN} (i.e. the argument passed to the \langle function \rangle) consists of characters of category “other” (category code 12), with the exception of spaces which have category “space” (category code 10). This means that the output does not expand further when written to a file.

\TeX{}hackers note: Internally, \texttt{\iow_wrap:nnnN} carries out an x-type expansion on the \langle text \rangle to expand it. This is done in such a way that \texttt{\exp_not:N} or \texttt{\exp_not:n could} be used to prevent expansion of material. However, this is less conceptually clear than conversion to a string, which is therefore the supported method for handling expandable material in the \langle text \rangle.

\iow_indent:n \iow_indent:n \iow_indent:n \langle text \rangle

In the first argument of \texttt{\iow_wrap:nnnN} (for instance in messages), indents \langle text \rangle by four spaces. This function does not cause a line break, and only affects lines which start within the scope of the \langle text \rangle. In case the indented \langle text \rangle should appear on separate lines from the surrounding text, use \texttt{\textbackslash{}} to force line breaks.

\texttt{l_iow_line_count_int}

The maximum number of characters in a line to be written by the \texttt{\iow_wrap:nnnN} function. This value depends on the \TeX{} system in use: the standard value is 78, which is typically correct for unmodified \TeX{} Live and MiKTeX systems.
12.1.4 Constant input–output streams, and variables

Scratch input stream for global use. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\begin{verbatim}
\g_tmpa_ior \g_tmpb_ior
Rev: 2017-12-11
\end{verbatim}

Constant output streams for writing to the log and to the terminal (plus the log), respectively.

\begin{verbatim}
\c_log_iow \c_term_iow
\end{verbatim}

Scratch output stream for global use. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\begin{verbatim}
\g_tmpa_iow \g_tmpb_iow
Rev: 2017-12-11
\end{verbatim}

12.1.5 Primitive conditionals

\begin{verbatim}
\if_eof:w * \if_eof:w (stream)
  (true code)
\else:
  (false code)
\fi:
\end{verbatim}

Tests if the \langle stream \rangle returns “end of file”, which is true for non-existent files. The \texttt{\else:} branch is optional.

\TeXhackers note: This is the \TeX primitive \texttt{\if_eof}.

12.2 File operation functions

Contain the directory, name and extension of the current file. The directory is empty if the file was loaded without an explicit path (\textit{i.e.} if it is in the \TeX search path), and does not end in / other than the case that it is exactly equal to the root directory. The \langle name \rangle and \langle ext \rangle parts together make up the file name, thus the \langle name \rangle part may be thought of as the “job name” for the current file. Note that \TeX does not provide information on the \langle ext \rangle part for the main (top level) file and that this file always has an empty \langle dir \rangle component. Also, the \langle name \rangle here will be equal to \texttt{\c_sys_jobname_str}, which may be different from the real file name (if set using \texttt{--jobname}, for example).

\begin{verbatim}
\g_file_curr_dir_str \g_file_curr_name_str \g_file_curr_ext_str
Rev: 2017-06-21
\end{verbatim}
Each entry is the path to a directory which should be searched when seeking a file. Each path can be relative or absolute, and should not include the trailing slash. The entries are not expanded when used so may contain active characters but should not feature any variable content. Spaces need not be quoted.

**\TeXhackers note:** When working as a package in \LaTeX\ expl3 will automatically append the current \texttt{\input@path} to the set of values from \texttt{\l_file_search_path_seq}.

\file_if_exist:nTF{⟨\texttt{file name}⟩}{⟨\texttt{true code}⟩}{⟨\texttt{false code}⟩}

Searches for ⟨\texttt{file name}⟩ using the current \TeX\ search path and the additional paths controlled by \texttt{\l_file_search_path_seq}.

\file_get:nnN{⟨\texttt{filename}⟩}{⟨\texttt{setup}⟩}{⟨\texttt{tl}⟩}

\file_get:nnNTF{⟨\texttt{filename}⟩}{⟨\texttt{setup}⟩}{⟨\texttt{tl}⟩}{⟨\texttt{true code}⟩}{⟨\texttt{false code}⟩}

Defines ⟨\texttt{tl}⟩ to the contents of ⟨\texttt{filename}⟩. Category codes may need to be set appropriately via the ⟨\texttt{setup}⟩ argument. The non-branching version sets the ⟨\texttt{tl}⟩ to \texttt{\q_no_value} if the file is not found. The branching version runs the ⟨\texttt{true code}⟩ after the assignment to ⟨\texttt{tl}⟩ if the file is found, and ⟨\texttt{false code}⟩ otherwise.

\file_get_full_name:nN
\file_get_full_name:VN
\file_get_full_name:nNTF
\file_get_full_name:VN

\file_get_full_name:nN{⟨\texttt{file name}⟩}{⟨\texttt{tl}⟩}
\file_get_full_name:nNTF{⟨\texttt{file name}⟩}{⟨\texttt{tl}⟩}{⟨\texttt{true code}⟩}{⟨\texttt{false code}⟩}

Searches for ⟨\texttt{file name}⟩ in the path as detailed for \texttt{\file_if_exist:nTF}, and if found sets the ⟨\texttt{tl var}⟩ the fully-qualified name of the file, i.e.~the path and file name. This includes an extension '.tex' when the given ⟨\texttt{file name}⟩ has no extension but the file found has that extension. In the non-branching version, the ⟨\texttt{tl var}⟩ will be set to \texttt{\q_no_value} in the case that the file does not exist.

\file_full_name:n
\file_full_name:V

\file_full_name:n{⟨\texttt{file name}⟩}

Searches for ⟨\texttt{file name}⟩ in the path as detailed for \texttt{\file_if_exist:nTF}, and if found leaves the fully-qualified name of the file, i.e.~the path and file name, in the input stream. This includes an extension '.tex' when the given ⟨\texttt{file name}⟩ has no extension but the file found has that extension. If the file is not found on the path, the expansion is empty.
\file_parse_full_name:nNNN \file_parse_full_name:VNNN

\file_parse_full_name:nNNN \file_parse_full_name:VNNN

\file_parse_full_name:n {\langle full name \rangle} \langle dir \rangle \langle name \rangle \langle ext \rangle

\file_parse_full_name:n {\langle full name \rangle}

\file_parse_full_name_apply:nN {\langle full name \rangle} \langle function \rangle

\file_hex_dump:n {\langle file name \rangle}
\file_hex_dump:nnn {\langle file name \rangle} \langle start index \rangle \langle end index \rangle

\file_get_hex_dump:nN {\langle file name \rangle} \langle tl var \rangle
\file_get_hex_dump:nnn {\langle file name \rangle} \langle start index \rangle \langle end index \rangle \langle tl var \rangle

Parses the \langle full name \rangle and splits it into three parts, each of which is returned by setting the appropriate local string variable:

- The \langle dir \rangle: everything up to the last / (path separator) in the \langle file path \rangle. As with system PATH variables and related functions, the \langle dir \rangle does not include the trailing / unless it points to the root directory. If there is no path (only a file name), \langle dir \rangle is empty.

- The \langle name \rangle: everything after the last / up to the last ., where both of those characters are optional. The \langle name \rangle may contain multiple . characters. It is empty if \langle full name \rangle consists only of a directory name.

- The \langle ext \rangle: everything after the last . (including the dot). The \langle ext \rangle is empty if there is no . after the last /

Before parsing, the \langle full name \rangle is expanded until only non-expandable tokens remain, except that active characters are also not expanded. Quotes (") are invalid in file names and are discarded from the input.

\file_parse_full_name:n *
\file_parse_full_name:nnn *

\file_parse_full_name.apply:nN *
\file_parse_full_name.apply:nnn *

\file_hex_dump:n *
\file_hex_dump:nnn *

\file_get_hex_dump:nN *
\file_get_hex_dump:nnnN *

Sets the \langle tl var \rangle to the result of applying \file_hex_dump:n/\file_hex_dump:nnn to the \langle file \rangle. If the file is not found, the \langle tl var \rangle will be set to \q_no_value.

93
\file_mdfive_hash:n \(\star\) \hspace{1em} \file_mdfive_hash:n \{\langle \text{file name} \rangle\}

Searches for \(\langle \text{file name} \rangle\) using the current \TeX{} search path and the additional paths controlled by \l_file_search_path_seq. It then expands to leave the MD5 sum generated from the contents of the file in the input stream. The file is read as bytes, which means that in contrast to most \TeX{} behaviour there will be a difference in result depending on the line endings used in text files. The same file will produce the same result between different engines: the algorithm used is the same in all cases. When the file is not found, the result of expansion is empty.

\file_get_mdfive_hash:nN \{\langle \text{file name} \rangle\} \langle \tl var \rangle

Sets the \(\langle \tl var \rangle\) to the result of applying \file_mdfive_hash:n to the \(\langle \text{file} \rangle\). If the file is not found, the \(\langle \tl var \rangle\) will be set to \q_no_value.

\file_size:n \{\langle \text{file name} \rangle\}

Searches for \(\langle \text{file name} \rangle\) using the current \TeX{} search path and the additional paths controlled by \l_file_search_path_seq. It then expands to leave the size of the file in bytes in the input stream. When the file is not found, the result of expansion is empty.

\file_get_size:nN \{\langle \text{file name} \rangle\} \langle \tl var \rangle

Sets the \(\langle \tl var \rangle\) to the result of applying \file_size:n to the \(\langle \text{file} \rangle\). If the file is not found, the \(\langle \tl var \rangle\) will be set to \q_no_value. This is not available in older versions of \XeTeX{}.

\file_timestamp:n \(\star\) \hspace{1em} \file_timestamp:n \{\langle \text{file name} \rangle\}

Searches for \(\langle \text{file name} \rangle\) using the current \TeX{} search path and the additional paths controlled by \l_file_search_path_seq. It then expands to leave the modification timestamp of the file in the input stream. The timestamp is of the form \(D:\langle \text{year} \rangle\langle \text{month} \rangle\langle \text{day} \rangle\langle \text{hour} \rangle\langle \text{minute} \rangle\langle \text{second} \rangle\langle \text{offset} \rangle\), where the latter may be \(Z\) (UTC) or \(\langle \text{plus-minus} \rangle\langle \text{hours} \rangle'\langle \text{minutes} \rangle''. When the file is not found, the result of expansion is empty. This is not available in older versions of \XeTeX{}.

\file_get_timestamp:nN \{\langle \text{file name} \rangle\} \langle \tl var \rangle

Sets the \(\langle \tl var \rangle\) to the result of applying \file_timestamp:n to the \(\langle \text{file} \rangle\). If the file is not found, the \(\langle \tl var \rangle\) will be set to \q_no_value. This is not available in older versions of \XeTeX{}.  

94
\file_compare_timestamp_p:nNn *  
\file_compare_timestamp:nNnT *  
\file_compare_timestamp:nNnTF *  

\texttt{\file_compare_timestamp:nNnT \{source-file\} > \{derived-file\}}  
\{
  \% Code to regenerate derived file
\}

to work when the derived file is entirely absent. The timestamp of two absent files is
regarded as different. This is not available in older versions of \TeX.

\file_input:n  
\file_input:n {\{file name\}}

Searches for \{file name\} in the path as detailed for \file_if_exist:nTF, and if found
reads in the file as additional \LaTeX source. All files read are recorded for information
and the file name stack is updated by this function. An error is raised if the file is not
found.

\file_if_exist_input:n  
\file_if_exist_input:n {\{file name\}}  
\file_if_exist_input:nF {\{file name\}} {\{false code\}}

Searches for \{file name\} using the current \TeX search path and the additional paths
included in \l_file_search_path_seq. If found then reads in the file as additional \LaTeX source as described for \file_input:n, otherwise inserts the \{false code\}. Note
that these functions do not raise an error if the file is not found, in contrast to \file_-
input:n.

\file_input_stop:  
\file_input_stop:

Ends the reading of a file started by \file_input:n or similar before the end of the
file is reached. Where the file reading is being terminated due to an error, \msg_-critical:n(n) should be preferred.

\textbf{\TeX hackers note:} This function must be used on a line on its own: \TeX reads files
line-by-line and so any additional tokens in the “current” line will still be read.

This is also true if the function is hidden inside another function (which will be the normal
case), i.e., all tokens on the same line in the source file are still processed. Putting it on a line
by itself in the definition doesn’t help as it is the line where it is used that counts!

\file_show_list:  
\file_log_list:  
\file_input:nn:nn

These functions list all files loaded by \LaTeX commands that populate \@filelist or
by \file_input:n. While \file_show_list: displays the list in the terminal, \file_-
log_list: outputs it to the log file only.
Chapter 13

The \texttt{l3luatex} package: \LaTeX\-specific functions

The \LaTeX{} engine provides access to the Lua programming language, and with it access to the “internals” of \TeX{}. In order to use this within the framework provided here, a family of functions is available. When used with \texttt{pdf\TeX{}}, \texttt{\LaTeX{}}, \texttt{up\TeX{}} or \texttt{Xe\TeX{}} these raise an error: use \texttt{\sys_if_engine_luatex:T} to avoid this. Details on using Lua with the \LaTeX{} engine are given in the \LaTeX{} manual.

13.1 Breaking out to Lua

\texttt{\lua_now:n \{token list\}}

The \{token list\} is first tokenized by \TeX{}, which includes converting line ends to spaces in the usual \TeX{} manner and which respects currently-applicable \TeX{} category codes. The resulting \{Lua input\} is passed to the Lua interpreter for processing. Each \texttt{\lua_now:n} block is treated by Lua as a separate chunk. The Lua interpreter executes the \{Lua input\} immediately, and in an expandable manner.

\TeX{}hackers note: \texttt{\lua_now:e} is a macro wrapper around \texttt{\directlua}: when Lua\TeX{} is in use two expansions are required to yield the result of the Lua code.

\texttt{\lua_shipout:n \{token list\}}

The \{token list\} is first tokenized by \TeX{}, which includes converting line ends to spaces in the usual \TeX{} manner and which respects currently-applicable \TeX{} category codes. The resulting \{Lua input\} is passed to the Lua interpreter when the current page is finalised (\emph{i.e.} at shipout). Each \texttt{\lua_shipout:n} block is treated by Lua as a separate chunk. The Lua interpreter will execute the \{Lua input\} during the page-building routine: no \TeX{} expansion of the \{Lua input\} will occur at this stage.

In the case of the \texttt{\lua_shipout_e:n} version the input is fully expanded by \TeX{} in an \texttt{e}-type manner during the shipout operation.

\TeX{}hackers note: At a \TeX{} level, the \{Lua input\} is stored as a “whatsit”.

\texttt{\lua_shipout_e:n \lua_shipout:n}

New: 2018-06-18
\lua_escape:n \{token list\}
\lua_escape:e

Converts the \{token list\} such that it can safely be passed to Lua: embedded backslashes, double and single quotes, and newlines and carriage returns are escaped. This is done by prepending an extra token consisting of a backslash with category code 12, and for the line endings, converting them to \n and \r, respectively.

\textbf{\LaTeX} hackers note: \lua_escape:e is a macro wrapper around \luaescapestring: when \LaTeX\L with \luaescapestring\ is in use two expansions are required to yield the result of the Lua code.

### 13.2 Lua interfaces

As well as interfaces for \TeX, there are a small number of Lua functions provided here.

\texttt{ltx.utils}

Most public interfaces provided by the module are stored within the \texttt{ltx.utils} table.

\texttt{l3kernel}

For compatibility reasons, there are also some deprecated interfaces provided in the \texttt{l3kernel} table. These do not return their result as Lua values but instead print them to \TeX.

\texttt{l3kernel.charcat}

\texttt{l3kernel.charcat(\{charcode\}, \{catcode\})}

Constructs a character of \{charcode\} and \{catcode\} and returns the result to \TeX.

\texttt{l3kernel.elapsedtime}

\texttt{l3kernel.elapsedtime()}

Returns the CPU time in \langle scaled seconds \rangle since the start of the \TeX run or since \texttt{l3kernel.resettimer} was issued. This only measures the time used by the CPU, not the real time, e.g., waiting for user input.

\texttt{ltx.utils.filedump}

\texttt{ltx.utils.filedump(\{file\}, \{offset\}, \{length\})}

\texttt{l3kernel.filedump(\{file\}, \{offset\}, \{length\})}

Returns the uppercase hexadecimal representation of the content of the \{file\} read as bytes. If the \{length\} is given, only this part of the file is returned; similarly, one may specify the \{offset\} from the start of the file. If the \{length\} is not given, the entire file is read starting at the \{offset\}.

\texttt{ltx.utils.filemd5sum}

\texttt{ltx.utils.filemd5sum(\{file\})}

\texttt{l3kernel.filemd5sum(\{file\})}

Returns the MD5 sum of the file contents read as bytes; note that the result will depend on the nature of the line endings used in the file, in contrast to normal \TeX behaviour. If the \{file\} is not found, nothing is returned with \textit{no error raised}. 

97
<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>ltx.utils.filemoddate</code></td>
<td>[(\text{date}) = \text{ltx.utils.filemoddate}(\langle\text{file}\rangle)] Returns the date/time of last modification of the \langle\text{file}\rangle in the format D:⟨year⟩⟨month⟩⟨day⟩⟨hour⟩⟨minute⟩⟨second⟩⟨offset⟩ where the latter may be Z (UTC) or \langle\text{plus-minus}\rangle⟨hours⟩\langle\text{minutes}\rangle. If the \langle\text{file}\rangle is not found, nothing is returned with \textit{no error raised}.</td>
</tr>
<tr>
<td><code>l3kernel.filemoddate</code></td>
<td>[\text{l3kernel.filemoddate}(\langle\text{file}\rangle)]</td>
</tr>
<tr>
<td><code>ltx.utils.filesize</code></td>
<td>[\text{size} = \text{ltx.utils.filesize}(\langle\text{file}\rangle)] Returns the size of the \langle\text{file}\rangle in bytes. If the \langle\text{file}\rangle is not found, nothing is returned with \textit{no error raised}.</td>
</tr>
<tr>
<td><code>l3kernel.filesize</code></td>
<td>[\text{l3kernel.filesize}(\langle\text{file}\rangle)]</td>
</tr>
<tr>
<td><code>l3kernel.resettimer()</code></td>
<td>[\text{l3kernel.resettimer}()] Resets the timer used by \texttt{l3kernel.elapsetime}.</td>
</tr>
<tr>
<td><code>l3kernel.shellescape()</code></td>
<td>[\text{l3kernel.shellescape}(\langle\text{cmd}\rangle)] Executes the \langle\text{cmd}\rangle and prints to the log as for \texttt{pdftex}.</td>
</tr>
<tr>
<td><code>l3kernel.strcmp()</code></td>
<td>[\text{l3kernel.strcmp}(\langle\text{str one}\rangle, \langle\text{str two}\rangle)] Compares the two strings and returns 0 to \texttt{TeX} if the two are identical.</td>
</tr>
</tbody>
</table>
Chapter 14

The \texttt{l3legacy} package

Interfaces to legacy concepts

There are a small number of \LaTeX{} or \LaTeXe{} concepts which are not used in \pkg{expl3} code but which need to be manipulated when working as a \LaTeXe{} package. To allow these to be integrated cleanly into \pkg{expl3} code, a set of legacy interfaces are provided here.

\begin{verbatim}
\legacy_if_p:n \legacy_if:nTF *
\end{verbatim}

Tests if the \TeX{} or \LaTeXe{} conditional (generated by \texttt{\newif}) if true or false and branches accordingly. The \texttt{name} of the conditional should omit the leading \texttt{if}.

\begin{verbatim}
\legacy_if_set_true:n \legacy_if_set_false:n \legacy_if_gset_true:n \legacy_if_gset_false:n
\end{verbatim}

Sets the \LaTeXe{} or plain \TeX{} conditional \texttt{if(name)} (generated by \texttt{\newif}) to be true or false.

\begin{verbatim}
\legacy_if_set:nn \legacy_if_gset:nn
\end{verbatim}

Sets the \LaTeXe{} or plain \TeX{} conditional \texttt{if(name)} (generated by \texttt{\newif}) to the result of evaluating the \texttt{boolean expression}.  

New: 2021-05-10
Part IV
Data types
Chapter 15

The l3tl package

Token lists

TEX works with tokens, and LaTeX therefore provides a number of functions to deal with lists of tokens. Token lists may be present directly in the argument to a function:

\foo:n { a collection of \tokens }

or may be stored in a so-called “token list variable”, which have the suffix tl: a token list variable can also be used as the argument to a function, for example

\foo:N \l_{\text{some}\_tl}

In both cases, functions are available to test and manipulate the lists of tokens, and these have the module prefix tl. In many cases, functions which can be applied to token list variables are paired with similar functions for application to explicit lists of tokens: the two “views” of a token list are therefore collected together here.

A token list (explicit, or stored in a variable) can be seen either as a list of “items”, or a list of “tokens”. An item is whatever \use:n would grab as its argument: a single non-space token or a brace group, with optional leading explicit space characters (each item is thus itself a token list). A token is either a normal N argument, or \, \{, or \} (assuming normal TEX category codes). Thus for example

\{ Hello \} - world

contains six items (Hello, w, o, r, l and d), but thirteen tokens (\{, H, e, l, l, o, \},, \w, o, r, l and d). Functions which act on items are often faster than their analogue acting directly on tokens.

15.1 Creating and initialising token list variables

\tl_new:N \tl_new:N \tl_new:c

Creates a new \langle tl var \rangle or raises an error if the name is already taken. The declaration is global. The \langle tl var \rangle is initially empty.
\tl_const:Nn \tl_const:(Nx|cn|cx) \tl_const:Nn \tl_var \{\token_list\}

Creates a new constant \langle \tl_var \rangle or raises an error if the name is already taken. The value of the \langle \tl_var \rangle is set globally to the \langle \token_list \rangle.

\tl_clear:N \tl_clear:c \tl_gclear:N \tl_gclear:c

\tl_clear:N \tl_var
Clears all entries from the \langle \tl_var \rangle.

\tl_clear_new:N \tl_clear_new:c \tl_gclear_new:N \tl_gclear_new:c

\tl_clear_new:N \tl_var
Ensures that the \langle \tl_var \rangle exists globally by applying \tl_new:N if necessary, then applies \tl_(g)clear:N to leave the \langle \tl_var \rangle empty.

\tl_set_eq:NN \tl_set_eq:ccc \tl_gset_eq:NN \tl_gset_eq:ccc

\tl_set_eq:NN \tl_var \tl_var_1 \tl_var_2
Sets the content of \langle \tl_var_1 \rangle equal to that of \langle \tl_var_2 \rangle.

\tl_concat:NNN \tl_concat:ccc \tl_gconcat:NNN \tl_gconcat:ccc

\tl_concat:NNN \tl_concat:ccc \tl_var \tl_var_2 \tl_var_3
Concatenates the content of \langle \tl_var_2 \rangle and \langle \tl_var_3 \rangle together and saves the result in \langle \tl_var_1 \rangle. The \langle \tl_var_2 \rangle is placed at the left side of the new token list.

\tl_set:Nn \tl_set:NV \tl_set:Nv \tl_set:No \tl_set:Nf \tl_set:Nx \tl_set:cn \tl_set:cV \tl_set:cv \tl_set:co \tl_set:cf \tl_set:cx

\tl_clear_p:N \tl_clear_p:c \tl_gclear_p:N \tl_gclear_p:c

\tl_if_exist_p:N \tl_if_exist_p:c \tl_if_exist:NTF \tl_if_exist:cTF

Tests whether the \langle \tl_var \rangle is currently defined. This does not check that the \langle \tl_var \rangle really is a token list variable.

New: 2012-05-18

New: 2012-03-03

15.2 Adding data to token list variables

\tl_set:Nn \tl_set:NV \tl_set:Nv \tl_set:No \tl_set:Nf \tl_set:Nx \tl_set:cn \tl_set:cV \tl_set:cv \tl_set:co \tl_set:cf \tl_set:cx

\tl_set:Nn \tl_var \{\tokens\}

Sets \langle \tl_var \rangle to contain \langle \tokens \rangle, removing any previous content from the variable.

\tl_put_left:Nn \tl_put_left:NV \tl_put_left:Nv \tl_put_left:No \tl_put_left:Nf \tl_put_left:Nx \tl_put_left:cn \tl_put_left:cV \tl_put_left:cv \tl_put_left:co \tl_put_left:cf \tl_put_left:cx

\tl_put_left:Nn \tl_put_left:NV \tl_put_left:Nv \tl_put_left:No \tl_put_left:Nf \tl_put_left:Nx \tl_put_left:cn \tl_put_left:cV \tl_put_left:cv \tl_put_left:co \tl_put_left:cf \tl_put_left:cx

Appends \langle \tokens \rangle to the left side of the current content of \langle \tl_var \rangle.
\tl_put_right:Nn \tl_put_right:(NV|No|Nx|cn|cV|co|cx) \tl_gput_right:Nn \tl_gput_right:(NV|No|Nx|cn|cV|co|cx)

Appends \textit{tokens} to the right side of the current content of \textit{tl var}.

### 15.3 Token list conditionals

\begin{verbatim}
\tl_if_blank_p:n \tl_if_blank:nTF \tl_if_empty_p:N \tl_if_empty:NTF \tl_if_eq_p:NN \tl_if_eq:NNTF
\end{verbatim}

Tests if the \textit{token list variable} is entirely empty \textit{(i.e. contains no tokens at all)}.

\begin{verbatim}
\tl_if_blank_p:n \tl_if_blank:nTF \tl_if_empty_p:N \tl_if_empty:NTF
\end{verbatim}

Tests if the \textit{token list variable} is entirely empty \textit{(i.e. contains no tokens at all)}.

\begin{verbatim}
\tl_if_eq_p:NN \tl_if_eq:NNTF
\end{verbatim}

Compares the content of two \textit{token list variables} and is logically \textit{true} if the two contain the same list of tokens \textit{(i.e. identical in both the list of characters they contain and the category codes of those characters)}. Thus for example

\begin{verbatim}
\tl_set:Nn \l_tmpa_tl { abc }
\tl_set:Nx \l_tmpb_tl { \tl_to_str:n { abc } }
\tl_if_eq:NNTF \l_tmpa_tl \l_tmpb_tl { true } { false }
\end{verbatim}

yields \textit{false}. See also \texttt{str_if_eq:nnTF} for a comparison that ignores category codes.

\begin{verbatim}
\tl_if_eq:NnTF \tl_if_eq:cnTF
\end{verbatim}

Tests if the \textit{token list variable} and the \textit{token list} contain the same list of tokens, both in respect of character codes and category codes. This conditional is not expandable: see \texttt{\tl_if_eq:NNTF} for an expandable version when both token lists are stored in variables, or \texttt{str_if_eq:nnTF} if category codes are not important.

Updated: 2019-09-04

New: 2012-05-24

Updated: 2012-06-05

New: 2020-07-14

Updated: 2019-09-04

New: 2012-05-24

Updated: 2012-06-05

New: 2020-07-14

103
\tl_if_eq:nnTF \{\langle \text{token list}_1 \rangle \} \{\langle \text{token list}_2 \rangle \} \{\langle \text{true code} \rangle \} \{\langle \text{false code} \rangle \}

Tests if \langle \text{token list}_1 \rangle and \langle \text{token list}_2 \rangle contain the same list of tokens, both in respect of character codes and category codes. This conditional is not expandable: see \tl_if_eq:NNTF for an expandable version when token lists are stored in variables, or \str_if_eq:nnTF if category codes are not important.

\tl_if_in:NnTF \tl_if_in:NnTF \langle \text{tl var} \rangle \{\langle \text{token list} \rangle \} \{\langle \text{true code} \rangle \} \{\langle \text{false code} \rangle \}

Tests if the \langle \text{token list} \rangle is found in the content of the \langle \text{tl var} \rangle. The \langle \text{token list} \rangle cannot contain the tokens { }, or # (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6).

\tl_if_novalue_p:n \tl_if_novalue_p:n \{\langle \text{token list} \rangle \}

Tests if the \langle \text{token list} \rangle is exactly equal to the special \c_novalue_tl marker. This function is intended to allow construction of flexible document interface structures in which missing optional arguments are detected.

\tl_if_single_p:N \tl_if_single_p:c \tl_if_single:N \tl_if_single:c

Tests if the content of the \langle \text{tl var} \rangle consists of a single \langle \text{item} \rangle, i.e. is a single normal token (neither an explicit space character nor a begin-group character) or a single brace group, surrounded by optional spaces on both sides. In other words, such a token list has token count 1 according to \tl_count:N.

\tl_if_single_p:n \tl_if_single_p:n \{\langle \text{token list} \rangle \}

Tests if the \langle \text{token list} \rangle has exactly one \langle \text{item} \rangle, i.e. is a single normal token (neither an explicit space character nor a begin-group character) or a single brace group, surrounded by optional spaces on both sides. In other words, such a token list has token count 1 according to \tl_count:n.

\tl_if_single_token_p:n \tl_if_single_token_p:n \{\langle \text{token list} \rangle \}

Tests if the token list consists of exactly one token, i.e. is either a single space character or a single normal token. Token groups \{ { } \} are not single tokens.
This function compares the \textit{test token list variable} in turn with each of the \textit{token list variable cases}. If the two are equal (as described for \texttt{tl_if_eq:NNTF}) then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{false code} is inserted. The function \texttt{tl_case:Nn}, which does nothing if there is no match, is also available.

\subsection*{15.3.1 Testing the first token}

Tests if the first \textit{token} in the \textit{token list} has the same category code as the \textit{test token}. In the case where the \textit{token list} is empty, the test is always \textit{false}.

Tests if the first \textit{token} in the \textit{token list} has the same character code as the \textit{test token}. In the case where the \textit{token list} is empty, the test is always \textit{false}.

Tests if the first \textit{token} in the \textit{token list} has the same meaning as the \textit{test token}. In the case where \textit{token list} is empty, the test is always \textit{false}.

Tests if the first \textit{token} in the \textit{token list} is an explicit begin-group character (with category code 1 and any character code), in other words, if the \textit{token list} starts with a brace group. In particular, the test is \textit{false} if the \textit{token list} starts with an implicit token such as \texttt{vc_group_begin_token}, or if it is empty. This function is useful to implement actions on token lists on a token by token basis.
Tests if the first \textit{token} in the \textit{token list} is a normal N-type argument. In other words, it is neither an explicit space character (explicit token with character code 32 and category code 10) nor an explicit begin-group character (with category code 1 and any character code). An empty argument yields \texttt{false}, as it does not have a normal first token. This function is useful to implement actions on token lists on a token by token basis.

Tests if the first \textit{token} in the \textit{token list} is an explicit space character (explicit token with character code 12 and category code 10). In particular, the test is \texttt{false} if the \textit{token list} starts with an implicit token such as \textbackslash{c_space_token}, or if it is empty. This function is useful to implement actions on token lists on a token by token basis.

15.4 Working with token lists as a whole

15.4.1 Using token lists

Converts the \textit{token list} to a \textit{string}, leaving the resulting character tokens in the input stream. A \textit{string} is a series of tokens with category code 12 (other) with the exception of spaces, which retain category code 10 (space). This function requires only a single expansion. Its argument must be braced.

\texttt{\tl_to_str:n} \texttt{\tl_to_str:V} \texttt{\tl_to_str:N} \texttt{\tl_to_str:c} \texttt{\tl_to_str:V}

Converts the content of the \textit{tl var} into a series of characters with category code 12 (other) with the exception of spaces, which retain category code 10 (space). This \textit{string} is then left in the input stream. For low-level details, see the notes given for \texttt{\tl_to_str:n}.
Recovers the content of a \langle tl var \rangle and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Note that it is possible to use a \langle tl var \rangle directly without an accessor function.

### 15.4.2 Counting and reversing token lists

\begin{verbatim}
l_count:n \{\langle items \rangle\}
\end{verbatim}

Counts the number of \langle items \rangle in \langle tokens \rangle and leaves this information in the input stream. Unbraced tokens count as one element as do each token group \langle \{ \ldots \} \rangle. This process ignores any unprotected spaces within \langle tokens \rangle. See also \tl_count:n. This function requires three expansions, giving an \langle integer denotation \rangle.

\begin{verbatim}
l_count:N \langle tl var \rangle
\end{verbatim}

Counts the number of \langle items \rangle in the \langle tl var \rangle and leaves this information in the input stream. Unbraced tokens count as one element as do each token group \langle \{ \ldots \} \rangle. This process ignores any unprotected spaces within the \langle tl var \rangle. See also \tl_count:n. This function requires three expansions, giving an \langle integer denotation \rangle.

\begin{verbatim}
l_count_tokens:n \{\langle tokens \rangle\}
\end{verbatim}

Counts the number of TEX tokens in the \langle tokens \rangle and leaves this information in the input stream. Every token, including spaces and braces, contributes one to the total; thus for instance, the token count of \texttt{a~\{bc\}} is 6.

\begin{verbatim}
l_reverse:n \{\langle token list \rangle\}
\end{verbatim}

Reverses the order of the \langle items \rangle in the \langle token list \rangle, so that \langle item_1 \rangle\langle item_2 \rangle\langle item_3 \rangle \ldots \langle item_n \rangle becomes \langle item_n \rangle\ldots\langle item_2 \rangle\langle item_1 \rangle. This process preserves unprotected space within the \langle token list \rangle. Tokens are not reversed within braced token groups, which keep their outer set of braces. In situations where performance is important, consider \tl_reverse_items:n. See also \tl_reverse:n.

\textbf{\LaTeXhackers note:} The result is returned within \texttt{\unexpanded}, which means that the token list does not expand further when appearing in an \texttt{x}-type argument expansion.

\begin{verbatim}
l_reverse:N \langle tl var \rangle
\end{verbatim}

Sets the \langle tl var \rangle to contain the result of reversing the order of its \langle items \rangle, so that \langle item_1 \rangle\langle item_2 \rangle\langle item_3 \rangle \ldots \langle item_n \rangle becomes \langle item_n \rangle\ldots\langle item_2 \rangle\langle item_1 \rangle. This process preserves unprotected spaces within the \langle token list variable \rangle. Braced token groups are copied without reversing the order of tokens, but keep the outer set of braces. This is equivalent to a combination of an assignment and \tl_reverse:V. See also \tl_reverse_items:n for improved performance.
\tl_reverse_items:n \{\langle token list\rangle\} \\

Reverses the order of the \langle items\rangle stored in \langle tl var\rangle, so that \{\langle item_1\rangle\}{\langle item_2\rangle}\{\langle item_3\rangle\} \ldots \{\langle item_n\rangle\} becomes \{\langle item_n\rangle\} \ldots \{\langle item_3\rangle\}{\langle item_2\rangle}\{\langle item_1\rangle\}. This process removes any unprotected space within the \langle token list\rangle. Braced token groups are copied without reversing the order of tokens, and keep the outer set of braces. Items which are initially not braced are copied with braces in the result. In cases where preserving spaces is important, consider the slower function \tl_reverse:n.

\textbf{\TeXhackers note:} The result is returned within \unexpanded, which means that the token list does not expand further when appearing in an \textit{x}-type argument expansion.

\tl_trim_spaces:n \star \tl_trim_spaces:o \star \\

\{\langle token list\rangle\} \\

Removes any leading and trailing explicit space characters (explicit tokens with character code 32 and category code 10) from the \langle token list\rangle and leaves the result in the input stream.

\textbf{\TeXhackers note:} The result is returned within \unexpanded, which means that the token list does not expand further when appearing in an \textit{x}-type argument expansion.

\tl_trim_spaces_apply:nN \star \tl_trim_spaces_apply:oN \star \\

\{\langle token list\rangle\} \{\langle function\rangle\} \\

Removes any leading and trailing explicit space characters (explicit tokens with character code 32 and category code 10) from the \langle token list\rangle and passes the result to the \langle function\rangle as an \textit{n}-type argument.

\tl_trim_spaces:N \langle tl var\rangle \\

Sets the \langle tl var\rangle to contain the result of removing any leading and trailing explicit space characters (explicit tokens with character code 32 and category code 10) from its contents.

\section*{15.4.3 Viewing token lists}

\tl_show:N \langle tl var\rangle \\

Displays the content of the \langle tl var\rangle on the terminal.

\textbf{\TeXhackers note:} This is similar to the \TeX primitive \texttt{\show}, wrapped to a fixed number of characters per line.

\tl_show:n \{\langle token list\rangle\} \\

Displays the \langle token list\rangle on the terminal.

\textbf{\TeXhackers note:} This is similar to the \texttt{\showtokens} primitive of \TeX, wrapped to a fixed number of characters per line.
\tl_log:N \tl_log:c

Writes the content of the ⟨tl var⟩ in the log file. See also \tl_show:N which displays the result in the terminal.

\tl_log:n ⟨token list⟩

Writes the ⟨token list⟩ in the log file. See also \tl_show:n which displays the result in the terminal.

15.5 Manipulating items in token lists

15.5.1 Mapping over token lists

All mappings are done at the current group level, i.e. any local assignments made by the ⟨function⟩ or ⟨code⟩ discussed below remain in effect after the loop.

\tl_map_function:NN \tl_map_function:cn

Applies ⟨function⟩ to every ⟨item⟩ in the ⟨tl var⟩. The ⟨function⟩ receives one argument for each iteration. This may be a number of tokens if the ⟨item⟩ was stored within braces. Hence the ⟨function⟩ should anticipate receiving n-type arguments. See also \tl_map_function:nN.

\tl_map_function:nN

Applies ⟨function⟩ to every ⟨item⟩ in the ⟨token list⟩. The ⟨function⟩ receives one argument for each iteration. This may be a number of tokens if the ⟨item⟩ was stored within braces. Hence the ⟨function⟩ should anticipate receiving n-type arguments. See also \tl_map_function:NN.

\tl_map_inline:Nn \tl_map_inline:cn

Applies the ⟨inline function⟩ to every ⟨item⟩ stored within the ⟨tl var⟩. The ⟨inline function⟩ should consist of code which receives the ⟨item⟩ as #1. See also \tl_map_function:NN.

\tl_map_inline:nn

Applies the ⟨inline function⟩ to every ⟨item⟩ stored within the ⟨token list⟩. The ⟨inline function⟩ should consist of code which receives the ⟨item⟩ as #1. See also \tl_map_function:nN.
\tl_map_tokens:Nn \tl_map_tokens:cn \tl_map_tokens:nn

Analogue of \tl_map_function:NN which maps several tokens instead of a single function. The \texttt{code} receives each \texttt{item} in the \texttt{(tl var)} or in \texttt{(tokens)} as a trailing brace group. For instance,

\begin{verbatim}
\tl_map_tokens:Nn \l_my_tl \prg_replicate:nn \{ 2 \}
\end{verbatim}

expands to twice each \texttt{item} in the \texttt{(tl var)}: for each \texttt{item} in \texttt{\l_my_tl} the function \texttt{\prg_replicate:nn} receives 2 and \texttt{item} as its two arguments. The function \texttt{\tl_map_inline:Nn} is typically faster but is not expandable.

\begin{verbatim}
\tl_map_variable:NNn \tl_map_variable:cNn
Updated: 2012-06-29
\end{verbatim}

\texttt{\tl_map_variable:NNn} \texttt{\tl_map_variable:cNn}

Stores each \texttt{item} of the \texttt{(token list)} \texttt{variable} and applies the \texttt{code}. The \texttt{code} will usually make use of the \texttt{variable}, but this is not enforced. The assignments to the \texttt{variable} are local. Its value after the loop is the last \texttt{item} in the \texttt{(tl var)}, or its original value if the \texttt{(tl var)} is blank. See also \texttt{\tl_map_inline:Nn}.

\begin{verbatim}
\tl_map_break:
Updated: 2012-06-29
\end{verbatim}

\texttt{\tl_map_break}:

Used to terminate a \texttt{\tl_map...} function before all entries in the \texttt{(token list variable)} have been processed. This normally takes place within a conditional statement, for example

\begin{verbatim}
\tl_map_inline:Nn \l_my_tl
{\str_if_eq:nnT \#1 \{ bingo \} \tl_map_break: }
% Do something useful
\end{verbatim}

See also \texttt{\tl_map_break:n}. Use outside of a \texttt{\tl_map...} scenario leads to low level \TeX errors.

\textbf{\TeX hackers note:} When the mapping is broken, additional tokens may be inserted before the \texttt{tokens} are inserted into the input stream. This depends on the design of the mapping function.
\tl_map_break:n \{\textit{code}\}

Used to terminate a \tl_map_function before all entries in the \textit{token list variable} have been processed, inserting the \textit{\texttt{code}} after the mapping has ended. This normally takes place within a conditional statement, for example

\begin{verbatim}
\tl_map_inline:Nn \l_my_tl
 { \str_if_eq:nnT { #1 } { bingo } 
 { \tl_map_break:n { <code> } } 
 % Do something useful
 }
\end{verbatim}

Use outside of a \tl_map_function scenario leads to low level \TeX{} errors.

\textbf{\TeX{}hackers note:} When the mapping is broken, additional tokens may be inserted before the \textit{(code)} is inserted into the input stream. This depends on the design of the mapping function.

### 15.5.2 Head and tail of token lists

Functions which deal with either only the very first item (balanced text or single normal token) in a token list, or the remaining tokens.

\begin{verbatim}
\tl_head:n \{\textit{token list}\}
\tl_head:n \{ \textit{token list} \} \textit{V|v|f}
\end{verbatim}

Leaves in the input stream the first \textit{(item)} in the \textit{\texttt{token list}}, discarding the rest of the \textit{\texttt{token list}}. All leading explicit space characters (explicit tokens with character code 32 and category code 10) are discarded; for example

\begin{verbatim}
\tl_head:n \{ abc \}
\end{verbatim}

and

\begin{verbatim}
\tl_head:n \{ - abc \}
\end{verbatim}

both leave \texttt{a} in the input stream. If the “head” is a brace group, rather than a single token, the braces are removed, and so

\begin{verbatim}
\tl_head:n \{ - \{ - ab \} c \}
\end{verbatim}

yields \texttt{ab}. A blank \textit{\texttt{token list}} (see \texttt{\tl_if_blank:nTF}) results in \texttt{\tl_head:n} leaving nothing in the input stream.

\textbf{\TeX{}hackers note:} The result is returned within \texttt{\exp_not:n}, which means that the token list does not expand further when appearing in an \texttt{x}-type argument expansion.
Leaves in the input stream the first \texttt{item} in the \texttt{token list}, discarding the rest of the \texttt{token list}. All leading explicit space characters (explicit tokens with character code 32 and category code 10) are discarded. A blank \texttt{token list} (which consists only of space characters) results in a low-level \TeX{} error, which may be avoided by the inclusion of an empty group in the input (as shown), without the need for an explicit test. Alternatively, \texttt{\tl_if_blank:nF} may be used to avoid using the function with a “blank” argument. This function requires only a single expansion, and thus is suitable for use within an \texttt{o}-type expansion. In general, \texttt{\tl_head:n} should be preferred if the number of expansions is not critical.

Discards all leading explicit space characters (explicit tokens with character code 32 and category code 10) and the first \texttt{item} in the \texttt{token list}, and leaves the remaining tokens in the input stream. Thus for example

\begin{verbatim}
\tl_tail:n { a - {bc} d }
\end{verbatim}

and

\begin{verbatim}
\tl_tail:n { - a - {bc} d }
\end{verbatim}

both leave \((bc)d\) in the input stream. A blank \texttt{token list} (see \texttt{\tl_if_blank:nF}) results in \texttt{\tl_tail:n} leaving nothing in the input stream.

\textbf{\TeX{}hackers note}: The result is returned within \texttt{\exp_not:n}, which means that the token list does not expand further when appearing in an \texttt{x}-type argument expansion.

\subsection{Items and ranges in token lists}

Indexing items in the \texttt{token list} from 1 on the left, this function evaluates the \texttt{integer expression} and leaves the appropriate item from the \texttt{token list} in the input stream. If the \texttt{integer expression} is negative, indexing occurs from the right of the token list, starting at \(-1\) for the right-most item. If the index is out of bounds, then the function expands to nothing.

\textbf{\TeX{}hackers note}: The result is returned within the \texttt{\unexpanded} primitive (\texttt{\exp_not:n}), which means that the \texttt{item} does not expand further when appearing in an \texttt{x}-type argument expansion.
\tl_rand_item:N \tl_rand_item:c \tl_rand_item:n

Selects a pseudo-random item of the \textit{token list}. If the \textit{token list} is blank, the result is empty. This is not available in older versions of \LaTeX{}.

\textbf{TeXhackers note:} The result is returned within the \texttt{\unexpanded primitive \exp_not:n}, which means that the \textit{item} does not expand further when appearing in an $\times$-type argument expansion.
Leaves in the input stream the items from the ⟨start index⟩ to the ⟨end index⟩ inclusive. Spaces and braces are preserved between the items returned (but never at either end of the list). Here ⟨start index⟩ and ⟨end index⟩ should be ⟨integer expressions⟩. For describing in detail the functions’ behavior, let m and n be the start and end index respectively. If either is 0, the result is empty. A positive index means ‘start counting from the left end’, and a negative index means ‘from the right end’. Let l be the count of the token list.

The actual start point is determined as \( M = m \) if \( m > 0 \) and as \( M = l + m + 1 \) if \( m < 0 \). Similarly the actual end point is \( N = n \) if \( n > 0 \) and \( N = l + n + 1 \) if \( n < 0 \). If \( M > N \), the result is empty. Otherwise it consists of all items from position \( M \) to position \( N \) inclusive; for the purpose of this rule, we can imagine that the token list extends at infinity on either side, with void items at positions \( s \) for \( s \leq 0 \) or \( s > l \).

Spaces in between items in the actual range are preserved. Spaces at either end of the token list will be removed anyway (think to the token list being passed to \tl_trim_spaces:n to begin with.

Thus, with \( l = 7 \) as in the examples below, all of the following are equivalent and result in the whole token list

\[
\begin{align*}
\text{\texttt{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ 1 \} \{ 7 \}}} \\
\text{\texttt{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ 1 \} \{ 12 \}}} \\
\text{\texttt{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ -7 \} \{ 7 \}}} \\
\text{\texttt{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ -12 \} \{ 7 \}}}
\end{align*}
\]

Here are some more interesting examples. The calls

\[
\begin{align*}
\text{\texttt{\iow_term:x \{ \text{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ 2 \} \{ 5 \}} \}}} \\
\text{\texttt{\iow_term:x \{ \text{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ 2 \} \{ -3 \}} \}}} \\
\text{\texttt{\iow_term:x \{ \text{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ -6 \} \{ 5 \}} \}}} \\
\text{\texttt{\iow_term:x \{ \text{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ -6 \} \{ -3 \}} \}}}
\end{align*}
\]

are all equivalent and will print bcd\{e{}\} on the terminal; similarly

\[
\begin{align*}
\text{\texttt{\iow_term:x \{ \text{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ 2 \} \{ 5 \}} \}}} \\
\text{\texttt{\iow_term:x \{ \text{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ 2 \} \{ -3 \}} \}}} \\
\text{\texttt{\iow_term:x \{ \text{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ -6 \} \{ 5 \}} \}}} \\
\text{\texttt{\iow_term:x \{ \text{\tl_range:nnn \{ \text{abcd{-e{}}fg} \} \{ -6 \} \{ -3 \}} \}}}
\end{align*}
\]

are all equivalent and will print bcd \{e{\}} on the terminal (note the space in the middle). To the contrary,

\[
\text{\texttt{\tl_range:nnn \{ \text{abcd{-e{}}f} \} \{ 2 \} \{ 4 \}}}
\]

will discard the space after ‘d’.

If we want to get the items from, say, the third to the last in a token list <tl>, the call is \texttt{\tl_range:nnn \{ <tl> \} \{ 3 \} \{ -1 \}}. Similarly, for discarding the last item, we can do \texttt{\tl_range:nnn \{ <tl> \} \{ 1 \} \{ -2 \}}.

For better performance, see \texttt{\tl_range_braced:nnn} and \texttt{\tl_range_unbraced:nnn}.

\textbf{\TeXhackers note:} The result is returned within the \texttt{\unexpanded} primitive \texttt{\exp_not:n}, which means that the ⟨item⟩ does not expand further when appearing in an x-type argument expansion.
15.5.4 Sorting token lists

\tl_sort:Nn \tl_sort:cn
\tl_gsort:Nn \tl_gsort:cn

Sorts the items in the \((tl\ var)\) according to the \((comparison\ code)\), and assigns the result to \((tl\ var)\). The details of sorting comparison are described in Section 6.1.

New: 2017-02-06

\tl_sort:nN *

Sorts the items in the \((token\ list)\), using the \((conditional)\) to compare items, and leaves the result in the input stream. The \((conditional)\) should have signature :nnTF, and return true if the two items being compared should be left in the same order, and false if the items should be swapped. The details of sorting comparison are described in Section 6.1.

\TeX{hackers note}: The result is returned within \exp_not:n, which means that the token list does not expand further when appearing in an \texttt{x-type} or \texttt{e-type} argument expansion.

15.6 Manipulating tokens in token lists

15.6.1 Replacing tokens

\tl_replace_once:Nnn \tl_replace_once:cn
\tl_greplace_once:Nnn \tl_greplace_once:cn

Replaces the first (leftmost) occurrence of \((old\ tokens)\) in the \((tl\ var)\) with \((new\ tokens)\). \((Old\ tokens)\) cannot contain \{, \} or \# (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6).

\tl_replace_all:Nnn \tl_replace_all:cn
\tl_greplace_all:Nnn \tl_greplace_all:cn

Replaces all occurrences of \((old\ tokens)\) in the \((tl\ var)\) with \((new\ tokens)\). \((Old\ tokens)\) cannot contain \{, \} or \# (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6). As this function operates from left to right, the pattern \((old\ tokens)\) may remain after the replacement (see \tl_remove_all:Nn for an example).

\tl_remove_once:Nn \tl_remove_once:cn
\tl_gremove_once:Nn \tl_gremove_once:cn

Removes the first (leftmost) occurrence of \((tokens)\) from the \((tl\ var)\). \((Tokens)\) cannot contain \{, \} or \# (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6).
\tl_remove_all:Nn \tl_remove_all:cn \tl_gremove_all:Nn \tl_gremove_all:cn

Updated: 2011-08-11

15.6.2 Reassigning category codes

These functions allow the rescanning of tokens: re-apply \TeX{}’s tokenization process to apply category codes different from those in force when the tokens were absorbed. Whilst this functionality is supported, it is often preferable to find alternative approaches to achieving outcomes rather than rescanning tokens (for example construction of token lists token-by-token with intervening category code changes or using \char_generate:nn).

\tl_set_rescan:Nnn \tl_set_rescan:Nn \tl_gset_rescan:Nnn \tl_gset_rescan:Nn

Updated: 2015-08-11

Sets \langle tl var \rangle to contain \langle tokens \rangle, applying the category code régime specified in the \langle setup \rangle before carrying out the assignment. (Category codes applied to tokens not explicitly covered by the \langle setup \rangle are those in force at the point of use of \tl_set_rescan:Nnn.) This allows the \langle tl var \rangle to contain material with category codes other than those that apply when \langle tokens \rangle are absorbed. The \langle setup \rangle is run within a group and may contain any valid input, although only changes in category codes, such as uses of \cctab_select:N, are relevant. See also \tl_rescan:nn.

\TeX{}hackers note: The \langle tokens \rangle are first turned into a string (using \tl_to_str:n). If the string contains one or more characters with character code \newlinechar (set equal to \endlinechar unless that is equal to 32, before the user \langle setup \rangle), then it is split into lines at these characters, then read as if reading multiple lines from a file, ignoring spaces (catcode 10) at the beginning and spaces and tabs (character code 32 or 9) at the end of every line. Otherwise, spaces (and tabs) are retained at both ends of the single-line string, as if it appeared in the middle of a line read from a file.
\tl_rescan:nn \tl_rescan:nn \{\langle setup\rangle\} \{\langle tokens\rangle\}

Rescans \langle tokens\rangle applying the category code régime specified in the \langle setup\rangle, and leaves the resulting tokens in the input stream. (Category codes applied to tokens not explicitly covered by the \langle setup\rangle are those in force at the point of use of \tl_rescan:nn.) The \langle setup\rangle is run within a group and may contain any valid input, although only changes in category codes, such as uses of \cctab_select:N, are relevant. See also \tl_set_rescan:NNn, which is more robust than using \tl_set:Nn in the \langle tokens\rangle argument of \tl_rescan:nn.

\textbf{\TeXhackers note:} The\langle tokens\rangle are first turned into a string (using \tl_to_str:n). If the string contains one or more characters with character code \newlinechar (set equal to \endlinechar unless that is equal to 32, before the user \langle setup\rangle), then it is split into lines at these characters, then read as if reading multiple lines from a file, ignoring spaces (catcode 10) at the beginning and spaces and tabs (character code 32 or 9) at the end of every line. Otherwise, spaces (and tabs) are retained at both ends of the single-line string, as if it appeared in the middle of a line read from a file.

Contrarily to the \scantokens primitive, \tl_rescan:nn tokenizes the whole string in the same category code regime rather than one token at a time, so that directives such as \verb that rely on changing category codes will not function properly.

## 15.7 Constant token lists

\c_empty_tl Constant that is always empty.

\c_novaluel_tl A marker for the absence of an argument. This constant tbl can safely be typeset \textit{(cf.} \c_n愈nil\textit{)}, with the result being \textit{-NoValue-}. It is important to note that \c_novaluel_tl is constructed such that it will \textit{not} match the simple text input \textit{-NoValue-}, \textit{i.e.} that

\tl_if_eq:NnTF \c_novaluel_tl { -NoValue- } \textit{is logically false}. The \c_novaluel_tl marker is intended for use in creating document-level interfaces, where it serves as an indicator that an (optional) argument was omitted. In particular, it is distinct from a simple empty tbl.

\c_space_tl An explicit space character contained in a token list \textit{(compare this with \c_space_token)}. For use where an explicit space is required.

## 15.8 Scratch token lists

\l_tmpa_tl \l_tmpb_tl Scratch token lists for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
\texttt{\_g\_tmpa\_tl} Scratch token lists for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Chapter 16

The \texttt{l3str} package: Strings

\TeX{} associates each character with a category code: as such, there is no concept of a “string” as commonly understood in many other programming languages. However, there are places where we wish to manipulate token lists while in some sense “ignoring” category codes: this is done by treating token lists as strings in a \TeX{} sense.

A \TeX{} string (and thus an \texttt{expl3} string) is a series of characters which have category code 12 (“other”) with the exception of space characters which have category code 10 (“space”). Thus at a technical level, a \TeX{} string is a token list with the appropriate category codes. In this documentation, these are simply referred to as strings.

String variables are simply specialised token lists, but by convention should be named with the suffix \texttt{...str}. Such variables should contain characters with category code 12 (other), except spaces, which have category code 10 (blank space). All the functions in this module which accept a token list argument first convert it to a string using \texttt{\tl_to_\texttt{str}:n} for internal processing, and do not treat a token list or the corresponding string representation differently.

As a string is a subset of the more general token list, it is sometimes unclear when one should be used over the other. Use a string variable for data that isn’t primarily intended for typesetting and for which a level of protection from unwanted expansion is suitable. This data type simplifies comparison of variables since there are no concerns about expansion of their contents.

The functions \texttt{\cs_to_str:N}, \texttt{\tl_to_str:N}, \texttt{\tl_to_str:n} and \texttt{\token_to_str:N} (and variants) generate strings from the appropriate input: these are documented in \texttt{l3basics}, \texttt{l3tl} and \texttt{l3token}, respectively.

Most expandable functions in this module come in three flavours:

- \texttt{\str_:N}, which expect a token list or string variable as their argument;
- \texttt{\str_:n}, taking any token list (or string) as an argument;
- \texttt{\str_:ignore_spaces:n}, which ignores any space encountered during the operation: these functions are typically faster than those which take care of escaping spaces appropriately.
16.1 Creating and initialising string variables

\str_new:N (\str var)
Creates a new (\str var) or raises an error if the name is already taken. The declaration
is global. The (\str var) is initially empty.

\str_new:c
New: 2015-09-18
Updated: 2018-07-28

\str_const:Nn (\str var) {\{token list\}}
Creates a new constant (\str var) or raises an error if the name is already taken. The
value of the (\str var) is set globally to the \{token list\}, converted to a string.

\str_const:Nn \str_const:
(NV|Nz|cn|cV|cx)
New: 2015-09-18
Updated: 2018-07-28

\str_clear:N (\str var)
Clears the content of the (\str var).
\str_clear:c
\str_gclear:N
\str_gclear:c
New: 2015-09-18

\str_clear_new:N (\str var)
Ensures that the (\str var) exists globally by applying \str_new:N if necessary, then
applies \str(g)clear:N to leave the (\str var) empty.
\str_clear_new:c
New: 2015-09-18

\str_set_eq:NN (\str var 1) (\str var 2)
Sets the content of (\str var 1) equal to that of (\str var 2).
\str_set_eq:NN \str_set_eq:c
(Nc|Nc|cc)
\str_gset_eq:NN \str_gset_eq:c
(Nc|Nc|cc)
New: 2015-09-18

\str_concat:NNN (\str var 1) (\str var 2) (\str var 3)
Concatenates the content of (\str var 2) and (\str var 3) together and saves the result in
(\str var 1). The (\str var 2) is placed at the left side of the new string variable. The
(\str var 2) and (\str var 3) must indeed be strings, as this function does not convert their
contents to a string.
\str_concat:NNN \str_concat:ccc
\str_gconcat:NNN \str_gconcat:ccc
New: 2017-10-08

\str_if_exist_p:N * (\str var)
\str_if_exist_p:c * \str_if_exist:NTF
\str_if_exist:CTF *
New: 2015-09-18

Tests whether the (\str var) is currently defined. This does not check that the (\str var)
really is a string.
16.2 Adding data to string variables

\str_set:Nn \str_set:(NV|Nx|cn|cV|cx) \str_set:Nn \str_set:(NV|Nx|cn|cV|cx)
New: 2015-09-18
Updated: 2018-07-28

Converts the ⟨token list⟩ to a ⟨string⟩, and stores the result in ⟨str var⟩.

\str_gset:Nn \str_gset:(NV|Nx|cn|cV|cx)
New: 2015-09-18
Updated: 2018-07-28

Converts the ⟨token list⟩ to a ⟨string⟩, and stores the result in ⟨str var⟩.

\str_put_left:Nn \str_put_left:(NV|Nx|cn|cV|cx) \str_put_left:Nn \str_put_left:(NV|Nx|cn|cV|cx)
New: 2015-09-18
Updated: 2018-07-28

Converts the ⟨token list⟩ to a ⟨string⟩, and prepends the result to ⟨str var⟩. The current contents of the ⟨str var⟩ are not automatically converted to a string.

\str_gput_left:Nn \str_gput_left:(NV|Nx|cn|cV|cx)
New: 2015-09-18
Updated: 2018-07-28

Converts the ⟨token list⟩ to a ⟨string⟩, and prepends the result to ⟨str var⟩. The current contents of the ⟨str var⟩ are not automatically converted to a string.

\str_put_right:Nn \str_put_right:(NV|Nx|cn|cV|cx) \str_put_right:Nn \str_put_right:(NV|Nx|cn|cV|cx)
New: 2015-09-18
Updated: 2018-07-28

Converts the ⟨token list⟩ to a ⟨string⟩, and appends the result to ⟨str var⟩. The current contents of the ⟨str var⟩ are not automatically converted to a string.

\str_gput_right:Nn \str_gput_right:(NV|Nx|cn|cV|cx)
New: 2015-09-18
Updated: 2018-07-28

Converts the ⟨token list⟩ to a ⟨string⟩, and appends the result to ⟨str var⟩. The current contents of the ⟨str var⟩ are not automatically converted to a string.

16.3 String conditionals

\str_if_empty_p:N \str_if_empty_p:c \str_if_empty:N \str_if_empty:c
New: 2015-09-18

Tests if the ⟨string variable⟩ is entirely empty (i.e. contains no characters at all).

\str_if_eq_p:NN \str_if_eq_p:cN \str_if_eq:NN \str_if_eq:cN
New: 2015-09-18

Compares the content of two ⟨str variables⟩ and is logically true if the two contain the same characters in the same order. See \tl\_if\_eq:NNTF to compare tokens (including their category codes) rather than characters.
\str_if_eq_p:nn \str_if_eq_p:{Vn|on|no|VV|vn|nv|ee} \str_if_eq:nnTF \str_if_eq:{Vn|on|no|nV|VV|vn|nv|ee}TF Updated: 2018-06-18

Compares the two \textit{token lists} on a character by character basis (namely after converting them to strings), and is \texttt{true} if the two \textit{strings} contain the same characters in the same order. Thus for example

\str_if_eq_p:no \{ abc \} \{ \tl_to_str:n \{ abc \} \}

is logically \texttt{true}. See \tl_if_eq:nnTF to compare tokens (including their category codes) rather than characters.

\str_if_in:NnTF \str_if_in:cn TF \str_if_in:nTF \str_if_in:nnTF Updated: 2017-10-08

Converts both \textit{token lists} to \textit{strings} and tests whether \textit{string2} is found inside \textit{string1}.

\str_case:nn \str_case:{Vn|on|nV|nv} \str_case:nnTF \str_case:{Vn|on|nV|nv}TF
\str_case:nn Updated: 2015-02-28

Compares the \textit{test string} in turn with each of the \textit{string cases} (all token lists are converted to strings). If the two are equal (as described for \str_if_eq:nnTF) then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \texttt{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \texttt{false code} is inserted. The function \str_case:nn, which does nothing if there is no match, is also available.
\begin{verbatim}
\str_case_e:nn \str_case_e:nnTF \{⟨test string⟩\}
\{ ⟨string case1⟩ ⟨code case1⟩ \}
\{ ⟨string case2⟩ ⟨code case2⟩ \}
\ldots
\{ ⟨string case_n⟩ ⟨code case_n⟩ \}
\{ ⟨true code⟩ \}
\{ ⟨false code⟩ \}
\end{verbatim}

Compares the full expansion of the \textit{(test string)} in turn with the full expansion of the \textit{(string cases)} (all token lists are converted to strings). If the two full expansions are equal (as described for \str_if_eq:nnTF) then the associated \textit{(code)} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{(true code)} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{(false code)} is inserted. The function \str_case_e:nn, which does nothing if there is no match, is also available. The \textit{(test string)} is expanded in each comparison, and must always yield the same result: for example, random numbers must not be used within this string.

\begin{verbatim}
\str_compare_p:nNn \str_compare_p:eNe \str_compare:nNnTF \str_compare:eNeTF
\end{verbatim}

Compares the two \textit{(token lists)} on a character by character basis (namely after converting them to strings) in a lexicographic order according to the character codes of the characters. The \textit{(relation)} can be <, =, or > and the test is \texttt{true} under the following conditions:

- for <, if the first string is earlier than the second in lexicographic order;
- for =, if the two strings have exactly the same characters;
- for >, if the first string is later than the second in lexicographic order.

Thus for example the following is logically \texttt{true}:

\begin{verbatim}
\str_compare_p:nNn \{ ab \} < \{ abc \}
\end{verbatim}

\TeX{}hackers note: This is a wrapper around the \TeX{} primitive \texttt{(pdf)strcmp}. It is meant for programming and not for sorting textual contents, as it simply considers character codes and not more elaborate considerations of grapheme clusters, locale, etc.

### 16.4 Mapping over strings

All mappings are done at the current group level, \textit{i.e.} any local assignments made by the \textit{(function)} or \textit{(code)} discussed below remain in effect after the loop.

\begin{verbatim}
\str_map_function:nN \str_map_function:NN \str_map_function:cN
\end{verbatim}

Converts the \textit{(token list)} to a \textit{(string)} then applies \textit{(function)} to every \textit{(character)} in the \textit{(string)} including spaces.
\str_map_inline:nn \str_map_inline:Nn \str_map_inline:cn

Converting the \langle token list \rangle to a \langle string \rangle then applies the \langle inline function \rangle to every \langle character \rangle in the \langle str var \rangle including spaces. The \langle inline function \rangle should consist of code which receives the \langle character \rangle as #1.

\str_map_tokens:nn \str_map_tokens:Nn \str_map_tokens:cn

Converting the \langle token list \rangle to a \langle string \rangle then applies \langle code \rangle to every \langle character \rangle in the \langle string \rangle including spaces. The \langle code \rangle receives each character as a trailing brace group. This is equivalent to \str_map_function:nN if the \langle code \rangle consists of a single function.

\str_map_variable:nNn \str_map_variable:NNn \str_map_variable:cNn

Converting the \langle token list \rangle to a \langle string \rangle then stores each \langle character \rangle in the \langle string \rangle (including spaces) in turn in the (string or token list) \langle variable \rangle and applies the \langle code \rangle. The \langle code \rangle will usually make use of the \langle variable \rangle, but this is not enforced. The assignments to the \langle variable \rangle are local. Its value after the loop is the last \langle character \rangle in the \langle string \rangle, or its original value if the \langle string \rangle is empty. See also \str_map_inline:Nn.

\str_map_break: \str_map_break:

Used to terminate a \str_map_... function before all characters in the \langle string \rangle have been processed. This normally takes place within a conditional statement, for example

\str_map_inline:Nn \_my_str
{\str_if_eq:nnT { #1 } { bingo } { \str_map_break: }
% Do something useful
}

See also \str_map_break:n. Use outside of a \str_map_... scenario leads to low level \TeX{} errors.

\TeXhackers_note: When the mapping is broken, additional tokens may be inserted before continuing with the code that follows the loop. This depends on the design of the mapping function.
\str_map_break:n \{\langle code\rangle\}

Used to terminate a \str_map: function before all characters in the \langle string\rangle have been processed, inserting the \langle code\rangle after the mapping has ended. This normally takes place within a conditional statement, for example

\str_map_inline:Nn \l_my_str
{\str_if_eq:nnT { #1 } { bingo }{ \str_map_break:n { <code> } }%
% Do something useful}

Use outside of a \str_map: scenario leads to low level \TeX{} errors.

\TeX{}hackers note: When the mapping is broken, additional tokens may be inserted before the \langle code\rangle is inserted into the input stream. This depends on the design of the mapping function.

\section{16.5 Working with the content of strings}

\str_use:N \str_use:c

Recovers the content of a \langle str var\rangle and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Note that it is possible to use a \langle str\rangle directly without an accessor function.

\str_count:n \str_count:n \str_count:n \str_count:n

Leaves in the input stream the number of characters in the string representation of \langle token list\rangle, as an integer denotation. The functions differ in their treatment of spaces. In the case of \str_count:N and \str_count:n, all characters including spaces are counted. The \str_count_ignore_spaces:n function leaves the number of non-space characters in the input stream.

\str_count_spaces:n \str_count_spaces:n \str_count_spaces:n

Leaves in the input stream the number of space characters in the string representation of \langle token list\rangle, as an integer denotation. Of course, this function has no \_ignore_spaces variant.
Converts the \langle token list⟩ into a \langle string⟩. The first character in the \langle string⟩ is then left in the input stream, with category code “other”. The functions differ if the first character is a space: \str_head:N and \str_head:n return a space token with category code 10 (blank space), while the \str_head_ignore_spaces:n function ignores this space character and leaves the first non-space character in the input stream. If the \langle string⟩ is empty (or only contains spaces in the case of the _ignore_spaces function), then nothing is left on the input stream.

Converts the \langle token list⟩ to a \langle string⟩, removes the first character, and leaves the remaining characters (if any) in the input stream, with category codes 12 and 10 (for spaces). The functions differ in the case where the first character is a space: \str_tail:N and \str_tail:n only trim that space, while \str_tail_ignore_spaces:n removes the first non-space character and any space before it. If the \langle token list⟩ is empty (or blank in the case of the _ignore_spaces variant), then nothing is left on the input stream.

Converts the \langle token list⟩ to a \langle string⟩, and leaves in the input stream the character in position \langle integer expression⟩ of the \langle string⟩, starting at 1 for the first (left-most) character. In the case of \str_item:Nn and \str_item:nn, all characters including spaces are taken into account. The \str_item_ignore_spaces:nn function skips spaces when counting characters. If the \langle integer expression⟩ is negative, characters are counted from the end of the \langle string⟩. Hence, −1 is the right-most character, etc.
\str_range:Nnn \str_range:nnn \{\text{token list}\} \{\text{start index}\} \{\text{end index}\}
\str_range:nnn \star \str_range:nnn \star \str_range:nnn \star \str_range:nnn \star \str_range_ignore_spaces:nnn \star

Converts the \textit{token list} to a \textit{string}, and leaves in the input stream the characters from the \textit{start index} to the \textit{end index} inclusive. Spaces are preserved and counted as items (contrast this with \tl_range:nnn where spaces are not counted as items and are possibly discarded from the output).

Here \textit{start index} and \textit{end index} should be integer denotations. For describing in detail the functions’ behavior, let \(m\) and \(n\) be the start and end index respectively. If either is 0, the result is empty. A positive index means ‘start counting from the left end’, a negative index means ‘start counting from the right end’. Let \(l\) be the count of the token list.

The \textit{actual start point} is determined as \(M = m\) if \(m > 0\) and as \(M = l + m + 1\) if \(m < 0\). Similarly the \textit{actual end point} is \(N = n\) if \(n > 0\) and \(N = l + n + 1\) if \(n < 0\). If \(M > N\), the result is empty. Otherwise it consists of all items from position \(M\) to position \(N\) inclusive; for the purpose of this rule, we can imagine that the token list extends at infinity on either side, with void items at positions \(s\) for \(s \leq 0\) or \(s > l\). For instance,

\begin{verbatim}
\iow_term:x { \str_range:nnn { abcdef } { 2 } { 5 } }
\iow_term:x { \str_range:nnn { abcdef } { -4 } { -1 } }
\iow_term:x { \str_range:nnn { abcdef } { -2 } { -1 } }
\iow_term:x { \str_range:nnn { abcdef } { 0 } { -1 } }
\end{verbatim}

prints \textit{bcde}, \textit{cdef}, \textit{ef}, and an empty line to the terminal. The \textit{start index} must always be smaller than or equal to the \textit{end index}: if this is not the case then no output is generated. Thus

\begin{verbatim}
\iow_term:x { \str_range:nnn { abcdef } { 5 } { 2 } }
\iow_term:x { \str_range:nnn { abcdef } { -1 } { -4 } }
\end{verbatim}

both yield empty strings.

The behavior of \texttt{\str_range_ignore_spaces:nnn} is similar, but spaces are removed before starting the job. The input

\begin{verbatim}
\iow_term:x { \str_range:nnn { abcdefg } { 2 } { 5 } }
\iow_term:x { \str_range:nnn { abcdefg } { 2 } { -3 } }
\iow_term:x { \str_range:nnn { abcdefg } { -6 } { 5 } }
\iow_term:x { \str_range:nnn { abcdefg } { -6 } { -3 } }
\iow_term:x { \str_range:nnn { abc-efg } { 2 } { 5 } }
\iow_term:x { \str_range:nnn { abc-efg } { 2 } { -3 } }
\iow_term:x { \str_range:nnn { abc-efg } { -6 } { 5 } }
\iow_term:x { \str_range:nnn { abc-efg } { -6 } { -3 } }
\iow_term:x { \str_range_ignore_spaces:nnn { abcdefg } { 2 } { 5 } }
\iow_term:x { \str_range_ignore_spaces:nnn { abcdefg } { 2 } { -3 } }
\iow_term:x { \str_range_ignore_spaces:nnn { abcdefg } { -6 } { 5 } }
\iow_term:x { \str_range_ignore_spaces:nnn { abcdefg } { -6 } { -3 } }
\end{verbatim}

127
will print four instances of bcde, four instances of be and eight instances of bcde.

16.6 Modifying string variables

\str_replace_once:Nnn \str_replace_once:cnn \str_greplace_once:Nnn \str_greplace_once:cnn

New: 2017-10-08

\str_replace_all:Nnn \str_replace_all:cnn \str_greplace_all:Nnn \str_greplace_all:cnn

New: 2017-10-08

\str_remove_once:Nn \str_remove_once:cn \str_gremove_once:Nn \str_gremove_once:cn

New: 2017-10-08

\str_remove_all:Nn \str_remove_all:cn \str_gremove_all:Nn \str_gremove_all:cn

New: 2017-10-08

\str_set:Nn \l_tmpa_str {abbccd} \str_remove_all:Nn \l_tmpa_str \{bc\}

results in \l_tmpa_str containing abcd.
16.7 String manipulation

\str_lowercase:n \str_lowercase:f \str_uppercase:n \str_uppercase:f

Converts the input (tokens) to their string representation, as described for \tl_to_str:n, and then to the lower or upper case representation using a one-to-one mapping as described by the Unicode Consortium file UnicodeData.txt.

These functions are intended for case changing programmatic data in places where upper/lower case distinctions are meaningful. One example would be automatically generating a function name from user input where some case changing is needed. In this situation the input is programmatic, not textual, case does have meaning and a language-independent one-to-one mapping is appropriate. For example

\cs_new_protected:Npn \myfunc:nn #1#2
{\cs_set_protected:cpn
 {user \str_uppercase:f \tl_head:n {#1}} \str_lowercase:f \tl_tail:n {#1}
 }
{#2}
}

would be used to generate a function with an auto-generated name consisting of the upper case equivalent of the supplied name followed by the lower case equivalent of the rest of the input.

These functions should not be used for

- Caseless comparisons: use \str_foldcase:n for this situation (case folding is distinct from lower casing).
- Case changing text for typesetting: see the \text_lowercase:n(n), \text_uppercase:n(n) and \text_titlecase:n(n) functions which correctly deal with context-dependence and other factors appropriate to text case changing.

\TeXhackers note: As with all expl3 functions, the input supported by \str_foldcase:n is engine-native characters which are or interoperate with UTF-8. As such, when used with pd\TeX only the Latin alphabet characters A–Z are case-folded (i.e. the ASCII range which coincides with UTF-8). Full UTF-8 support is available with both Xe\TeX and Lua\TeX.
\str_foldcase:n \langle tokens \rangle
\str_foldcase:v

Converts the input \langle tokens \rangle to their string representation, as described for $\texttt{tl_to_str:n}$, and then folds the case of the resulting \langle string \rangle to remove case information. The result of this process is left in the input stream.

String folding is a process used for material such as identifiers rather than for “text”. The folding provided by $\texttt{str_foldcase:n}$ follows the mappings provided by the Unicode Consortium, who state:

Case folding is primarily used for caseless comparison of text, such as identifiers in a computer program, rather than actual text transformation. Case folding in Unicode is based on the lowercase mapping, but includes additional changes to the source text to help make it language-insensitive and consistent. As a result, case-folded text should be used solely for internal processing and generally should not be stored or displayed to the end user.

The folding approach implemented by $\texttt{str_foldcase:n}$ follows the “full” scheme defined by the Unicode Consortium (e.g. SSfolds to SS). As case-folding is a language-insensitive process, there is no special treatment of Turkic input (i.e. $I$ always folds to $i$ and not to $ı$).

\texttt{TeXhacker} note: As with all \texttt{expl3} functions, the input supported by $\texttt{str_foldcase:n}$ is engine-native characters which are or interoperate with UTF-8. As such, when used with pdf\TeX, only the Latin alphabet characters A–Z are case-folded (i.e. the ASCII range which coincides with UTF-8). Full UTF-8 support is available with both X\TeX{} and Lua\TeX{}, subject only to the fact that X\TeX{} in particular has issues with characters of code above hexadecimal 0xFFFF when interacting with $\texttt{tl_to_str:n}$.

### 16.8 Viewing strings

\texttt{str_show:N \langle str var \rangle}
\texttt{str_show:c}
\texttt{str_show:n}

Displays the content of the \langle str var \rangle on the terminal.

\texttt{str_log:N \langle str var \rangle}
\texttt{str_log:c}
\texttt{str_log:n}

Writes the content of the \langle str var \rangle in the log file.
### 16.9 Constant strings

Constant strings, containing a single character token, with category code 12.

- `\c_ampersand_str`
- `\c_atsign_str`
- `\c_backslash_str`
- `\c_left_brace_str`
- `\c_right_brace_str`
- `\c_circumflex_str`
- `\c_colon_str`
- `\c_dollar_str`
- `\c_hash_str`
- `\c_percent_str`
- `\c_tilde_str`
- `\c_underscore_str`
- `\c_zero_str`

**New:** 2015-09-19  
**Updated:** 2020-12-22

### 16.10 Scratch strings

Scratch strings for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

- `\l_tmpa_str`
- `\l_tmpb_str`

Scratch strings for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

- `\g_tmpa_str`
- `\g_tmpb_str`
Chapter 17

The \texttt{l3str-convert} package: string encoding conversions

17.1 Encoding and escaping schemes

Traditionally, string encodings only specify how strings of characters should be stored as bytes. However, the resulting lists of bytes are often to be used in contexts where only a restricted subset of bytes are permitted (e.g., PDF string objects, URLs). Hence, storing a string of characters is done in two steps.

- The code points (“character codes”) are expressed as bytes following a given “encoding”. This can be \texttt{utf-16}, \texttt{iso 8859-1}, \textit{etc}. See Table 1 for a list of supported encodings.\footnote{Encodings and escapings will be added as they are requested.}

- Bytes are translated to \LaTeX\ tokens through a given “escaping”. Those are defined for the most part by the \texttt{pdf} file format. See Table 2 for a list of escaping methods supported.\footnote{Encodings and escapings will be added as they are requested.}
Table 1: Supported encodings. Non-alphanumeric characters are ignored, and capital letters are lower-cased before searching for the encoding in this list.

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>utf8</td>
<td>UTF-8</td>
</tr>
<tr>
<td>utf16</td>
<td>UTF-16, with byte-order mark</td>
</tr>
<tr>
<td>utf16be</td>
<td>UTF-16, big-endian</td>
</tr>
<tr>
<td>utf16le</td>
<td>UTF-16, little-endian</td>
</tr>
<tr>
<td>utf32</td>
<td>UTF-32, with byte-order mark</td>
</tr>
<tr>
<td>utf32be</td>
<td>UTF-32, big-endian</td>
</tr>
<tr>
<td>utf32le</td>
<td>UTF-32, little-endian</td>
</tr>
<tr>
<td>iso88591, latin1</td>
<td>ISO 8859-1</td>
</tr>
<tr>
<td>iso88592, latin2</td>
<td>ISO 8859-2</td>
</tr>
<tr>
<td>iso88593, latin3</td>
<td>ISO 8859-3</td>
</tr>
<tr>
<td>iso88594, latin4</td>
<td>ISO 8859-4</td>
</tr>
<tr>
<td>iso88595</td>
<td>ISO 8859-5</td>
</tr>
<tr>
<td>iso88596</td>
<td>ISO 8859-6</td>
</tr>
<tr>
<td>iso88597</td>
<td>ISO 8859-7</td>
</tr>
<tr>
<td>iso88598</td>
<td>ISO 8859-8</td>
</tr>
<tr>
<td>iso88599, latin5</td>
<td>ISO 8859-9</td>
</tr>
<tr>
<td>iso885910, latin6</td>
<td>ISO 8859-10</td>
</tr>
<tr>
<td>iso885911</td>
<td>ISO 8859-11</td>
</tr>
<tr>
<td>iso885913, latin7</td>
<td>ISO 8859-13</td>
</tr>
<tr>
<td>iso885914, latin8</td>
<td>ISO 8859-14</td>
</tr>
<tr>
<td>iso885915, latin9</td>
<td>ISO 8859-15</td>
</tr>
<tr>
<td>iso885916, latin10</td>
<td>ISO 8859-16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cllist</td>
<td>Comma-list of integers</td>
</tr>
<tr>
<td>⟨empty⟩</td>
<td>Native (Unicode) string</td>
</tr>
<tr>
<td>default</td>
<td>Like utf8 with 8-bit engines, and like native with unicode-engines</td>
</tr>
</tbody>
</table>

Table 2: Supported escapings. Non-alphanumeric characters are ignored, and capital letters are lower-cased before searching for the escaping in this list.

<table>
<thead>
<tr>
<th>Escaping</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bytes, or empty</td>
<td>Arbitrary bytes</td>
</tr>
<tr>
<td>hex, hexadecimal name</td>
<td>Byte = two hexadecimal digits, see \pdfescauname</td>
</tr>
<tr>
<td>string</td>
<td>See \pdfescapestring</td>
</tr>
<tr>
<td>url</td>
<td>Encoding used in URLs</td>
</tr>
</tbody>
</table>
17.2 Conversion functions

\texttt{\str_set_convert:Nnnn} \str_set_convert:Nnnn \str_gset_convert:Nnnn

This function converts the \langle string \rangle from the encoding given by \langle name 1 \rangle to the encoding given by \langle name 2 \rangle, and stores the result in the \langle str var \rangle. Each \langle name \rangle can have the form \langle encoding \rangle or \langle encoding \rangle/\langle escaping \rangle, where the possible values of \langle encoding \rangle and \langle escaping \rangle are given in Tables 1 and 2, respectively. The default escaping is to input and output bytes directly. The special case of an empty \langle name \rangle indicates the use of “native” strings, 8-bit for pdfT\TeX, and Unicode strings for the other two engines.

For example,

\texttt{\str_set_convert:Nnnn } \texttt{\l\_foo_str} \texttt{ \{ Hello! \} \{ \} \{ utf16/hex \}}

results in the variable \texttt{\l\_foo_str} holding the string FEFF00480065006C006C006F0021. This is obtained by converting each character in the (native) string Hello! to the UTF-16 encoding, and expressing each byte as a pair of hexadecimal digits. Note the presence of a (big-endian) byte order mark “FEFF, which can be avoided by specifying the encoding utf16be/hex.

An error is raised if the \langle string \rangle is not valid according to the \langle escaping 1 \rangle and \langle encoding 1 \rangle, or if it cannot be reencoded in the \langle encoding 2 \rangle and \langle escaping 2 \rangle (for instance, if a character does not exist in the \langle encoding 2 \rangle). Erroneous input is replaced by the Unicode replacement character \texttt{\textordmasculine{FFFD}}, and characters which cannot be reencoded are replaced by either the replacement character \texttt{\textordmasculine{FFFD}} if it exists in the \langle encoding 2 \rangle, or an encoding-specific replacement character, or the question mark character.

\texttt{\str_set_convert:NnnnTF} \str_set_convert:NnnnTF \str_gset_convert:NnnnTF

As \texttt{\str_set_convert:Nnnn}, converts the \langle string \rangle from the encoding given by \langle name 1 \rangle to the encoding given by \langle name 2 \rangle, and assigns the result to \langle str var \rangle. Contrarily to \texttt{\str_set_convert:Nnnn}, the conditional variant does not raise errors in case the \langle string \rangle is not valid according to the \langle name 1 \rangle encoding, or cannot be expressed in the \langle name 2 \rangle encoding. Instead, the \langle false code \rangle is performed.

17.3 Conversion by expansion (for PDF contexts)

A small number of expandable functions are provided for use in PDF string/name contexts. These assume UTF-8 and no escaping in the input.

\texttt{\str_convert_pdfname:n} \str_convert_pdfname:n

As \texttt{\str_convert:Nnnn}, converts the \langle string \rangle on a byte-by-byte basis with non-ASCII codepoints escaped using hashes.

17.4 Possibilities, and things to do

Encoding/escaping-related tasks.
In Xe\TeX/Lua\TeX, would it be better to use the \ldots approach to build a string from a given list of character codes? Namely, within a group, assign 0-9a-f and all characters we want to category “other”, then assign ~ the category superscript, and use \scantokens.

- Change \str_set_convert:Nnnn to expand its last two arguments.
- Describe the internal format in the code comments. Refuse code points in ["D800,"DFFF] in the internal representation?
- Add documentation about each encoding and escaping method, and add examples.
- The hex unescaping should raise an error for odd-token count strings.
- Decide what bytes should be escaped in the uri escaping. Perhaps the characters !’()*-./0123456789_ are safe, and all other characters should be escaped?
- Automate generation of 8-bit mapping files.
- Change the framework for 8-bit encodings: for decoding from 8-bit to Unicode, use 256 integer registers; for encoding, use a tree-box.
- More encodings (see Heiko’s stringenc). CESU?
- More escapings: ascii85, shell escapes, lua escapes, etc.?
Chapter 18

The \texttt{l3quark} package

Quarks

Two special types of constants in \LaTeXe are “quarks” and “scan marks”. By convention all constants of type quark start out with \texttt{\q}, and scan marks start with \texttt{\s}.

18.1 Quarks

Quarks are control sequences (and in fact, token lists) that expand to themselves and should therefore never be executed directly in the code. This would result in an endless loop!

They are meant to be used as delimiter in weird functions, the most common use case being the ‘stop token’ (i.e. \texttt{\q\_stop}). For example, when writing a macro to parse a user-defined date

\begin{verbatim}
\date_parse:n \{19/June/1981\}
\end{verbatim}

one might write a command such as

\begin{verbatim}
\cs_new:Npn \date_parse:n #1 \{ \date_parse_aux:w #1 \q\_stop \}
\cs_new:Npn \date_parse_aux:w #1 / #2 / #3 \q\_stop
\{ \texttt{<do something with the date>} \}
\end{verbatim}

Quarks are sometimes also used as error return values for functions that receive erroneous input. For example, in the function \texttt{\prop\_get:Nnn} to retrieve a value stored in some key of a property list, if the key does not exist then the return value is the quark \texttt{\q\_no\_value}. As mentioned above, such quarks are extremely fragile and it is imperative when using such functions that code is carefully written to check for pathological cases to avoid leakage of a quark into an uncontrolled environment.

Quarks also permit the following ingenious trick when parsing tokens: when you pick up a token in a temporary variable and you want to know whether you have picked up a particular quark, all you have to do is compare the temporary variable to the quark using \texttt{\tl\_if\_eq:NNTF}. A set of special quark testing functions is set up below. All the quark testing functions are expandable although the ones testing only single tokens are much faster.
18.2 Defining quarks

\texttt{\textbackslash quark\_new:N} \quad \texttt{\quark\_new:N (quark)}

Creates a new \texttt{\textasciitilde quark\textbackslash text\_new} which expands only to \texttt{\textasciitilde quark}. The \texttt{\textasciitilde quark} is defined globally, and an error message is raised if the name was already taken.

\texttt{\textbackslash q\_stop} \quad \texttt{\q\_stop}

Used as a marker for delimited arguments, such as

\texttt{\cs\_set:Npn \textasciitilde tmp:w \#1\#2 \textasciitilde q\_stop \{\#1}}

\texttt{\textbackslash q\_mark} \quad \texttt{\q\_mark}

Used as a marker for delimited arguments when \texttt{\textasciitilde q\_stop} is already in use.

\texttt{\textbackslash q\_nil} \quad \texttt{\q\_nil}

Quark to mark a null value in structured variables or functions. Used as an end delimiter when this may itself need to be tested (in contrast to \texttt{\textasciitilde q\_stop}, which is only ever used as a delimiter).

\texttt{\textbackslash q\_no\_value} \quad \texttt{\q\_no\_value}

A canonical value for a missing value, when one is requested from a data structure. This is therefore used as a “return” value by functions such as \texttt{\textasciitilde prop\_get:NnN} if there is no data to return.

18.3 Quark tests

The method used to define quarks means that the single token (N) tests are faster than the multi-token (n) tests. The latter should therefore only be used when the argument can definitely take more than a single token.

\texttt{\quark\_if\_nil\_p:N} \quad \texttt{\quark\_if\_nil\_p:N (token)}

Tests if the \texttt{\textasciitilde token} is equal to \texttt{\textasciitilde q\_nil}.

\texttt{\quark\_if\_nil\_p:n} \quad \texttt{\quark\_if\_nil\_p:n (token list)}

Tests if the \texttt{\textasciitilde token list} contains only \texttt{\textasciitilde q\_nil} (distinct from \texttt{\textasciitilde token list} being empty or containing \texttt{\textasciitilde q\_nil} plus one or more other tokens).

\texttt{\quark\_if\_no\_value\_p:N} \quad \texttt{\quark\_if\_no\_value\_p:N (token)}

Tests if the \texttt{\textasciitilde token} is equal to \texttt{\textasciitilde q\_no\_value}.

\texttt{\quark\_if\_no\_value\_p:n} \quad \texttt{\quark\_if\_no\_value\_p:n (token list)}

Tests if the \texttt{\textasciitilde token list} contains only \texttt{\textasciitilde q\_no\_value} (distinct from \texttt{\textasciitilde token list} being empty or containing \texttt{\textasciitilde q\_no\_value} plus one or more other tokens).
18.4 Recursion

This module provides a uniform interface to intercepting and terminating loops as when one is doing tail recursion. The building blocks follow below and an example is shown in Section 18.4.1.

\q_recursion_tail
This quark is appended to the data structure in question and appears as a real element there. This means it gets any list separators around it.

\q_recursion_stop
This quark is added after the data structure. Its purpose is to make it possible to terminate the recursion at any point easily.

\quark_if_recursion_tail_stop:N \token \quark_if_recursion_tail_stop:N\star
Tests if \token contains only the marker \q_recursion_tail, and if so uses \use-none_delimit_by_q_recursion_stop:w to terminate the recursion that this belongs to. The recursion input must include the marker tokens \q_recursion_tail and \q_recursion_stop as the last two items.

\quark_if_recursion_tail_stop:n \token \{\token list\} \quark_if_recursion_tail_stop:n\star
\quark_if_recursion_tail_stop:o \star
\quark_if_recursion_tail_stop:do:Nn \token \{\token\} \{\insertion\}
Tests if the \token list contains only \q_recursion_tail, and if so uses \use-i_delimit_by_q_recursion_stop:w to terminate the recursion that this belongs to. The recursion input must include the marker tokens \q_recursion_tail and \q_recursion_stop as the last two items. The \insertion code is then added to the input stream after the recursion has ended.

\quark_if_recursion_tail_stop:do:nn \token \{\token list\} \{\insertion\}
\quark_if_recursion_tail_stop:do:on \star
Updated: 2011-09-06
Tests if the \token list contains only \q_recursion_tail, and if so uses \use-i_delimit_by_q_recursion_stop:w to terminate the recursion that this belongs to. The recursion input must include the marker tokens \q_recursion_tail and \q_recursion_stop as the last two items. The \insertion code is then added to the input stream after the recursion has ended.
Tests if \texttt{\{token list\}} contains only \texttt{\q_recursion_tail}, and if so terminates the recursion using \texttt{\(\langle\text{type}\rangle\_map\_break:\)}. The recursion end should be marked by \texttt{\prg_break_point:Nn \langle\text{type}\rangle\_map\_break:}.

### 18.4.1 An example of recursion with quarks

Quarks are mainly used internally in the expl3 code to define recursion functions such as \texttt{\tl_map_inline:nn} and so on. Here is a small example to demonstrate how to use quarks in this fashion. We shall define a command called \texttt{\my_map_dbl:nn} which takes a token list and applies an operation to every pair of tokens. For example, \texttt{\my_map_dbl:nn \{abcd\} \{-#1--#2--\}} would produce \texttt{\{-a--b--\} \{-c--d--\}}. Using quarks to define such functions simplifies their logic and ensures robustness in many cases.

Here's the definition of \texttt{\my_map_dbl:nn}. First of all, define the function that does the processing based on the inline function argument \texttt{#2}. Then initiate the recursion using an internal function. The token list \texttt{#1} is terminated using \texttt{\q_recursion_tail}, with delimiters according to the type of recursion (here a pair of \texttt{\q_recursion_tail}), concluding with \texttt{\q_recursion_stop}. These quarks are used to mark the end of the token list being operated upon.

\begin{verbatim}
\cs_new:Npn \my_map_dbl:nn #1#2 
{ 
  \cs_set:Npn \__my_map_dbl_fn:nn ##1 ##2 {#2} 
  \__my_map_dbl:nn #1 \q_recursion_tail \q_recursion_tail 
  \q_recursion_stop 
}
\end{verbatim}

The definition of the internal recursion function follows. First check if either of the input tokens are the termination quarks. Then, if not, apply the inline function to the two arguments.

\begin{verbatim}
\cs_new:Nn \__my_map_dbl:nn 
{ 
  \quark_if_recursion_tail_stop:n {#1} 
  \quark_if_recursion_tail_stop:n {#2} 
  \__my_map_dbl_fn:nn {#1} {#2} 
}
\end{verbatim}

Finally, recurse:

\begin{verbatim}
\__my_map_dbl:nn 
}\end{verbatim}

Note that contrarily to \LaTeX{}X3 built-in mapping functions, this mapping function cannot be nested, since the second map would overwrite the definition of \texttt{\__my_map_dbl_fn:nn}. 
18.5 Scan marks

Scan marks are control sequences set equal to \texttt{\textbackslash scan\_stop}; hence never expand in an expansion context and are (largely) invisible if they are encountered in a typesetting context.

Like quarks, they can be used as delimiters in weird functions and are often safer to use for this purpose. Since they are harmless when executed by \TeX in non-expandable contexts, they can be used to mark the end of a set of instructions. This allows to skip to that point if the end of the instructions should not be performed (see \texttt{l3regex}).

\begin{verbatim}
\scan_new:N \langle scan mark \rangle
\end{verbatim}

\texttt{\scan_new:N \langle scan mark \rangle}

Creates a new \texttt{\langle scan mark \rangle} which is set equal to \texttt{\textbackslash scan\_stop}: The \texttt{\langle scan mark \rangle} is defined globally, and an error message is raised if the name was already taken by another scan mark.

\begin{verbatim}
\s_stop
\end{verbatim}

\texttt{\s_stop}

Used at the end of a set of instructions, as a marker that can be jumped to using \texttt{\use\_none\_delimit\_by\_s\_stop:w}.

\begin{verbatim}
\use\_none\_delimit\_by\_s\_stop:w \langle tokens \rangle \s_stop
\end{verbatim}

\texttt{\use\_none\_delimit\_by\_s\_stop:w \langle tokens \rangle \s_stop}

Removes the \texttt{\langle tokens \rangle} and \texttt{\s_stop} from the input stream. This leads to a low-level \TeX error if \texttt{\s_stop} is absent.
Chapter 19

The \texttt{l3seq} package
Sequences and stacks

\LaTeX\XeLaTeX{} implements a “sequence” data type, which contain an ordered list of entries which may contain any ⟨balanced text⟩. It is possible to map functions to sequences such that the function is applied to every item in the sequence.

Sequences are also used to implement stack functions in \LaTeX\XeLaTeX{}. This is achieved using a number of dedicated stack functions.

19.1 Creating and initialising sequences

\begin{verbatim}
\seq_new:N \seq_new:c
\seq_clear:N \seq_clear:c
\seq_gclear:N \seq_gclear:c
\seq_clear_new:N \seq_clear_new:c
\seq_gclear_new:N \seq_gclear_new:c
\end{verbatim}

\texttt{\seq_new:N \langle sequence⟩}
Creates a new ⟨sequence⟩ or raises an error if the name is already taken. The declaration is global. The ⟨sequence⟩ initially contains no items.

\texttt{\seq_clear:N \langle sequence⟩}
Clears all items from the ⟨sequence⟩.

\texttt{\seq_clear_new:N \langle sequence⟩}
Ensures that the ⟨sequence⟩ exists globally by applying \texttt{\seq_new:N} if necessary, then applies \texttt{\seq_(g)clear:N} to leave the ⟨sequence⟩ empty.

\texttt{\seq_set_eq:NN \langle sequence₁⟩ \langle sequence₂⟩}
Sets the content of ⟨sequence₁⟩ equal to that of ⟨sequence₂⟩.
\texttt{\textbackslash seq\_set\_from\_clist:NN} \texttt{(sequence) \{comma-list\)}

Converts the data in the \{comma list\} into a \{sequence\}: the original \{comma list\} is unchanged.

\texttt{\textbackslash seq\_const\_from\_clist:NN} \texttt{(seq\ var) \{\textbackslash comma\_list\}}

Creates a new constant \{seq\ var\} or raises an error if the name is already taken. The \{seq\ var\} is set globally to contain the items in the \{comma list\}.

\texttt{\textbackslash seq\_set\_split:NNn} \texttt{(sequence) \{\textbackslash delimiter\}} \texttt{\{\textbackslash token\ list\}}

Splits the \{token list\} into \{items\} separated by \{delimiter\}, and assigns the result to the \{sequence\}. Spaces on both sides of each \{item\} are ignored, then one set of outer braces is removed (if any): this space trimming behaviour is identical to that of \texttt{l3clist} functions. Empty \{items\} are preserved by \texttt{\textbackslash seq\_set\_split:NNn}, and can be removed afterwards using \texttt{\textbackslash seq\_remove\_all:NN} \{sequence\} \{\}. The \{delimiter\} may not contain \{,\} or \# (assuming \TeX\’s normal category code régime). If the \{delimiter\} is empty, the \{token list\} is split into \{items\} as a \{token list\}. See also \texttt{\textbackslash seq\_set\_split\_keep\_spaces:NNn}, which omits space stripping.

\texttt{\textbackslash seq\_set\_split\_keep\_spaces:NNn} \texttt{(sequence) \{\textbackslash delimiter\}} \texttt{\{\textbackslash token\ list\}}

Splits the \{token list\} into \{items\} separated by \{delimiter\}, and assigns the result to the \{sequence\}. One set of outer braces is removed (if any) but any surrounding spaces are retained: any braces inside one or more spaces are therefore kept. Empty \{items\} are preserved by \texttt{\textbackslash seq\_set\_split\_keep\_spaces:NNn}, and can be removed afterwards using \texttt{\textbackslash seq\_remove\_all:NN} \{sequence\} \{\}. The \{delimiter\} may not contain \{,\} or \# (assuming \TeX\’s normal category code régime). If the \{delimiter\} is empty, the \{token list\} is split into \{items\} as a \{token list\}. See also \texttt{\textbackslash seq\_set\_split:NNn}, which removes spaces around the delimiters.

\texttt{\textbackslash seq\_concat:NNW} \texttt{(sequence\}) \texttt{(sequence\_2)} \texttt{(sequence\_3)}

Concatenates the content of \{sequence\_2\} and \{sequence\_3\} together and saves the result in \{sequence\_1\}. The items in \{sequence\_2\} are placed at the left side of the new sequence.
Tests whether the \textit{sequence} is currently defined. This does not check that the \textit{sequence} really is a sequence variable.

19.2 Appending data to sequences

\begin{align*}
\seq_put_left:Nn & \seq_put_left:NV \text{(item)} \\
\seq_put_left:Nn & \seq_put_left:NV \text{(item)} \\
\seq_gput_left:Nn & \seq_gput_left:NV \text{(item)} \\
\seq_put_right:Nn & \seq_put_right:NV \text{(item)} \\
\seq_put_right:Nn & \seq_put_right:NV \text{(item)} \\
\seq_gput_right:Nn & \seq_gput_right:NV \text{(item)} \\
\end{align*}

Appends the \textit{item} to the left of the \textit{sequence}.

19.3 Recovering items from sequences

Items can be recovered from either the left or the right of sequences. For implementation reasons, the actions at the left of the sequence are faster than those acting on the right. These functions all assign the recovered material locally, i.e. setting the \textit{token list variable} used with \texttt{\tl_set:Nn} and never \texttt{\tl_gset:Nn}.

\begin{align*}
\seq_get_left:NN & \seq_get_left:NN \text{(token list variable)} \\
\seq_get_left:NN & \seq_get_left:NN \text{(token list variable)} \\
\seq_get_left:NN & \seq_get_left:NN \text{(token list variable)} \\
\seq_pop_left:NN & \seq_pop_left:NN \text{(token list variable)} \\
\seq_pop_left:NN & \seq_pop_left:NN \text{(token list variable)} \\
\seq_pop_left:NN & \seq_pop_left:NN \text{(token list variable)} \\
\end{align*}

Stores the left-most item from a \textit{sequence} in the \textit{token list variable} without removing it from the \textit{sequence}. The \textit{token list variable} is assigned locally. If \textit{sequence} is empty the \textit{token list variable} is set to the special marker \texttt{\q_no_value}.

Stores the right-most item from a \textit{sequence} in the \textit{token list variable} without removing it from the \textit{sequence}. The \textit{token list variable} is assigned locally. If \textit{sequence} is empty the \textit{token list variable} is set to the special marker \texttt{\q_no_value}.

Pops the left-most item from a \textit{sequence} into the \textit{token list variable}, i.e. removes the item from the sequence and stores it in the \textit{token list variable}. Both of the variables are assigned locally. If \textit{sequence} is empty the \textit{token list variable} is set to the special marker \texttt{\q_no_value}.
\seq_gpop_left:NN \seq_gpop_left:cN
\seq_pop_right:NN \seq_pop_right:cN
\seq_gpop_right:NN \seq_gpop_right:cN
\seq_item:Nn \seq_item:cn
\seq_rand_item:N \seq_rand_item:c

19.4 Recovering values from sequences with branching

The functions in this section combine tests for non-empty sequences with recovery of an item from the sequence. They offer increased readability and performance over separate testing and recovery phases.
\seq_get_left:NNTF \seq_get_left:cNNTF
\seq_get_right:NNTF \seq_get_right:cNNTF
\seq_pop_left:NNTF \seq_pop_left:cNNTF
\seq_gpop_left:NNTF \seq_gpop_left:cNNTF
\seq_pop_right:NNTF \seq_pop_right:cNNTF
\seq_gpop_right:NNTF \seq_gpop_right:cNNTF

\seq_get_left:NNTF \sequence \token \list \variable \{(true \ code}\} \{(false \ code)\}

If the \sequence is empty, leaves the \false \code \in \the \input \stream. \The \value \of \the \token \list \variable \is \not \defined \in \this \case \and \should \not \be \relied \upon. \If \the \sequence \is \non-empty, \stores \the \left-most \item \from \the \sequence \in \the \token \list \variable \without \removing \it \from \the \sequence, \then \leaves \the \true \code \in \the \input \stream. \The \token \list \variable \is \assigned \locally.

\seq_get_right:NNTF \sequence \token \list \variable \{(true \ code}\} \{(false \ code)\}

If the \sequence is empty, leaves the \false \code \in \the \input \stream. \The \value \of \the \token \list \variable \is \not \defined \in \this \case \and \should \not \be \relied \upon. \If \the \sequence \is \non-empty, \stores \the \right-most \item \from \the \sequence \in \the \token \list \variable \without \removing \it \from \the \sequence, \then \leaves \the \true \code \in \the \input \stream. \The \token \list \variable \is \assigned \locally.

\seq_pop_left:NNTF \sequence \token \list \variable \{(true \ code}\} \{(false \ code)\}

If the \sequence is empty, leaves the \false \code \in \the \input \stream. \The \value \of \the \token \list \variable \is \not \defined \in \this \case \and \should \not \be \relied \upon. \If \the \sequence \is \non-empty, \pops \the \left-most \item \from \the \sequence \in \the \token \list \variable, \i.e. \removes \the \item \from \the \sequence, \then \leaves \the \true \code \in \the \input \stream. \Both \the \sequence \and \the \token \list \variable \are \assigned \locally.

\seq_gpop_left:NNTF \sequence \token \list \variable \{(true \ code}\} \{(false \ code)\}

If the \sequence is empty, leaves the \false \code \in \the \input \stream. \The \value \of \the \token \list \variable \is \not \defined \in \this \case \and \should \not \be \relied \upon. \If \the \sequence \is \non-empty, \stores \the \right-most \item \from \the \sequence \in \the \token \list \variable, \i.e. \removes \the \item \from \the \sequence, \then \leaves \the \true \code \in \the \input \stream. \The \sequence \is \modified \globally, \while \the \token \list \variable \is \assigned \locally.

\seq_pop_right:NNTF \sequence \token \list \variable \{(true \ code}\} \{(false \ code)\}

If the \sequence is empty, leaves the \false \code \in \the \input \stream. \The \value \of \the \token \list \variable \is \not \defined \in \this \case \and \should \not \be \relied \upon. \If \the \sequence \is \non-empty, \pops \the \right-most \item \from \the \sequence \in \the \token \list \variable, \i.e. \removes \the \item \from \the \sequence, \then \leaves \the \true \code \in \the \input \stream. \Both \the \sequence \and \the \token \list \variable \are \assigned \locally.

\seq_gpop_right:NNTF \sequence \token \list \variable \{(true \ code}\} \{(false \ code)\}

If the \sequence is empty, leaves the \false \code \in \the \input \stream. \The \value \of \the \token \list \variable \is \not \defined \in \this \case \and \should \not \be \relied \upon. \If \the \sequence \is \non-empty, \pops \the \right-most \item \from \the \sequence \in \the \token \list \variable, \i.e. \removes \the \item \from \the \sequence, \then \leaves \the \true \code \in \the \input \stream. \The \sequence \is \modified \globally, \while \the \token \list \variable \is \assigned \locally.
19.5 Modifying sequences

While sequences are normally used as ordered lists, it may be necessary to modify the content. The functions here may be used to update sequences, while retaining the order of the unaffected entries.

\seq_remove_duplicates:N \seq_remove_duplicates:c \seq_gremove_duplicates:N \seq_gremove_duplicates:c

Removes duplicate items from the \seq, leaving the left most copy of each item in the \seq. The \item comparison takes place on a token basis, as for \tl_if_eq:nnTF.

\seq_remove_all:Nn \seq_remove_all:cn \seq_gremove_all:Nn \seq_gremove_all:cn

Removes every occurrence of \item from the \seq. The \item comparison takes place on a token basis, as for \tl_if_eq:nnTF.

\seq_reverse:N \seq_reverse:c \seq_greverse:N \seq_greverse:c

Reverses the order of the items stored in the \seq.

\seq_sort:Nn \seq_sort:cn \seq_gsort:Nn \seq_gsort:cn

Sorts the items in the \seq according to the \comparison code, and assigns the result to \seq. The details of sorting comparison are described in Section 6.1.

\seq_shuffle:N \seq_shuffle:c \seq_gshuffle:N \seq_gshuffle:c

Sets the \seq to the result of placing the items of the \seq in a random order. Each item is (roughly) as likely to end up in any given position.

\seq_if_empty_p:N \seq_if_empty_p:c \seq_if_empty:NTF \seq_if_empty:cTF

Tests if the \seq is empty (containing no items).
Tests if the \langle item \rangle is present in the \langle sequence \rangle.

## 19.7 Mapping over sequences

All mappings are done at the current group level, \textit{i.e.} any local assignments made by the \langle function \rangle or \langle code \rangle discussed below remain in effect after the loop.

\seq_if_in:NnTF \langle sequence \rangle \langle item \rangle \{ \langle true code \rangle \} \{ \langle false code \rangle \}

\seq_if_in:Nn \langle \text{NV|Nv|No|Nx|cn|cV|cv|co|cx}\rangle TF

Tests if the \langle item \rangle is present in the \langle sequence \rangle.

\seq_map_function:NN \langle sequence \rangle \langle function \rangle

Applies \langle function \rangle to every \langle item \rangle stored in the \langle sequence \rangle. The \langle function \rangle will receive one argument for each iteration. The \langle items \rangle are returned from left to right. To pass further arguments to the \langle function \rangle, see \seq_map_tokens:Nn. The function \seq_map_inline:Nn is faster than \seq_map_function:NN for sequences with more than about 10 items.

\seq_map_inline:Nn \langle sequence \rangle \{ \langle inline function \rangle \}

Applies \langle inline function \rangle to every \langle item \rangle stored within the \langle sequence \rangle. The \langle inline function \rangle should consist of code which will receive the \langle item \rangle as \#1. The \langle items \rangle are returned from left to right.

\seq_map_tokens:Nn \langle sequence \rangle \{ \langle code \rangle \}

Analogue of \seq_map_function:NN which maps several tokens instead of a single function. The \langle code \rangle receives each item in the \langle sequence \rangle as a trailing brace group. For instance,

\seq_map_tokens:Nn \l_my_seq \{ \prg_replicate:nn \{ 2 \} \}

expands to twice each item in the \langle sequence \rangle: for each item in \l_my_seq the function \prg_replicate:nn receives 2 and \langle item \rangle as its two arguments. The function \seq_map_inline:Nn is typically faster but it is not expandable.

\seq_map_variable:NNn \langle sequence \rangle \langle variable \rangle \{ \langle code \rangle \}

Stores each \langle item \rangle of the \langle sequence \rangle in turn in the (token list) \langle variable \rangle and applies the \langle code \rangle. The \langle code \rangle will usually make use of the \langle variable \rangle, but this is not enforced. The assignments to the \langle variable \rangle are local. Its value after the loop is the last \langle item \rangle in the \langle sequence \rangle, or its original value if the \langle sequence \rangle is empty. The \langle items \rangle are returned from left to right.

\seq_map_indexed_function:NN \langle sequence variable \rangle \langle function \rangle

Applies \langle function \rangle to every entry in the \langle sequence variable \rangle. The \langle function \rangle should have signature :nn. It receives two arguments for each iteration: the \langle index \rangle (namely 1 for the first entry, then 2 and so on) and the \langle item \rangle.
\seq_map_indexed_inline:Nn \seq_map_indexed_inline:Nn \seq_map_indexed_inline:Nn \seq_map_indexed_inline:Nn \seq_map_indexed_inline:Nn \seq_map_indexed_inline:Nn \seq_map_indexed_inline:Nn

Applies \texttt{<inline function>} to every entry in the \texttt{<sequence variable>}. The \texttt{<inline function>} should consist of code which receives the \texttt{<index>} (namely 1 for the first entry, then 2 and so on) as \#1 and the \texttt{<item>} as \#2.

\seq_map_break: ⭐

\seq_map_break:

Updated: 2012-06-29

Used to terminate a \texttt{\seq_map...} function before all entries in the \texttt{<sequence>} have been processed. This normally takes place within a conditional statement, for example

\seq_map_inline:Nn \l_my_seq
{\str_if_eq:nnTF { #1 } { bingo }{ \seq_map_break: }
{ % Do something useful
}
}

Use outside of a \texttt{\seq_map...} scenario leads to low level \TeX{} errors.

\textbf{\TeX{}hackers note}: When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.

\seq_map_break:n ⭐

\seq_map_break:n \seq_map_break:n \seq_map_break:n

Updated: 2012-06-29

Used to terminate a \texttt{\seq_map...} function before all entries in the \texttt{<sequence>} have been processed, inserting the \texttt{<code>} after the mapping has ended. This normally takes place within a conditional statement, for example

\seq_map_inline:Nn \l_my_seq
{\str_if_eq:nTF { #1 } { bingo }{ \seq_map_break:n { <code> } }{ % Do something useful
}
}

Use outside of a \texttt{\seq_map...} scenario leads to low level \TeX{} errors.

\textbf{\TeX{}hackers note}: When the mapping is broken, additional tokens may be inserted before the \texttt{<code>} is inserted into the input stream. This depends on the design of the mapping function.
\seq_set_map:NNn \seq_gset_map:NNn

Applies \textit{inline function} to every \textit{item} stored within the \textit{sequence}. The \textit{inline function} should consist of code which will receive the \textit{item} as \texttt.id. The sequence resulting applying \textit{inline function} to each \textit{item} is assigned to \textit{sequence}.\textbf{\textsc{\texttt{\LaTeX}}hackers note:} Contrarily to other mapping functions, \seq_map_break: cannot be used in this function, and would lead to low-level \LaTeX{} errors.

\seq_set_map_x:NNn \seq_gset_map_x:NNn

Applies \textit{inline function} to every \textit{item} stored within the \textit{sequence}. The \textit{inline function} from \texttt.x-expanding \textit{inline function} applied to each \textit{item} is assigned to \textit{sequence}. As such, the code in \textit{inline function} should be expandable.\textbf{\textsc{\texttt{\LaTeX}}hackers note:} Contrarily to other mapping functions, \seq_map_break: cannot be used in this function, and would lead to low-level \LaTeX{} errors.

\seq_count:N \star \seq_count:c\star

Leaves the number of items in the \textit{sequence} in the input stream as an \textit{integer denotation}. The total number of items in a \textit{sequence} includes those which are empty and duplicates, \textit{i.e.} every item in a \textit{sequence} is unique.

\textbf{19.8 Using the content of sequences directly}

\seq_set_split:Nnn \l_tmpa_seq { | } { a | b | c | \{de\} | f }
\seq_use:Nnnn \l_tmpa_seq { \texttt{}and\texttt{}} { ,\texttt{}} { ,\texttt{and\texttt{}}} inserts “a, b, c, de, and f” in the input stream. The first separator argument is not used in this case because the sequence has more than 2 items.

\textbf{\textsc{\texttt{\LaTeX}}hackers note:} The result is returned within the \texttt.unexpanded primitive (\exp_not:n), which means that the \textit{items} do not expand further when appearing in an \texttt.x-type argument expansion.
\seq_use:Nn \seq_use:cn *
\seq_set_split:Nnn \l_tmpa_seq { | } { a | b | c | {de} | f }
\seq_use:Nn \l_tmpa_seq { ~and~ }

\TeXhackers note: The result is returned within the \unexpanded primitive (\exp_not:n),
which means that the ⟨items⟩ do not expand further when appearing in an x-type argument expansion.

19.9 Sequences as stacks

Sequences can be used as stacks, where data is pushed to and popped from the top of
the sequence. (The left of a sequence is the top, for performance reasons.) The stack
functions for sequences are not intended to be mixed with the general ordered data
functions detailed in the previous section: a sequence should either be used as an ordered
data type or as a stack, but not in both ways.

\seq_get:NN \seq_get:cN
\seq_pop:NN \seq_pop:cN
\seq_gpop:NN \seq_gpop:cN
\seq_get:NNF \seq_get:cnF

If the (sequence) is empty, leaves the (false code) in the input stream. The value of the
(token list variable) is not defined in this case and should not be relied upon. If the
(sequence) is non-empty, stores the top item from a (sequence) in the (token list variable)
without removing it from the (sequence). The (token list variable) is assigned locally.
\seq_pop:NNTF \seq_pop:cNNTF
\seq_gpop:NNTF \seq_gpop:cNNTF
\seq_gpush:Nn \seq_gpush:NV|

If the \langle sequence \rangle is empty, leaves the \langle false code \rangle in the input stream. The value of the \langle token list variable \rangle is not defined in this case and should not be relied upon. If the \langle sequence \rangle is non-empty, pops the top item from the \langle sequence \rangle in the \langle token list variable \rangle, i.e. removes the item from the \langle sequence \rangle. Both the \langle sequence \rangle and the \langle token list variable \rangle are assigned locally.

\seq_push:Nn \seq_push:NV|
\seq_gpush:Nn \seq_gpush:NV|

Adds the \{\langle item \rangle\} to the top of the \langle sequence \rangle.

19.10 Sequences as sets

Sequences can also be used as sets, such that all of their items are distinct. Usage of sequences as sets is not currently widespread, hence no specific set function is provided. Instead, it is explained here how common set operations can be performed by combining several functions described in earlier sections. When using sequences to implement sets, one should be careful not to rely on the order of items in the sequence representing the set.

Sets should not contain several occurrences of a given item. To make sure that a \langle sequence variable \rangle only has distinct items, use \seq_remove_duplicates:N \langle sequence variable \rangle. This function is relatively slow, and to avoid performance issues one should only use it when necessary.

Some operations on a set \langle seq var \rangle are straightforward. For instance, \seq_count:N \langle seq var \rangle expands to the number of items, while \seq_if_in:NnTF \langle seq var \rangle \{\langle item \rangle\} tests if the \langle item \rangle is in the set.

Adding an \langle item \rangle to a set \langle seq var \rangle can be done by appending it to the \langle seq var \rangle if it is not already in the \langle seq var \rangle:

\seq_if_in:NnF \seq_remove_all:Nn \langle seq var \rangle \{\langle item \rangle\}

Removing an \langle item \rangle from a set \langle seq var \rangle can be done using \seq_remove_all:Nn,

The intersection of two sets \langle seq var1 \rangle and \langle seq var2 \rangle can be stored into \langle seq var3 \rangle by collecting items of \langle seq var1 \rangle which are in \langle seq var2 \rangle.
\seq_clear:N \seq_var_3
\seq_map_inline:Nn \seq_var_1
{
\seq_if_in:NnT \seq_var_2 {#1}
{ \seq_put_right:Nn \seq_var_3 {#1} }
}

The code as written here only works if \seq_var_3 is different from the other two sequence variables. To cover all cases, items should first be collected in a sequence \l__\pkg\_internal_seq, then \seq_var_3 should be set equal to this internal sequence. The same remark applies to other set functions.

The union of two sets \seq_var_1 and \seq_var_2 can be stored into \seq_var_3 through
\seq_concat:NNN \seq_var_3 \seq_var_1 \seq_var_2
\seq_remove_duplicates:N \seq_var_3

or by adding items to (a copy of) \seq_var_1 one by one
\seq_set_eq:NN \seq_var_3 \seq_var_1
\seq_map_inline:Nn \seq_var_2
{ \seq_put_right:Nn \seq_var_3 {#1} }

The second approach is faster than the first when the \seq_var_2 is short compared to \seq_var_1.

The difference of two sets \seq_var_1 and \seq_var_2 can be stored into \seq_var_3 by removing items of the \seq_var_2 from (a copy of) the \seq_var_1 one by one.
\seq_set_eq:NN \seq_var_3 \seq_var_1
\seq_map_inline:Nn \seq_var_2
{ \seq_remove_all:Nn \seq_var_3 {#1} }

The symmetric difference of two sets \seq_var_1 and \seq_var_2 can be stored into \seq_var_3 by computing the difference between \seq_var_1 and \seq_var_2 and storing the result as \l__\pkg\_internal_seq, then the difference between \seq_var_2 and \seq_var_1, and finally concatenating the two differences to get the symmetric differences.
\seq_set_eq:NN \seq_var_1 \l__\pkg\_internal_seq \seq_var_1
\seq_set_eq:NN \seq_var_2 \l__\pkg\_internal_seq \seq_var_2
\seq_set_eq:NN \seq_var_3 \seq_var_1
\seq_set_eq:NN \seq_var_4 \seq_var_2
\seq_set_eq:NN \seq_var_5 \seq_var_3
\seq_concat:NNN \seq_var_3 \seq_var_4 \seq_var_5

19.11 Constant and scratch sequences

\c_empty_seq Constant that is always empty.

See: 2012-07-02

152
Scratch sequences for local assignment. These are never used by the kernel code, and so are safe for use with any \TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\l_tmpa_seq
\l_tmpb_seq

New: 2012-04-26

Scratch sequences for global assignment. These are never used by the kernel code, and so are safe for use with any \TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_seq
\g_tmpb_seq

New: 2012-04-26

## 19.12 Viewing sequences

\seq_show:N \seq_show:c

Updated: 2021-04-29

\seq_log:N \seq_log:c

New: 2014-08-12
Updated: 2021-04-29

\seq_show:N \sequence
Displays the entries in the \sequence in the terminal.

\seq_log:N \sequence
Writes the entries in the \sequence in the log file.
Chapter 20

The \texttt{l3int} package

Integers

Calculation and comparison of integer values can be carried out using literal numbers, \texttt{int} registers, constants and integers stored in token list variables. The standard operators $+$, $-$, $/$ and $*$ and parentheses can be used within such expressions to carry arithmetic operations. This module carries out these functions on integer expressions ("\texttt{intexpr}").

20.1 Integer expressions

Throughout this module, (almost) all \texttt{n}-type argument allow for an \texttt{(intexpr)} argument with the following syntax. The \texttt{(integer expression)} should consist, after expansion, of $+$, $-$, $*$, $/$, $($, $)$ and of course integer operands. The result is calculated by applying standard mathematical rules with the following peculiarities:

\begin{itemize}
  \item $/$ denotes division rounded to the closest integer with ties rounded away from zero;
  \item there is an error and the overall expression evaluates to zero whenever the absolute value of any intermediate result exceeds $2^{31} - 1$, except in the case of scaling operations $a*b/c$, for which $a*b$ may be arbitrarily large (but the operands $a$, $b$, $c$ are still constrained to an absolute value at most $2^{31} - 1$);
  \item parentheses may not appear after unary $+$ or $-$, namely placing $+($ or $-($ at the start of an expression or after $+$, $-$, $*$, $/$ or $($ leads to an error.
\end{itemize}

Each integer operand can be either an integer variable (with no need for \texttt{\int_use:N}) or an integer denotation. For example both

\begin{verbatim}
\int_show:n { 5 + 4 * 3 - ( 3 + 4 * 5 ) }

\tl_new:N \l_my_tl
\tl_set:Nn \l_my_tl { 5 }
\int_new:N \l_my_int
\int_set:Nn \l_my_int { 4 }
\int_show:n { \l_my_tl + \l_my_int * 3 - ( 3 + 4 * 5 ) }
\end{verbatim}

and

\begin{verbatim}
\tl_new:N \l_my_tl
\tl_set:Nn \l_my_tl { 5 }
\int_new:N \l_my_int
\int_set:Nn \l_my_int { 4 }
\int_show:n { \l_my_tl + \l_my_int * 3 - ( 3 + 4 * 5 ) }
\end{verbatim}
show the same result −6 because \_my\_ti expands to the integer denotation 5 while the
integer variable \_my\_int takes the value 4. As the \textit{integer expression} is fully expanded
from left to right during evaluation, fully expandable and restricted-expandable functions
can both be used, and \texttt{exp\_not:n} and its variants have no effect while \texttt{exp\_not:N}
may incorrectly interrupt the expression.

\textbf{\TeX}hackers note: Exactly two expansions are needed to evaluate \texttt{int\_eval:n}. The
result is \textit{not} an \textit{internal integer}, and therefore should be terminated by a space if used in
\texttt{int\_value:w} or in a \TeX-style integer assignment.

As all \TeX\ integers, integer operands can also be: \texttt{value⟨\LaTeX\ 2e\ counter⟩}; dimension
or skip variables, converted to integers in sp; the character code of some character given as
\langle\texttt{char}\rangle or \langle\texttt{char}\rangle; octal numbers given as \texttt{'\textbackslash\texttt{0} to \texttt{7}}; or hexadecimal
numbers given as \texttt{\textbackslash\texttt{0} to \texttt{F}}.

\begin{verbatim}
\int_eval:n \langle integer expression \rangle
Evaluates the \textit{integer expression} and leaves the result in the input stream as an integer
denotation: for positive results an explicit sequence of decimal digits not starting with 0, for
negative results − followed by such a sequence, and 0 for zero.
\end{verbatim}

\begin{verbatim}
\int_eval:w \langle integer expression \rangle
Evaluate the \textit{integer expression} as described for \texttt{int\_eval:n}. The end of the expres-
sion is the first token encountered that cannot form part of such an expression. If that
token is \texttt{\textbackslash scan\_stop}: it is removed, otherwise not. Spaces do not terminate the expres-
sion. However, spaces terminate explicit integers, and this may terminate the expression:
for instance, \texttt{int\_eval:w 1\texttt{~}+\texttt{~}1\texttt{~}9} (with explicit space tokens inserted using − in a code
setting) expands to 29 since the digit 9 is not part of the expression.
\end{verbatim}

\begin{verbatim}
\int_sign:n \langle integer expression \rangle
Evaluates the \textit{integer expression} then leaves 1 or 0 or −1 in the input stream according
to the sign of the result.
\end{verbatim}

\begin{verbatim}
\int_abs:n \langle integer expression \rangle
Evaluates the \textit{integer expression} as described for \texttt{int\_eval:n} and leaves the absolute
value of the result in the input stream as an \textit{integer denotation} after two expansions.
\end{verbatim}

\begin{verbatim}
\int_div_round:nn \langle intexpr \rangle \{ \langle intexpr \rangle \}
Evaluates the two \textit{integer expressions} as described earlier, then divides the first value
by the second, and rounds the result to the closest integer. Ties are rounded away from
zero. Note that this is identical to using / directly in an \textit{integer expression}. The result
is left in the input stream as an \textit{integer denotation} after two expansions.
\end{verbatim}

\begin{verbatim}
\int_div_truncate:nn \langle intexpr \rangle \{ \langle intexpr \rangle \}
Evaluates the two \textit{integer expressions} as described earlier, then divides the first value by
the second, and rounds the result towards zero. Note that division using / rounds to the
closest integer instead. The result is left in the input stream as an \textit{integer denotation}
after two expansions.
\end{verbatim}
\int_max:nn \int_min:nn
\int_mod:nn

\int_max:nn {\langle intexpr_1 \rangle} {\langle intexpr_2 \rangle}
\int_min:nn {\langle intexpr_1 \rangle} {\langle intexpr_2 \rangle}
\int_mod:nn {\langle intexpr_1 \rangle} {\langle intexpr_2 \rangle}

Evaluates the \(\langle integer \ expressions \rangle\) as described for \int_eval:n and leaves either the larger or smaller value in the input stream as an \(\langle integer \ denotation \rangle\) after two expansions.

Evaluates the two \(\langle integer \ expressions \rangle\) as described earlier, then calculates the integer remainder of dividing the first expression by the second. This is obtained by subtracting \int_div_truncate:nn {\langle intexpr_1 \rangle} {\langle intexpr_2 \rangle} times \langle intexpr_2 \rangle \text{ from } \langle intexpr_1 \rangle. Thus, the result has the same sign as \langle intexpr_1 \rangle and its absolute value is strictly less than that of \langle intexpr_2 \rangle. The result is left in the input stream as an \(\langle integer \ denotation \rangle\) after two expansions.

20.2 Creating and initialising integers

\int_new:N \int_new:c
\int_const:Nn \int_const:cn
\int_zero:N \int_zero:c \int_gzero:N \int_gzero:c
\int_zero_new:N \int_zero_new:c \int_gzero_new:N \int_gzero_new:c
\int_set_eq:NN \int_set_eq: (cN)N (cN)cc \int_gset_eq:NN \int_gset_eq: (cN)N (cN)cc
\int_if_exist_p:N \int_if_exist_p:c \int_if_exist:NTF \int_if_exist:c TT

\int_new:N \langle integer \rangle
\int_new:c \langle integer \rangle
\int_const:Nn \langle integer \rangle \{\langle integer \ expression \rangle\}
\int_const:cn \langle integer \rangle \{\langle integer \ expression \rangle\}
!\int_zero:N \langle integer \rangle
\int_zero:c \langle integer \rangle \int_gzero:N \langle integer \rangle \int_gzero:c \langle integer \rangle
\int_zero_new:N \langle integer \rangle \int_zero_new:c \langle integer \rangle \int_gzero_new:N \langle integer \rangle \int_gzero_new:c \langle integer \rangle
\int_set_eq:NN \langle integer_1 \rangle \langle integer_2 \rangle \int_set_eq: (cN)N (cN)cc \int_gset_eq:NN \langle integer_1 \rangle \langle integer_2 \rangle \int_gset_eq: (cN)N (cN)cc
\int_if_exist_p:N \langle int \rangle \int_if_exist_p:c \langle int \rangle \int_if_exist:NTF \langle int \rangle \{\langle true \ code \rangle\} \{\langle false \ code \rangle\}
\int_if_exist:c TT \langle int \rangle \int_if_exist:NTF \langle int \rangle \{\langle true \ code \rangle\} \{\langle false \ code \rangle\}

Creates a new \(\langle integer \rangle\) or raises an error if the name is already taken. The declaration is global. The \(\langle integer \rangle\) is initially equal to 0.

Creates a new constant \(\langle integer \rangle\) or raises an error if the name is already taken. The value of the \(\langle integer \rangle\) is set globally to the \(\langle integer \ expression \rangle\).

Sets \(\langle integer \rangle\) to 0.

Ensures that the \(\langle integer \rangle\) exists globally by applying \int_new:N if necessary, then applies \int_(g)zero:N to leave the \(\langle integer \rangle\) set to zero.

Sets the content of \(\langle integer_1 \rangle\) equal to that of \(\langle integer_2 \rangle\).

Tests whether the \(\langle int \rangle\) is currently defined. This does not check that the \(\langle int \rangle\) really is an integer variable.
20.3 Setting and incrementing integers

\int_add:Nn \int_add:n \text{(integer)} \{\langle \text{integer expression} \rangle \}

Adds the result of the \langle \text{integer expression} \rangle to the current content of the \langle \text{integer} \rangle.

\int_decr:N \int_decr:n \int_gdecr:N \int_gdecr:n

\int_decr:N \langle \text{integer} \rangle

Decreases the value stored in \langle \text{integer} \rangle by 1.

\int_incr:N \int_incr:n \int_gincr:N \int_gincr:n

\int_incr:N \langle \text{integer} \rangle

Increases the value stored in \langle \text{integer} \rangle by 1.

\int_set:Nn \int_set:n \int_gset:Nn \int_gset:n

\int_set:Nn \langle \text{integer} \rangle \{\langle \text{integer expression} \rangle \}

Sets \langle \text{integer} \rangle to the value of \langle \text{integer expression} \rangle, which must evaluate to an integer (as described for \int_eval:n).

\int_sub:Nn \int_sub:n \int_gsub:Nn \int_gsub:n

\int_sub:Nn \langle \text{integer} \rangle \{\langle \text{integer expression} \rangle \}

Subtracts the result of the \langle \text{integer expression} \rangle from the current content of the \langle \text{integer} \rangle.

20.4 Using integers

\int_use:N \int_use:n \int_use:c

\int_use:N \langle \text{integer} \rangle

Recovers the content of an \langle \text{integer} \rangle and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Can be omitted in places where an \langle \text{integer} \rangle is required (such as in the first and third arguments of \int_compare:nNnTF).

\textbf{\TeX}hackers note: \texttt{\int_use:N} is the \TeX\ primitive \texttt{\the}: this is one of several \L\TeX\3 names for this primitive.
20.5 Integer expression conditionals

\int_compare_p:nNn  \\int_compare:nNnTF
\int_compare_p:nNn \{\textit{intexpr}_1\} \{\textit{relation}\} \{\textit{intexpr}_2\}
\int_compare:nNnTF \{\{\textit{intexpr}_1\}\} \{\{\textit{relation}\}\} \{\{\textit{intexpr}_2\}\}
\{\{\textit{true code}\}\} \{\{\textit{false code}\}\}

This function first evaluates each of the \textit{(integer expressions)} as described for \texttt{\int_eval:n}. The two results are then compared using the \textit{(relation)}:

\text{Equal} =
\text{Greater than} >
\text{Less than} <

This function is less flexible than \texttt{\int_compare:nTF} but around 5 times faster.

\int_compare_p:n  \\int_compare:nTF
\int_compare_p:n \{\textit{intexpr}_1\} \{\textit{relation}_1\}
\{\textit{intexpr}_2\} \{\textit{relation}_2\}
\{\textit{intexpr}_3\} \{\textit{relation}_3\}
\{\textit{intexpr}_N\} \{\textit{relation}_N\}
\{\textit{intexpr}_{N+1}\}
\int_compare:nTF \{\{\textit{true code}\}\} \{\{\textit{false code}\}\}

This function evaluates the \textit{(integer expressions)} as described for \texttt{\int_eval:n} and compares consecutive result using the corresponding \textit{(relation)}, namely it compares \textit{(intexpr}_1\} and \textit{(intexpr}_2\} using the \textit{(relation}_1\}, then \textit{(intexpr}_2\} and \textit{(intexpr}_3\} using the \textit{(relation}_2\}, until finally comparing \textit{(intexpr}_N\} and \textit{(intexpr}_{N+1}\} using the \textit{(relation}_N\}. The test yields \textit{true} if all comparisons are \textit{true}. Each \textit{(integer expression)} is evaluated only once, and the evaluation is lazy, in the sense that if one comparison is \textit{false}, then no other \textit{(integer expression)} is evaluated and no other comparison is performed. The \textit{(relations)} can be any of the following:

\text{Equal} = \text{or} \text{==}
\text{Greater than or equal to} \text{>=}
\text{Greater than} \text{>}
\text{Less than or equal to} \text{<=}
\text{Less than} \text{<}
\text{Not equal} \text{!=}

This function is more flexible than \texttt{\int_compare:nNnTF} but around 5 times slower.
This function evaluates the \( \textit{test integer expression} \) and compares this in turn to each of the \( \textit{integer expression cases} \). If the two are equal then the associated \( \textit{code} \) is left in the input stream and other cases are discarded. If any of the cases are matched, the \( \textit{true code} \) is also inserted into the input stream (after the code for the appropriate case), while if none match then the \( \textit{false code} \) is inserted. The function \texttt{\int_case:nn}, which does nothing if there is no match, is also available. For example

\[
\texttt{\int_case:nnF}\{2 \times 5\} = \{\text{Small}\} \{\text{Medium}\} \{\text{Negative}\} \{\text{No idea!}\}
\]

leaves “Medium” in the input stream.

This function first evaluates the \( \textit{integer expression} \) as described for \texttt{\int_eval:n}. It then evaluates if this is odd or even, as appropriate.

### 20.6 Integer expression loops

Placing the \( \textit{code} \) in the input stream for \LaTeX{} to process, and then evaluates the relationship between the two \( \textit{integer expressions} \) as described for \texttt{\int_compare:nNnTF}. If the test is \texttt{false} then the \( \textit{code} \) is inserted into the input stream again and a loop occurs until the \( \textit{relation} \) is \texttt{true}.

Placing the \( \textit{code} \) in the input stream for \LaTeX{} to process, and then evaluates the relationship between the two \( \textit{integer expressions} \) as described for \texttt{\int_compare:nNnTF}. If the test is \texttt{true} then the \( \textit{code} \) is inserted into the input stream again and a loop occurs until the \( \textit{relation} \) is \texttt{false}. 

...
Evaluates the relationship between the two \textit{integer expressions} as described for \texttt{\int_compare:nTF}, and then places the \textit{code} in the input stream if the \textit{relation} is \texttt{false}. After the \textit{code} has been processed by \LaTeX the test is repeated, and a loop occurs until the test is \texttt{true}.

Evaluates the relationship between the two \textit{integer expressions} as described for \texttt{\int_compare:nTF}, and then places the \textit{code} in the input stream if the \textit{relation} is \texttt{true}. After the \textit{code} has been processed by \LaTeX the test is repeated, and a loop occurs until the test is \texttt{false}.

Places the \textit{code} in the input stream for \LaTeX to process, and then evaluates the \textit{integer relation} as described for \texttt{\int_compare:nTF}. If the test is \texttt{false} then the \textit{code} is inserted into the input stream again and a loop occurs until the \textit{relation} is \texttt{true}.

Places the \textit{code} in the input stream for \LaTeX to process, and then evaluates the \textit{integer relation} as described for \texttt{\int_compare:nTF}. If the test is \texttt{true} then the \textit{code} is inserted into the input stream again and a loop occurs until the \textit{relation} is \texttt{false}.

Evaluates the \textit{integer relation} as described for \texttt{\int_compare:nTF}, and then places the \textit{code} in the input stream if the \textit{relation} is \texttt{false}. After the \textit{code} has been processed by \LaTeX the test is repeated, and a loop occurs until the test is \texttt{true}.

Evaluates the \textit{integer relation} as described for \texttt{\int_compare:nTF}, and then places the \textit{code} in the input stream if the \textit{relation} is \texttt{true}. After the \textit{code} has been processed by \LaTeX the test is repeated, and a loop occurs until the test is \texttt{false}.
20.7 Integer step functions

\begin{verbatim}
\int_step_function:nnN \{ (final value) \} \{ (function) \}
\int_step_function:nnN \{ (initial value) \} \{ (final value) \} \{ (function) \}
\int_step_function:nnN \{ (initial value) \} \{ (step) \} \{ (final value) \} \{ (function) \}
\end{verbatim}

This function first evaluates the \textit{(initial value)}, \textit{(step)} and \textit{(final value)}, all of which should be integer expressions. The \textit{(function)} is then placed in front of each \textit{(value)} from the \textit{(initial value)} to the \textit{(final value)} in turn (using \textit{(step)} between each \textit{(value)}). The \textit{(step)} must be non-zero. If the \textit{(step)} is positive, the loop stops when the \textit{(value)} becomes larger than the \textit{(final value)}. If the \textit{(step)} is negative, the loop stops when the \textit{(value)} becomes smaller than the \textit{(final value)}. The \textit{(function)} should absorb one numerical argument. For example

\begin{verbatim}
\cs_set:Npn \my_func:n #1 { [I saw \textast \#1] \quad }
\int_step_function:nnN \{ 1 \} \{ 1 \} \{ 5 \} \my_func:n
\end{verbatim}

would print

\begin{verbatim}
[I saw 1] [I saw 2] [I saw 3] [I saw 4] [I saw 5]
\end{verbatim}

The functions \texttt{\int_step_function:nnN} and \texttt{\int_step_function:nnN} both use a fixed \textit{(step)} of 1, and in the case of \texttt{\int_step_function:nnN} the \textit{(initial value)} is also fixed as 1. These functions are provided as simple short-cuts for code clarity.

\begin{verbatim}
\int_step_inline:nn { (final value) } { (code) }
\int_step_inline:nn { (initial value) } { (final value) } { (code) }
\int_step_inline:nn { (initial value) } { (step) } { (final value) } { (code) }
\end{verbatim}

This function first evaluates the \textit{(initial value)}, \textit{(step)} and \textit{(final value)}, all of which should be integer expressions. Then for each \textit{(value)} from the \textit{(initial value)} to the \textit{(final value)} in turn (using \textit{(step)} between each \textit{(value)}), the \textit{(code)} is inserted into the input stream with \texttt{\#1} replaced by the current \textit{(value)}. Thus the \textit{(code)} should define a function of one argument (\texttt{\#1}).

The functions \texttt{\int_step_inline:nn} and \texttt{\int_step_inline:nn} both use a fixed \textit{(step)} of 1, and in the case of \texttt{\int_step_inline:nn} the \textit{(initial value)} is also fixed as 1. These functions are provided as simple short-cuts for code clarity.

\begin{verbatim}
\int_step_variable:nN \{ (final value) \} { tl var } { (code) }
\int_step_variable:nN \{ (initial value) \} { (final value) } { tl var } { (code) }
\int_step_variable:nNn \{ (initial value) \} \{ (step) \} \{ (final value) \} { tl var } { (code) }
\end{verbatim}

This function first evaluates the \textit{(initial value)}, \textit{(step)} and \textit{(final value)}, all of which should be integer expressions. Then for each \textit{(value)} from the \textit{(initial value)} to the \textit{(final value)} in turn (using \textit{(step)} between each \textit{(value)}), the \textit{(code)} is inserted into the input stream, with the \texttt{\{tl var\}} defined as the current \textit{(value)}. Thus the \textit{(code)} should make use of the \texttt{\{tl var\}}.

The functions \texttt{\int_step_variable:nN} and \texttt{\int_step_variable:nNn} both use a fixed \textit{(step)} of 1, and in the case of \texttt{\int_step_variable:nN} the \textit{(initial value)} is also fixed as 1. These functions are provided as simple short-cuts for code clarity.
20.8 Formatting integers

Integers can be placed into the output stream with formatting. These conversions apply to any integer expressions.

\int_to_arabic:n \{integer expression\}

Places the value of the \(\text{integer expression}\) in the input stream as digits, with category code 12 (other).

\int_to_alph:n \{integer expression\}

Evaluates the \(\text{integer expression}\) and converts the result into a series of letters, which are then left in the input stream. The conversion rule uses the 26 letters of the English alphabet, in order, adding letters when necessary to increase the total possible range of representable numbers. Thus

\int_to_alph:n \{ 1 \}

places \(a\) in the input stream,

\int_to_alph:n \{ 26 \}

is represented as \(z\) and

\int_to_alph:n \{ 27 \}

is converted to \(aa\). For conversions using other alphabets, use \int_to_symbols:nnn to define an alphabet-specific function. The basic \int_to_alph:n and \int_to_Alph:n functions should not be modified. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\int_to_symbols:nnn \{(integer expression)\} \{(total symbols)\} \{(value to symbol mapping)\}

This is the low-level function for conversion of an \(\text{integer expression}\) into a symbolic form (often letters). The \(\text{total symbols}\) available should be given as an integer expression. Values are actually converted to symbols according to the \(\text{value to symbol mapping}\). This should be given as \(\text{total symbols}\) pairs of entries, a number and the appropriate symbol. Thus the \int_to_alph:n function is defined as

\cs_new:Npn \int_to_alph:n #1
{\int_to_symbols:nnn \#1 \{ 26 \}\{ 1 \} \{ a \}\{ 2 \} \{ b \}\{ 26 \} \{ z \}\}
\int_to_bin:n \{\text{integer expression}\}

Calculates the value of the \text{integer expression} and places the binary representation of the result in the input stream.

\int_to_hex:n \{\text{integer expression}\}

Calculates the value of the \text{integer expression} and places the hexadecimal (base 16) representation of the result in the input stream. Letters are used for digits beyond 9: lower case letters for \int_to_hex:n and upper case ones for \int_to_Hex:n. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\int_to_oct:n \{\text{integer expression}\}

Calculates the value of the \text{integer expression} and places the octal (base 8) representation of the result in the input stream. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\int_to_base:nn \{\text{integer expression}\} \{\text{base}\}

Calculates the value of the \text{integer expression} and converts it into the appropriate representation in the \text{base}; the later may be given as an integer expression. For bases greater than 10 the higher “digits” are represented by letters from the English alphabet: lower case letters for \int_to_base:n and upper case ones for \int_to_Base:n. The maximum \text{base} value is 36. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\int_to_roman:n \{\text{integer expression}\}

Places the value of the \text{integer expression} in the input stream as Roman numerals, either lower case \int_to_roman:n or upper case \int_to_Roman:n. If the value is negative or zero, the output is empty. The Roman numerals are letters with category code 11 (letter). The letters used are mdclxvi, repeated as needed: the notation with bars (such as \(\overline{V}\) for 5000) is not used. For instance \int_to_roman:n \{8249\} expands to mmmmmmcxlvi.

\int_from_alph:n \{\text{letters}\}

Converts the \text{letters} into the integer (base 10) representation and leaves this in the input stream. The \text{letters} are first converted to a string, with no expansion. Lower and upper case letters from the English alphabet may be used, with “a” equal to 1 through to “z” equal to 26. The function also accepts a leading sign, made of + and -. This is the inverse function of \int_to_alph:n and \int_to_Alph:n.

\textbf{20.9 Converting from other formats to integers}
\int_{\text{from bin}}:n \quad \text{(New: 2014-02-11, Updated: 2014-08-25)}

Converts the \textit{(binary number)} into the integer (base 10) representation and leaves this in the input stream. The \textit{(binary number)} is first converted to a string, with no expansion. The function accepts a leading sign, made of + and −, followed by binary digits. This is the inverse function of \int_{\text{to bin}}:n.

\int_{\text{from hex}}:n \quad \text{(New: 2014-02-11, Updated: 2014-08-25)}

Converts the \textit{(hexadecimal number)} into the integer (base 10) representation and leaves this in the input stream. Digits greater than 9 may be represented in the \textit{(hexadecimal number)} by upper or lower case letters. The \textit{(hexadecimal number)} is first converted to a string, with no expansion. The function also accepts a leading sign, made of + and −. This is the inverse function of \int_{\text{to hex}}:n and \int_{\text{to Hex}}:n.

\int_{\text{from oct}}:n \quad \text{(New: 2014-02-11, Updated: 2014-08-25)}

Converts the \textit{(octal number)} into the integer (base 10) representation and leaves this in the input stream. The \textit{(octal number)} is first converted to a string, with no expansion. The function accepts a leading sign, made of + and −, followed by octal digits. This is the inverse function of \int_{\text{to oct}}:n.

\int_{\text{from roman}}:n \quad \text{(Updated: 2014-08-25)}

Converts the \textit{(roman numeral)} into the integer (base 10) representation and leaves this in the input stream. The \textit{(roman numeral)} is first converted to a string, with no expansion. The \textit{(roman numeral)} may be in upper or lower case; if the numeral contains characters besides mdclxvi or MDCLXVI then the resulting value is −1. This is the inverse function of \int_{\text{to roman}}:n and \int_{\text{to Roman}}:n.

\int_{\text{from base}}:nn \quad \text{(Updated: 2014-08-25)}

Converts the \textit{(number)} expressed in \textit{(base)} into the appropriate value in base 10. The \textit{(number)} is first converted to a string, with no expansion. The \textit{(number)} should consist of digits and letters (either lower or upper case), plus optionally a leading sign. The maximum \textit{(base)} value is 36. This is the inverse function of \int_{\text{to base}}:nn and \int_{\text{to Base}}:nn.

### 20.10 Random integers

\int_{\text{rand}}:nn \quad \text{(New: 2016-12-06, Updated: 2018-04-27)}

Evaluates the two \textit{(integer expressions)} and produces a pseudo-random number between the two (with bounds included). This is not available in older versions of \TeX{}.

\int_{\text{rand}}:n \quad \text{(New: 2018-05-05)}

Evaluates the \textit{(integer expression)} then produces a pseudo-random number between 1 and the \textit{(integerexpr) (included)}. This is not available in older versions of \TeX{}.
20.11 Viewing integers

\int_show:N \int_show:C

Displays the value of the \(\text{integer}\) on the terminal.

\int_show:n

Displays the result of evaluating the \(\text{integer expression}\) on the terminal.

\int_log:N \int_log:C

Writes the value of the \(\text{integer}\) in the log file.

\int_log:n \int_log:{\text{integer expression}}

Writes the result of evaluating the \(\text{integer expression}\) in the log file.

20.12 Constant integers

\c_zero_int \c_one_int

Integer values used with primitive tests and assignments: their self-terminating nature makes these more convenient and faster than literal numbers.

\c_max_int

The maximum value that can be stored as an integer.

\c_max_register_int

Maximum number of registers.

\c_max_char_int

Maximum character code completely supported by the engine.

20.13 Scratch integers

\l_tmpa_int \l_tmpb_int

Scratch integer for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX{}3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_int \g_tmpb_int

Scratch integer for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX{}3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
20.14 Direct number expansion

\int_value:w * \int_value:w (integer)
\int_value:w (integer denotation) (optional space)

Expands the following tokens until an \langle integer \rangle is formed, and leaves a normalized form (no leading sign except for negative numbers, no leading digit 0 except for zero) in the input stream as category code 12 (other) characters. The \langle integer \rangle can consist of any number of signs (with intervening spaces) followed by

- an integer variable (in fact, any \TeX{} register except \texttt{\toks}) or
- explicit digits (or by \texttt{\langle octal digits \rangle} or \texttt{\langle hexadecimal digits \rangle} or \texttt{\langle character \rangle}).

In this last case expansion stops once a non-digit is found; if that is a space it is removed as in \texttt{f}-expansion, and so \texttt{\exp_stop_f:} may be employed as an end marker. Note that protected functions are expanded by this process.

This function requires exactly one expansion to produce a value, and so is suitable for use in cases where a number is required “directly”. In general, \texttt{\int_eval:n} is the preferred approach to generating numbers.

\TeX{}hackers note: This is the \TeX{} primitive \texttt{\number}.

20.15 Primitive conditionals

\texttt{\if_int_compare:w} * \if_int_compare:w (integer_1) (relation) (integer_2)
\langle true code \rangle
\else:
\langle false code \rangle
\fi:

Compare two integers using \langle relation \rangle, which must be one of \texttt{=}, \texttt{<} or \texttt{>} with category code 12. The \texttt{\else:} branch is optional.

\TeX{}hackers note: These are both names for the \TeX{} primitive \texttt{\ifnum}.

\texttt{\if_case:w} * \if_case:w (integer) (case_0)
\or: (case_1)
\or: ...
\or: (default)
\else: (default)
\fi:

Selects a case to execute based on the value of the \langle integer \rangle. The first case ((case_0)) is executed if \langle integer \rangle is 0, the second ((case_1)) if the \langle integer \rangle is 1, etc. The \langle integer \rangle may be a literal, a constant or an integer expression (e.g. using \texttt{\int_eval:n}).

\TeX{}hackers note: These are the \TeX{} primitives \texttt{\ifcase} and \texttt{\or}.
\if_int_odd:w *  \if_int_odd:w \langle tokens \rangle \ (optional space) 
\langle true code \rangle
\else:
\langle true code \rangle
\fi:

Expands \langle tokens \rangle until a non-numeric token or a space is found, and tests whether the resulting \langle integer \rangle is odd. If so, \langle true code \rangle is executed. The \else: branch is optional.

\TeXhackers note: This is the \TeX primitive \ifodd.
Chapter 21

The l3flag package: Expandable flags

Flags are the only data-type that can be modified in expansion-only contexts. This module is meant mostly for kernel use: in almost all cases, booleans or integers should be preferred to flags because they are very significantly faster.

A flag can hold any non-negative value, which we call its \langle height \rangle. In expansion-only contexts, a flag can only be “raised”: this increases the \langle height \rangle by 1. The \langle height \rangle can also be queried expandably. However, decreasing it, or setting it to zero requires non-expandable assignments.

Flag variables are always local. They are referenced by a \langle flag name \rangle such as str_-missing. The \langle flag name \rangle is used as part of \use:c constructions hence is expanded at point of use. It must expand to character tokens only, with no spaces.

A typical use case of flags would be to keep track of whether an exceptional condition has occurred during expandable processing, and produce a meaningful (non-expandable) message after the end of the expandable processing. This is exemplified by l3str-convert, which for performance reasons performs conversions of individual characters expandably and for readability reasons produces a single error message describing incorrect inputs that were encountered.

Flags should not be used without carefully considering the fact that raising a flag takes a time and memory proportional to its height. Flags should not be used unless unavoidable.

21.1 Setting up flags

\flag_new:n \flag_new:n {\langle flag name \rangle}

Creates a new flag with a name given by \langle flag name \rangle, or raises an error if the name is already taken. The \langle flag name \rangle may not contain spaces. The declaration is global, but flags are always local variables. The \langle flag \rangle initially has zero height.

\flag_clear:n \flag_clear:n {\langle flag name \rangle}

The \langle flag \rangle’s height is set to zero. The assignment is local.
\flag_clear_new:n \flag_clear_new:n \{flag name\}
Ensures that the ⟨flag⟩ exists globally by applying \flag_new:n if necessary, then applies \flag_clear:n, setting the height to zero locally.

\flag_show:n \flag_show:n \{\langle flag name\rangle\}
Displays the ⟨flag⟩’s height in the terminal.

\flag_log:n \flag_log:n \{\langle flag name\rangle\}
Writes the ⟨flag⟩’s height to the log file.

21.2 Expandable flag commands

\flag_if_exist_p:n * \flag_if_exist:n \{\langle flag name\rangle\}
This function returns true if the ⟨flag name⟩ references a flag that has been defined previously, and false otherwise.

\flag_if_raised_p:n * \flag_if_raised:n \{\langle flag name\rangle\}
This function returns true if the ⟨flag⟩ has non-zero height, and false if the ⟨flag⟩ has zero height.

\flag_height:n * \flag_height:n \{\langle flag name\rangle\}
Expands to the height of the ⟨flag⟩ as an integer denotation.

\flag_raise:n * \flag_raise:n \{\langle flag name\rangle\}
The ⟨flag⟩’s height is increased by 1 locally.
Chapter 22

The \texttt{l3clist} package
Comma separated lists

Comma lists (in short, \texttt{clist}) contain ordered data where items can be added to the left or right end of the list. This data type allows basic list manipulations such as adding/removing items, applying a function to every item, removing duplicate items, extracting a given item, using the comma list with specified separators, and so on. Sequences (defined in \texttt{l3seq}) are safer, faster, and provide more features, so they should often be preferred to comma lists. Comma lists are mostly useful when interfacing with \LaTeX{} or other code that expects or provides items separated by commas.

Several items can be added at once. To ease input of comma lists from data provided by a user outside an \texttt{\ExplSyntaxOn...\ExplSyntaxOff} block, spaces are removed from both sides of each comma-delimited argument upon input. Blank arguments are ignored, to allow for trailing commas or repeated commas (which may otherwise arise when concatenating comma lists “by hand”). In addition, a set of braces is removed if the result of space-trimming is braced: this allows the storage of any item in a comma list. For instance,

\begin{verbatim}
\clist_new:N \l_my_clist
\clist_put_left:Nn \l_my_clist { -a- , -\{b\} - , c-\d }
\clist_put_right:Nn \l_my_clist { -\{e\} , , \{f\} , }
\end{verbatim}

results in \texttt{\l_my_clist} containing \(a, b, c-\d, e\), \{f\} namely the five items \(a, b, c-\d, e\) and \{f\}. Comma lists normally do not contain empty or blank items so the following gives an empty comma list:

\begin{verbatim}
\clist_clear_new:N \l_my_clist
\clist_set:NN \l_my_clist { , , , } \clist_if_empty:NTF \l_my_clist \{ true \} \{ false \}
\end{verbatim}

and it leaves \texttt{true} in the input stream. To include an “unsafe” item (empty, or one that contains a comma, or starts or ends with a space, or is a single brace group), surround it with braces.

Any \texttt{n}-type token list is a valid comma list input for \texttt{l3clist} functions, which will split the token list at every comma and process the items as described above. On the other hand, \texttt{N}-type functions expect comma list variables, which are particular token list variables in which this processing of items (and removal of blank items) has already

170
occurred. Because comma list variables are token list variables, expanding them once
yields their items separated by commas, and \l\texttt{tl\_show:N} can be
applied to them. (These functions often have \l\texttt{l3\texttt{list}} analogues, which should be preferred.)

Almost all operations on comma lists are noticeably slower than those on sequences
so converting the data to sequences using \texttt{\seq\_set\_from\_clist:Nn} (see \texttt{l3\texttt{seq}}) may be
advisable if speed is important. The exception is that \texttt{\clist\_if\_in:NnTF} and \texttt{\clist\_remove\_duplicates:N}
may be faster than their sequence analogues for large lists. However, these functions
work slowly for “unsafe” items that must be braced, and may pro-
duce errors when their argument contains \{, \} or \# (assuming the usual \TeX category
codes apply). The sequence data type should thus certainly be preferred to comma lists
to store such items.

22.1 Creating and initialising comma lists

\begin{itemize}
\item \texttt{\clist\_new:N} \langle comma list \rangle
\item \texttt{\clist\_new:c} \langle comma list \rangle
\item \texttt{\clist\_const:Nn} \langle\langle comma list\rangle\rangle
\item \texttt{\clist\_const:c} \langle\langle comma list\rangle\rangle
\item \texttt{\clist\_clear:N} \langle comma list \rangle
\item \texttt{\clist\_clear:c} \langle comma list \rangle
\item \texttt{\clist\_gclear:N} \langle comma list \rangle
\item \texttt{\clist\_gclear:c} \langle comma list \rangle
\item \texttt{\clist\_clear\_new:N} \langle comma list \rangle
\item \texttt{\clist\_clear\_new:c} \langle comma list \rangle
\item \texttt{\clist\_gclear\_new:N} \langle comma list \rangle
\item \texttt{\clist\_gclear\_new:c} \langle comma list \rangle
\item \texttt{\clist\_set\_eq:NN} \langle comma list \rangle \langle comma list \rangle
\item \texttt{\clist\_set\_eq:NN} \langle\langle comma list\rangle\rangle \langle\langle comma list\rangle\rangle
\item \texttt{\clist\_set\_eq:NN} \langle\langle comma list\rangle\rangle \langle\langle comma list\rangle\rangle
\item \texttt{\clist\_set\_eq:NN} \langle\langle comma list\rangle\rangle \langle\langle comma list\rangle\rangle
\item \texttt{\clist\_set\_eq:NN} \langle\langle comma list\rangle\rangle \langle\langle comma list\rangle\rangle
\item \texttt{\clist\_set\_from\_seq:NN} \langle comma list \rangle \langle sequence \rangle
\item \texttt{\clist\_set\_from\_seq:NN} \langle\langle comma list\rangle\rangle \langle\langle comma list\rangle\rangle
\item \texttt{\clist\_set\_from\_seq:NN} \langle\langle comma list\rangle\rangle \langle\langle comma list\rangle\rangle
\item \texttt{\clist\_set\_from\_seq:NN} \langle\langle comma list\rangle\rangle \langle\langle comma list\rangle\rangle
\end{itemize}

\begin{itemize}
\item Creates a new \langle comma list \rangle or raises an error if the name is already taken. The declaration
is global. The \langle comma list \rangle initially contains no items.
\item Creates a new constant \langle\langle comma list\rangle\rangle or raises an error if the name is already taken. The
value of the \langle\langle comma list\rangle\rangle is set globally to the \langle comma list \rangle.
\item Clears all items from the \langle comma list \rangle.
\item Ensures that the \langle comma list \rangle exists globally by applying \texttt{\clist\_new:N} if necessary,
them \texttt{\clist\_\(g\)\texttt{clear}:\texttt{N}} to leave the list empty.
\item Sets the content of \langle comma list\rangle equal to that of \langle comma list\rangle. To set a token list
variable equal to a comma list variable, use \texttt{\tl\_set\_eq:NN}. Conversely, setting a comma
list variable to a token list is unadvisable unless one checks space-trimming and related
issues.
\item Converts the data in the \langle sequence \rangle into a \langle comma list \rangle: the original \langle sequence \rangle is un-
changed. Items which contain either spaces or commas are surrounded by braces.
\end{itemize}

\newpage

171
Concatenates the content of \langle comma list_2 \rangle and \langle comma list_3 \rangle together and saves the result in \langle comma list_1 \rangle. The items in \langle comma list_2 \rangle are placed at the left side of the new comma list.

\clist_if_exist_p:N \langle comma list \rangle
\clist_if_exist:NTF \langle comma list \rangle{\langle true code \rangle}{\langle false code \rangle}

Tests whether the \langle comma list \rangle is currently defined. This does not check that the \langle comma list \rangle really is a comma list.

\clist_set:Nn \langle comma list \rangle{\langle item_1 \rangle, \ldots, \langle item_n \rangle}
\clist_gset:Nn \langle comma list \rangle{\langle NV \rangle|\langle No \rangle|\langle Nx \rangle|\langle cn \rangle|\langle cV \rangle|\langle co \rangle|\langle cx \rangle}

Sets \langle comma list \rangle to contain the \langle items \rangle, removing any previous content from the variable. Blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. To store some \langle tokens \rangle as a single \langle item \rangle even if the \langle tokens \rangle contain commas or spaces, add a set of braces: \clist_set:Nn \langle comma list \rangle{\langle \langle tokens \rangle \rangle}.

\clist_put_left:Nn \langle comma list \rangle{\langle item_1 \rangle, \ldots, \langle item_n \rangle}
\clist_gput_left:Nn \langle comma list \rangle{\langle NV \rangle|\langle No \rangle|\langle Nx \rangle|\langle cn \rangle|\langle cV \rangle|\langle co \rangle|\langle cx \rangle}

Appends the \langle items \rangle to the left of the \langle comma list \rangle. Blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. To append some \langle tokens \rangle as a single \langle item \rangle even if the \langle tokens \rangle contain commas or spaces, add a set of braces: \clist_put_left:Nn \langle comma list \rangle{\langle \langle tokens \rangle \rangle}.

\clist_put_right:Nn \langle comma list \rangle{\langle item_1 \rangle, \ldots, \langle item_n \rangle}
\clist_gput_right:Nn \langle comma list \rangle{\langle NV \rangle|\langle No \rangle|\langle Nx \rangle|\langle cn \rangle|\langle cV \rangle|\langle co \rangle|\langle cx \rangle}

Appends the \langle items \rangle to the right of the \langle comma list \rangle. Blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. To append some \langle tokens \rangle as a single \langle item \rangle even if the \langle tokens \rangle contain commas or spaces, add a set of braces: \clist_put_right:Nn \langle comma list \rangle{\langle \langle tokens \rangle \rangle}.

\section{Adding data to comma lists}

Sets \langle comma list \rangle to contain the \langle items \rangle, removing any previous content from the variable. Blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. To store some \langle tokens \rangle as a single \langle item \rangle even if the \langle tokens \rangle contain commas or spaces, add a set of braces: \clist_set:Nn \langle comma list \rangle{\langle \langle tokens \rangle \rangle}.
22.3 Modifying comma lists

While comma lists are normally used as ordered lists, it may be necessary to modify the content. The functions here may be used to update comma lists, while retaining the order of the unaffected entries.

\clist_remove_duplicates:N  ⟨comma list⟩
\clist_remove_duplicates:c
\clist_gremove_duplicates:N
\clist_gremove_duplicates:c

Removes duplicate items from the ⟨comma list⟩, leaving the left most copy of each item in the ⟨comma list⟩. The ⟨item⟩ comparison takes place on a token basis, as for \tl_if_eq:nnTF.

\texttt{\textsc{Texhacker} note:} This function iterates through every item in the ⟨comma list⟩ and does a comparison with the ⟨items⟩ already checked. It is therefore relatively slow with large comma lists. Furthermore, it may fail if any of the items in the ⟨comma list⟩ contains ⟨,⟩, or ⟨#⟩ (assuming the usual \texttt{\textsc{TeX}} category codes apply).

\clist_remove_all:Nn ⟨comma list⟩ ⟨item⟩
\clist_remove_all:cn
\clist_gremove_all:Nn
\clist_gremove_all:cn

Removes every occurrence of ⟨item⟩ from the ⟨comma list⟩. The ⟨item⟩ comparison takes place on a token basis, as for \tl_if_eq:nnTF.

\texttt{\textsc{Texhacker} note:} The function may fail if the ⟨item⟩ contains ⟨,⟩, or ⟨#⟩ (assuming the usual \texttt{\textsc{TeX}} category codes apply).

\clist_reverse:N  ⟨comma list⟩
\clist_reverse:c
\clist_greverse:N
\clist_greverse:c

New: 2014-07-18

Reverses the order of items stored in the ⟨comma list⟩.

\clist_reverse:n ⟨comma list⟩

New: 2014-07-18

Leaves the items in the ⟨comma list⟩ in the input stream in reverse order. Contrarily to other what is done for other \texttt{n-type ⟨comma list⟩} arguments, braces and spaces are preserved by this process.

\texttt{\textsc{Texhacker} note:} The result is returned within \texttt{\texttt{unexpanded}}, which means that the comma list does not expand further when appearing in an \texttt{x-type} or \texttt{e-type} argument expansion.

\clist_sort:Nn ⟨clist var⟩ ⟨⟨comparison code⟩⟩
\clist_sort:cn
\clist_gsort:Nn
\clist_gsort:cn

New: 2017-02-06

Sorts the items in the ⟨clist var⟩ according to the ⟨comparison code⟩, and assigns the result to ⟨clist var⟩. The details of sorting comparison are described in Section 6.1.
22.4 Comma list conditionals

\clist_if_empty_p:N  \clist_if_empty_p:N (comma list)
\clist_if_empty_p:c  \clist_if_empty_p:c (comma list)
\clist_if_empty:NTF \clist_if_empty:NTF \{\true\ code\}\ \{(false\ code)\}

Tests if the \textit{comma list} is empty (containing no items).

\clist_if_empty_p:n  \clist_if_empty_p:n \{\comma\ list\}
\clist_if_empty:nTF \clist_if_empty:nTF \{\comma\ list\}\ \{(true\ code)\}\ \{(false\ code)\}

Tests if the \textit{comma list} is empty (containing no items). The rules for space trimming are as for other \textit{n}-type comma-list functions, hence the comma list \{-, -, -\} (without outer braces) is empty, while \{-, \}, \{\} (without outer braces) contains one element, which happens to be empty: the comma-list is not empty.

\clist_if_in:NnTF \clist_if_in:NnTF \{\comma\ list\}\ \{\item\}\ \{(true\ code)\}\ \{(false\ code)\}
\clist_if_in:NN \clist_if_in:NN \texttt{(NV|No|cn|cV|co|cV)} \texttt{TF}
\clist_if_in:nn \clist_if_in:nn \texttt{(nV|no)} \texttt{TF}
\clist_if_in:TF  \clist_if_in:TF
\clist_if_in:c \clist_if_in:c
\clist_if_in:n \clist_if_in:n

\texttt{Updated: 2011-09-06}

Tests if the \textit{item} is present in the \textit{comma list}. In the case of an \textit{n}-type \textit{comma list}, the usual rules of space trimming and brace stripping apply. Hence,

\clist_if_in:nnTF { a, \{b\}, \{\} , c } { b } \{true\} \{false\}
yields \texttt{true}.

\textbf{\LaTeX hack:} The function may fail if the \textit{item} contains \{, \}, or \# (assuming the usual \LaTeX category codes apply).

22.5 Mapping over comma lists

The functions described in this section apply a specified function to each item of a comma list. All mappings are done at the current group level, \textit{i.e.} any local assignments made by the \textit{function} or \textit{code} discussed below remain in effect after the loop.

When the comma list is given explicitly, as an \textit{n}-type argument, spaces are trimmed around each item. If the result of trimming spaces is empty, the item is ignored. Otherwise, if the item is surrounded by braces, one set is removed, and the result is passed to the mapped function. Thus, if the comma list that is being mapped is \{a, \{b\}, \{\} , c\}, then the arguments passed to the mapped function are \texttt{a}, \texttt{\{b\}}, an empty argument, and \texttt{c}.

When the comma list is given as an \textit{N}-type argument, spaces have already been trimmed on input, and items are simply stripped of one set of braces if any. This case is more efficient than using \textit{n}-type comma lists.

\clist_map_function:NN \clist_map_function:NN \{\comma\ list\}\ \{\function\}
\clist_map_function:CN \clist_map_function:CN
\clist_map_function:nN \clist_map_function:nN

\texttt{Updated: 2012-06-29}

Applies \textit{function} to every \textit{item} stored in the \textit{comma list}. The \textit{function} receives one argument for each iteration. The \textit{items} are returned from left to right. The function \clist_map_inline:Nn is in general more efficient than \clist_map_function:NN.
\clist_map_inline:Nn \clist_map_inline:cn \clist_map_inline:nNn

\clist_map_variable:NNn \clist_map_variable:cNn \clist_map_variable:nNn

\clist_map_tokens:Nn \clist_map_tokens:cn \clist_map_tokens:nn

\clist_map_break: ∗

\clist_map_inline:Nn \clist_map_inline:Nn \clist_map_inline:nn

\clist_map_variable:NNn \clist_map_variable:cNn \clist_map_variable:nNn

\clist_map_tokens:Nn \clist_map_tokens:cn \clist_map_tokens:nn

\clist_map_break:

\clist_map_inline:Nn \l_my_clist
{ \str_if_eq:nnTF { #1 } { bingo } { \clist_map_break: } % Do something useful }

Use outside of a \clist_map_... scenario leads to low level \TeX{} errors.

\TeX{}hackers note: When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.
\clist_map_break:n \(\langle\text{code}\rangle\)}

Used to terminate a \clist_map\_... function before all entries in the \assoc{comma list} have been processed, inserting the \assoc{code} after the mapping has ended. This normally takes place within a conditional statement, for example

\clist_map_inline:Nn \l_my_clist
{\str_if_eq:nnTF { #1 } { bingo } { \clist_map_break:n \{ <\text{code}> \} }
{ % Do something useful
}
}

Use outside of a \clist_map\_... scenario leads to low level \TeX{} errors.

\TeX{}hackers note: When the mapping is broken, additional tokens may be inserted before the \assoc{code} is inserted into the input stream. This depends on the design of the mapping function.

\clist_count:N \(\langle\text{comma list}\rangle\)

Leaves the number of items in the \assoc{comma list} in the input stream as an \assoc{integer denotation}. The total number of items in a \assoc{comma list} includes those which are duplicates, \ie{} every item in a \assoc{comma list} is counted.

22.6 Using the content of comma lists directly

\clist_use:Nnnn \clist_use:cnnn \clist_use:nnnn
\(\langle\text{clist var}\rangle\) \(\langle\text{separator between two}\rangle\) \(\langle\text{separator between more than two}\rangle\) \(\langle\text{separator between final two}\rangle\)

Places the contents of the \assoc{clist var} in the input stream, with the appropriate \assoc{separator} between the items. Namely, if the comma list has more than two items, the \assoc{separator between more than two} is placed between each pair of items except the last, for which the \assoc{separator between final two} is used. If the comma list has exactly two items, then they are placed in the input stream separated by the \assoc{separator between two}. If the comma list has a single item, it is placed in the input stream, and a comma list with no items produces no output. An error is raised if the variable does not exist or if it is invalid.

For example,

\clist_set:Nn \l_tmpa_clist { a , b , , c , \{de\} , f }
\clist_use:Nnnn \l_tmpa_clist { -and- } { , , } { , -and- }

inserts “a, b, c, de, and f” in the input stream. The first separator argument is not used in this case because the comma list has more than 2 items.

\TeX{}hackers note: The result is returned within the \\unexpanded primitive (\exp_not:n), which means that the \assoc{items} do not expand further when appearing in an \textit{x}-type argument expansion.
\clist_use:Nn \clist_use:cn \newline
New: 2013-05-26

\clist_use:Nn \clist_use:cn
\clist_use:nnnn \clist_use:nn
\cline{3-4}
New: 2021-05-10
Updated: 2021-05-10

\clist_use:nnnn \clist_use:nn
\clist_get:NN \clist_get:cN \clist_get:NN \clist_get:cN
\cline{3-4}
New: 2012-05-14
Updated: 2019-02-16

\clist_pop:NN \clist_pop:cN
\cline{3-4}
Updated: 2011-09-06

\clist_use:Nn \clist_use:cn \newline
New: 2013-05-26

Places the contents of the \clist var in the input stream, with the \separator between the items. If the comma list has a single item, it is placed in the input stream, and a comma list with no items produces no output. An error is raised if the variable does not exist or if it is invalid.

For example,

\clist_set:Nn \l_tmpa_clist \{ a , b , , c , \{de\} , f \}
\clist_use:Nn \l_tmpa_clist \{ -and- \}

inserts “a and b and c and de and f” in the input stream.

\textbf{\TeXhackers note:} The result is returned within the \unexpanded primitive (\exp_not:n), which means that the \items do not expand further when appearing in an x-type argument expansion.

\clist_use:nnnn \clist_use:nn
\cline{3-4}
New: 2021-05-10
Updated: 2021-05-10

\clist_get:NN \clist_get:cN \clist_get:NN \clist_get:cN
\cline{3-4}
New: 2012-05-14
Updated: 2019-02-16

\clist_pop:NN \clist_pop:cN
\cline{3-4}
Updated: 2011-09-06

\section{22.7 Comma lists as stacks}

Comma lists can be used as stacks, where data is pushed to and popped from the top of the comma list. (The left of a comma list is the top, for performance reasons.) The stack functions for comma lists are not intended to be mixed with the general ordered data functions detailed in the previous section: a comma list should either be used as an ordered data type or as a stack, but not in both ways.
\clist_gpop:NN \clist_gpop:cN

Pops the left-most item from a \langle \text{comma list} \rangle into the \langle \text{token list variable} \rangle, i.e. removes the item from the comma list and stores it in the \langle \text{token list variable} \rangle. The \langle \text{comma list} \rangle is modified globally, while the assignment of the \langle \text{token list variable} \rangle is local.

\clist_gpop:NNTF \clist_gpop:cN
\clist_pop:NNTF \clist_pop:cN

\clist_gpop:NNTF \clist_gpop:cN
\clist_pop:NNTF \clist_pop:cN

\clist_gpop:NNTF \clist_gpop:cN
\clist_pop:NNTF \clist_pop:cN

If the \langle \text{comma list} \rangle is empty, leaves the \langle \text{false code} \rangle in the input stream. The value of the \langle \text{token list variable} \rangle is not defined in this case and should not be relied upon. If the \langle \text{comma list} \rangle is non-empty, pops the top item from the \langle \text{comma list} \rangle in the \langle \text{token list variable} \rangle, i.e. removes the item from the \langle \text{comma list} \rangle. Both the \langle \text{comma list} \rangle and the \langle \text{token list variable} \rangle are assigned locally.

\clist_gpush:Nn \clist_gpush:cn
\clist_gpush:Nn \clist_gpush:cn

\clist_gpush:Nn \clist_gpush:cn
\clist_gpush:Nn \clist_gpush:cn

Adds the \{\langle \text{items} \rangle\} to the top of the \langle \text{comma list} \rangle. Spaces are removed from both sides of each item as for any n-type comma list.

22.8 Using a single item

\clist_item:Nn \clist_item:cn \clist_item:nn
\clist_item:Nn \clist_item:cn \clist_item:nn

Indexing items in the \langle \text{comma list} \rangle from 1 at the top (left), this function evaluates the \langle \text{integer expression} \rangle and leaves the appropriate item from the comma list in the input stream. If the \langle \text{integer expression} \rangle is negative, indexing occurs from the bottom (right) of the comma list. When the \langle \text{integer expression} \rangle is larger than the number of items in the \langle \text{comma list} \rangle (as calculated by \clist_count:N) then the function expands to nothing.

\TeXhackers note: The result is returned within the \texttt{\textbackslash unexpanded} primitive (\texttt{\textbackslash exp_not:n}), which means that the \langle \text{item} \rangle does not expand further when appearing in an x-type argument expansion.
\clist_rand_item:N \clist_rand_item:c \clist_rand_item:n

Selects a pseudo-random item of the (comma list). If the (comma list) has no item, the result is empty.

\textbf{\LaTeX{}hackers note:} The result is returned within the \texttt{\unexpanded} primitive (\texttt{\exp_not:n}), which means that the \texttt{item} does not expand further when appearing in an x-type argument expansion.

\section{Viewing comma lists}

\clist_show:N \clist_show:c

Displays the entries in the (comma list) in the terminal.

\clist_show:n

Displays the entries in the comma list in the terminal.

\clist_log:N \clist_log:c

Writes the entries in the (comma list) in the log file. See also \texttt{\clist_show:N} which displays the result in the terminal.

\clist_log:n

Writes the entries in the comma list in the log file. See also \texttt{\clist_show:n} which displays the result in the terminal.

\section{Constant and scratch comma lists}

\c_empty_clist

Constant that is always empty.

\l_tmpa_clist \l_tmpb_clist

Scratch comma lists for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX{}3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_clist \g_tmpb_clist

Scratch comma lists for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX{}3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Chapter 23

The l3token package
Token manipulation

This module deals with tokens. Now this is perhaps not the most precise description so let’s try with a better description: When programming in TeX, it is often desirable to know just what a certain token is: is it a control sequence or something else. Similarly one often needs to know if a control sequence is expandable or not, a macro or a primitive, how many arguments it takes etc. Another thing of great importance (especially when it comes to document commands) is looking ahead in the token stream to see if a certain character is present and maybe even remove it or disregard other tokens while scanning. This module provides functions for both and as such has two primary function categories: \token_ for anything that deals with tokens and \peek_ for looking ahead in the token stream.

Most functions we describe here can be used on control sequences, as those are tokens as well.

It is important to distinguish two aspects of a token: its “shape” (for lack of a better word), which affects the matching of delimited arguments and the comparison of token lists containing this token, and its “meaning”, which affects whether the token expands or what operation it performs. One can have tokens of different shapes with the same meaning, but not the converse.

For instance, \if:w, \if_charcode:w, and \tex_if:D are three names for the same internal operation of TeX, namely the primitive testing the next two characters for equality of their character code. They have the same meaning hence behave identically in many situations. However, TeX distinguishes them when searching for a delimited argument. Namely, the example function \show_until_if:w defined below takes everything until \if:w as an argument, despite the presence of other copies of \if:w under different names.

\cs_new:Npn \show_until_if:w #1 \if:w { \tl_show:n {#1} } \show_until_if:w \tex_if:D \if_charcode:w \if:w

A list of all possible shapes and a list of all possible meanings are given in section 23.7.
23.1 Creating character tokens

\char_set_active_eq:NN \langle char \rangle \langle function \rangle

Sets the behaviour of the \langle char \rangle in situations where it is active (category code 13) to be equivalent to that of the \langle function \rangle. The category code of the \langle char \rangle is unchanged by this process. The \langle function \rangle may itself be an active character.

\char_set_active_eq:nN \langle integer expression \rangle \langle function \rangle

Sets the behaviour of the \langle char \rangle which has character code as given by the \langle integer expression \rangle in situations where it is active (category code 13) to be equivalent to that of the \langle function \rangle. The category code of the \langle char \rangle is unchanged by this process. The \langle function \rangle may itself be an active character.

\char_generate:nn \langle charcode \rangle \langle catcode \rangle

Generates a character token of the given \langle charcode \rangle and \langle catcode \rangle (both of which may be integer expressions). The \langle catcode \rangle may be one of:

- 1 (begin group)
- 2 (end group)
- 3 (math toggle)
- 4 (alignment)
- 6 (parameter)
- 7 (math superscript)
- 8 (math subscript)
- 11 (letter)
- 12 (other)
- 13 (active)

and other values raise an error. The \langle charcode \rangle may be any one valid for the engine in use. Active characters cannot be generated in older versions of \TeX. Another way to build token lists with unusual category codes is \regex_replace:nnN \{.*\} \{\langle replacement \rangle\} \{tl var\}.

\TeXhackers note: Exactly two expansions are needed to produce the character.
\char_lowercase:N \char_uppercase:N \char_titlecase:N \char_foldcase:N \char_str_lowercase:N \char_str_uppercase:N \char_str_titlecase:N \char_str_foldcase:N

\char_lowercase:N \char_uppercase:N \char_titlecase:N \char_foldcase:N \char_str_lowercase:N \char_str_uppercase:N \char_str_titlecase:N \char_str_foldcase:N

\c_catcode_other_space_tl

\char_set_catcode_escape:N \char_set_catcode_letter:N \char_set_catcode_group_begin:N \char_set_catcode_group_end:N \char_set_catcode_math_toggle:N \char_set_catcode_alignment:N \char_set_catcode_end_line:N \char_set_catcode_parameter:N \char_set_catcode_math_superscript:N \char_set_catcode_math_subscript:N \char_set_catcode_ignore:N \char_set_catcode_space:N \char_set_catcode_letter:N \char_set_catcode_other:N \char_set_catcode_active:N \char_set_catcode_comment:N \char_set_catcode_invalid:N

\char_lowercase:N \langle \char \rangle

Converts the \langle \char \rangle to the equivalent case-changed character as detailed by the function name (see \texttt{\str_foldcase:n} and \texttt{\text_titlecase:n} for details of these terms). The case mapping is carried out with no context-dependence (cf. \texttt{\text_uppercase:n}, etc.) The \texttt{str} versions always generate “other” (category code 12) characters, whilst the standard versions generate characters with the category code of the \langle \char \rangle (i.e. only the character code changes).

\c_catcode_other_space_tl

Token list containing one character with category code 12, (“other”), and character code 32 (space).

23.2 Manipulating and interrogating character tokens

\char_set_catcode_escape:N \char_set_catcode_letter:N \langle \texttt{character} \rangle

Sets the category code of the \langle \texttt{character} \rangle to that indicated in the function name. Depending on the current category code of the \langle \texttt{token} \rangle the escape token may also be needed:

\char_set_catcode_other:N \%

The assignment is local.
Sets the category code of the (character) which has character code as given by the (integer expression). This version can be used to set up characters which cannot otherwise be given (cf. the \textfamilyname N-type variants). The assignment is local.

\char_set_catcode:nn \char_set_catcode:nn \{\textvar{intexpr1}\} \{\textvar{intexpr2}\}

These functions set the category code of the (character) which has character code as given by the (integer expression). The first (integer expression) is the character code and the second is the category code to apply. The setting applies within the current \TeX{} group. In general, the symbolic functions \char_set_catcode\textunderscore\textfamilyname{type} should be preferred, but there are cases where these lower-level functions may be useful.

\char_value_catcode:n
\char_value_catcode:n \{\textvar{integer expression}\}

Expands to the current category code of the (character) with character code given by the (integer expression).

\char_show_value_catcode:n
\char_show_value_catcode:n \{\textvar{integer expression}\}

Displays the current category code of the (character) with character code given by the (integer expression) on the terminal.

\char_set_lccode:nn
\char_set_lccode:nn \{\textvar{intexpr1}\} \{\textvar{intexpr2}\}

Sets up the behaviour of the (character) when found inside \text_lowercase:n, such that (character1) will be converted into (character2). The two (characters) may be specified using an (integer expression) for the character code concerned. This may include the \TeX{} ’(character) method for converting a single character into its character code:

\char_set_lccode:nn \{\textbackslash A\} \{\textbackslash a\} % Standard behaviour
\char_set_lccode:nn \{\textbackslash A\} \{\textbackslash A + 32\}
\char_set_lccode:nn \{50\} \{60\}

The setting applies within the current \TeX{} group.
\char_value_lccode:n \{\langle integer expression\rangle\} \n Expands to the current lower case code of the \langle character\rangle with character code given by the \langle integer expression\rangle.

\char_value_lccode:n \star \{\langle integer expression\rangle\}

\char_show_value_lccode:n \{\langle integer expression\rangle\} \n Displays the current lower case code of the \langle character\rangle with character code given by the \langle integer expression\rangle on the terminal.

\char_set_uccode:nn \{\langle intexpr\rangle\} \{\langle intexpr\rangle\}

\char_value_uccode:n \{\langle integer expression\rangle\} \n Expands to the current upper case code of the \langle character\rangle with character code given by the \langle integer expression\rangle.

\char_value_uccode:n \star \{\langle integer expression\rangle\}

\char_show_value_uccode:n \{\langle integer expression\rangle\} \n Displays the current upper case code of the \langle character\rangle with character code given by the \langle integer expression\rangle on the terminal.

\char_set_mathcode:nn \{\langle intexpr\rangle\} \{\langle intexpr\rangle\}

\char_value_mathcode:n \{\langle integer expression\rangle\} \n Expands to the current math code of the \langle character\rangle with character code given by the \langle integer expression\rangle.

\char_value_mathcode:n \star \{\langle integer expression\rangle\}

\char_set_sfcode:nn \{\langle intexpr\rangle\} \{\langle intexpr\rangle\}

\char_set_sfcode:nn \{\langle intexpr\rangle\} \{\langle intexpr\rangle\} \n This function sets up the space factor for the \langle character\rangle. The \langle character\rangle is specified as an \langle integer expression\rangle which will be used as the character code of the relevant character. The setting applies within the current \TeX{} group.
\char_value_sfcode:n \{\texttt{integer expression}\} \* 

Expands to the current space factor for the \langle character \rangle with character code given by the \langle integer expression \rangle.

\char_show_value_sfcode:n \{\texttt{integer expression}\}

Displays the current space factor for the \langle character \rangle with character code given by the \langle integer expression \rangle on the terminal.

\l_char_active_seq

\begin{verbatim}
Rev: 2012-01-23
Updated: 2015-11-11
\end{verbatim}

Used to track which tokens may require special handling at the document level as they are (or have been at some point) of category \langle active \rangle (catcode 13). Each entry in the sequence consists of a single escaped token, for example \texttt{\textbackslash -}. Active tokens should be added to the sequence when they are defined for general document use.

\l_char_special_seq

\begin{verbatim}
Rev: 2012-01-23
Updated: 2015-11-11
\end{verbatim}

Used to track which tokens will require special handling when working with verbatim-like material at the document level as they are not of categories \langle letter \rangle (catcode 11) or \langle other \rangle (catcode 12). Each entry in the sequence consists of a single escaped token, for example \texttt{\textbackslash \textbackslash} for the backslash or \texttt{\textbackslash \{} for an opening brace. Escaped tokens should be added to the sequence when they are defined for general document use.

## 23.3 Generic tokens

These are implicit tokens which have the category code described by their name. They are used internally for test purposes but are also available to the programmer for other uses.

\begin{verbatim}
\c_group_begin_token
\c_group_end_token
\c_math_toggle_token
\c_alignment_token
\c_parameter_token
\c_math_superscript_token
\c_math_subscript_token
\c_space_token
\end{verbatim}

These are implicit tokens which have the category code described by their name. They are used internally for test purposes and should not be used other than for category code tests.

\begin{verbatim}
\c_catcode_letter_token
\c_catcode_other_token
\end{verbatim}

A token list containing an active token. This is used internally for test purposes and should not be used other than in appropriately-constructed category code tests.
23.4 Converting tokens

\token_to_meaning:N \token_to_meaning:c

Inserts the current meaning of the \textlangle token\textrangle into the input stream as a series of characters of category code 12 (other). This is the primitive \TeX description of the \textlangle token\textrangle, thus for example both functions defined by \cs_set_nopar:Npn and token list variables defined using \tl_new:N are described as macros.

\TeXhackersnote: This is the \TeX primitive \meaning. The \textlangle token\textrangle can thus be an explicit space tokens or an explicit begin-group or end-group character token (\{ or \} when normal \TeX category codes apply) even though these are not valid N-type arguments.

\token_to_str:N \token_to_str:c

Converts the given \textlangle token\textrangle into a series of characters with category code 12 (other). If the \textlangle token\textrangle is a control sequence, this will start with the current escape character with category code 12 (the escape character is part of the \textlangle token\textrangle). This function requires only a single expansion.

\TeXhackersnote: \token_to_str:N is the \TeX primitive \string renamed. The \textlangle token\textrangle can thus be an explicit space tokens or an explicit begin-group or end-group character token (\{ or \} when normal \TeX category codes apply) even though these are not valid N-type arguments.

23.5 Token conditionals

\token_if_group_begin_p:N \token_if_group_begin:NTF

Tests if \textlangle token\textrangle has the category code of a begin group token (\{ when normal \TeX category codes are in force). Note that an explicit begin group token cannot be tested in this way, as it is not a valid N-type argument.

\token_if_group_end_p:N \token_if_group_end:NTF

Tests if \textlangle token\textrangle has the category code of an end group token (\} when normal \TeX category codes are in force). Note that an explicit end group token cannot be tested in this way, as it is not a valid N-type argument.

\token_if_math_toggle_p:N \token_if_math_toggle:NTF

Tests if \textlangle token\textrangle has the category code of a math shift token ($ when normal \TeX category codes are in force).
Tests if \( \langle \text{token} \rangle \) has the category code of an alignment token (& when normal TeX category codes are in force).

Tests if \( \langle \text{token} \rangle \) has the category code of a macro parameter token (# when normal TeX category codes are in force).

Tests if \( \langle \text{token} \rangle \) has the category code of a superscript token (^ when normal TeX category codes are in force).

Tests if \( \langle \text{token} \rangle \) has the category code of a subscript token (_) when normal TeX category codes are in force).

Tests if \( \langle \text{token} \rangle \) has the category code of a space token. Note that an explicit space token with character code 32 cannot be tested in this way, as it is not a valid N-type argument.

Tests if \( \langle \text{token} \rangle \) has the category code of a letter token.

Tests if \( \langle \text{token} \rangle \) has the category code of an “other” token.

Tests if \( \langle \text{token} \rangle \) has the category code of an active character.

Tests if the two \( \langle \text{tokens} \rangle \) have the same category code.

Tests if the two \( \langle \text{tokens} \rangle \) have the same character code.
\token_if_eq_meaning_{p:NN} * \token_if_eq_meaning_{p:NN {token}_1 {token}_2}
\token_if_eq_meaning_{NNTF} * \token_if_eq_meaning_{NNTF {token}_1 {token}_2} \{true code\} \{false code\}

Tests if the two (tokens) have the same meaning when expanded.

\token_if_macro_{p:N} * \token_if_macro_{p:N {token}_1}
\token_if_macro_{NTF} * \token_if_macro_{NTF {token}_1} \{true code\} \{false code\}

Tests if the (token) is a \TeX\ macro.

\token_if_cs_{p:N} * \token_if_cs_{p:N {token}_1}
\token_if_cs_{NTF} * \token_if_cs_{NTF {token}_1} \{true code\} \{false code\}

Tests if the (token) is a control sequence.

\token_if_expandable_{p:N} * \token_if_expandable_{p:N {token}_1}
\token_if_expandable_{NTF} * \token_if_expandable_{NTF {token}_1} \{true code\} \{false code\}

Tests if the (token) is expandable. This test returns (false) for an undefined token.

\token_if_long_macro_{p:N} * \token_if_long_macro_{p:N {token}_1}
\token_if_long_macro_{NTF} * \token_if_long_macro_{NTF {token}_1} \{true code\} \{false code\}

Tests if the (token) is a long macro.

\token_if_protected_macro_{p:N} * \token_if_protected_macro_{p:N {token}_1}
\token_if_protected_macro_{NTF} * \token_if_protected_macro_{NTF {token}_1} \{true code\} \{false code\}

Tests if the (token) is a protected macro: for a macro which is both protected and long this returns false.

\token_if_protected_long_macro_{p:N} * \token_if_protected_long_macro_{p:N {token}_1}
\token_if_protected_long_macro_{NTF} * \token_if_protected_long_macro_{NTF {token}_1} \{true code\} \{false code\}

Tests if the (token) is a protected long macro.

\token_if_chardef_{p:N} * \token_if_chardef_{p:N {token}_1}
\token_if_chardef_{NTF} * \token_if_chardef_{NTF {token}_1} \{true code\} \{false code\}

Tests if the (token) is defined to be a chardef.

\TeXhackers\ note: Booleans, boxes and small integer constants are implemented as \chardefs.

\token_if_mathchardef_{p:N} * \token_if_mathchardef_{p:N {token}_1}
\token_if_mathchardef_{NTF} * \token_if_mathchardef_{NTF {token}_1} \{true code\} \{false code\}

Tests if the (token) is defined to be a mathchardef.
\token_if_font_selection_p:N * \token_if_font_selection:N (token) \token_if_font_selection:NTF (token) \token_if_font_selection:N (true code) \token_if_font_selection:N (false code) \token_if_font_selection_p:N

Tests if the (token) is defined to be a font selection command.

\token_if_dim_register_p:N * \token_if_dim_register_p:N (token) \token_if_dim_register:NTF (token) \token_if_dim_register:N (true code) \token_if_dim_register:N (false code) \token_if_dim_register_p:N

Tests if the (token) is defined to be a dimension register.

\token_if_int_register_p:N * \token_if_int_register_p:N (token) \token_if_int_register:NTF (token) \token_if_int_register:N (true code) \token_if_int_register:N (false code) \token_if_int_register_p:N

Tests if the (token) is defined to be an integer register.

\TeXhacks{Note} Constant integers may be implemented as integer registers, \chardefs, or \mathchardefs depending on their value.

\token_if_muskip_register_p:N * \token_if_muskip_register_p:N (token) \token_if_muskip_register:NTF (token) \token_if_muskip_register:N (true code) \token_if_muskip_register:N (false code) \token_if_muskip_register_p:N

Tests if the (token) is defined to be a muskip register.

\token_if_skip_register_p:N * \token_if_skip_register_p:N (token) \token_if_skip_register:NTF (token) \token_if_skip_register:N (true code) \token_if_skip_register:N (false code) \token_if_skip_register_p:N

Tests if the (token) is defined to be a skip register.

\token_if_toks_register_p:N * \token_if_toks_register_p:N (token) \token_if_toks_register:NTF (token) \token_if_toks_register:N (true code) \token_if_toks_register:N (false code) \token_if_toks_register_p:N

Tests if the (token) is defined to be a toks register (not used by \LaTeX3).

\token_if_primitive_p:N * \token_if_primitive_p:N (token) \token_if_primitive:NTF (token) \token_if_primitive:N (true code) \token_if_primitive:N (false code) \token_if_primitive_p:N

Tests if the (token) is an engine primitive. In \LuaTeX this includes primitive-like commands defined using \token.set_lua.
This function compares the \textit{(test token)} in turn with each of the \textit{(token cases)}. If the two are equal (as described for \texttt{token_if_eq_catcode:NNTF}, \texttt{token_if_eq_charcode:NNTF}, and \texttt{token_if_eq_meaning:NNTF}, respectively) then the associated \textit{(code)} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{(true code)} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{(false code)} is inserted. The functions \texttt{token_case_catcode:Nn}, \texttt{token_case_charcode:Nn}, and \texttt{token_case_meaning:Nn}, which do nothing if there is no match, are also available.

23.6 Peeking ahead at the next token

There is often a need to look ahead at the next token in the input stream while leaving it in place. This is handled using the “peek” functions. The generic \texttt{peek_after:Nw} is provided along with a family of predefined tests for common cases. As peeking ahead does \textit{not} skip spaces the predefined tests include both a space-respecting and space-skipping version. In addition, using \texttt{peek_analysis_map_inline:n}, one can map through the following tokens in the input stream and repeatedly perform some tests.

\begin{verbatim}
\peek_after:Nw \peek_after:Nw (function) \token
Locally sets the test variable \texttt{l_peek_token} equal to \textit{(token)} (as an implicit token, \textit{not} as a token list), and then expands the \textit{(function)}. The \textit{(token)} remains in the input stream as the next item after the \textit{(function)}. The \textit{(token)} here may be \texttt{\textbackslash i}, \{ or \} (assuming normal \LaTeX{} category codes), \textit{i.e.} it is not necessarily the next argument which would be grabbed by a normal function.

\peek_gafter:Nw \peek_gafter:Nw (function) \token
Globally sets the test variable \texttt{g_peek_token} equal to \textit{(token)} (as an implicit token, \textit{not} as a token list), and then expands the \textit{(function)}. The \textit{(token)} remains in the input stream as the next item after the \textit{(function)}. The \textit{(token)} here may be \texttt{\textbackslash i}, \{ or \} (assuming normal \LaTeX{} category codes), \textit{i.e.} it is not necessarily the next argument which would be grabbed by a normal function.
\end{verbatim}

\texttt{l_peek_token} Token set by \texttt{peek_after:Nw} and available for testing as described above.

\texttt{g_peek_token} Token set by \texttt{peek_gafter:Nw} and available for testing as described above.
\peek_catcode:NTF \peek_catcode:NTF ⟨test token⟩ {⟨true code⟩} {⟨false code⟩}
Tests if the next ⟨token⟩ in the input stream has the same category code as the ⟨test token⟩ (as defined by the test \token_if_eq_catcode:NTF). Spaces are respected by the test and the ⟨token⟩ is left in the input stream after the ⟨true code⟩ or ⟨false code⟩ (as appropriate to the result of the test).

\peek_catcode_ignore_spaces:NTF \peek_catcode_ignore_spaces:NTF ⟨test token⟩ {⟨true code⟩} {⟨false code⟩}
Tests if the next non-space ⟨token⟩ in the input stream has the same category code as the ⟨test token⟩ (as defined by the test \token_if_eq_catcode:NTF). Explicit and implicit space tokens (with character code 32 and category code 10) are ignored and removed by the test and the ⟨token⟩ is left in the input stream after the ⟨true code⟩ or ⟨false code⟩ (as appropriate to the result of the test).

\peek_catcode_remove:NTF \peek_catcode_remove:NTF ⟨test token⟩ {⟨true code⟩} {⟨false code⟩}
Tests if the next ⟨token⟩ in the input stream has the same category code as the ⟨test token⟩ (as defined by the test \token_if_eq_catcode:NTF). Spaces are respected by the test and the ⟨token⟩ is removed from the input stream if the test is true. The function then places either the ⟨true code⟩ or ⟨false code⟩ in the input stream (as appropriate to the result of the test).

\peek_catcode_remove_ignore_spaces:NTF \peek_catcode_remove_ignore_spaces:NTF ⟨test token⟩ {⟨true code⟩} {⟨false code⟩}
Tests if the next non-space ⟨token⟩ in the input stream has the same category code as the ⟨test token⟩ (as defined by the test \token_if_eq_catcode:NTF). Explicit and implicit space tokens (with character code 32 and category code 10) are ignored and removed by the test and the ⟨token⟩ is removed from the input stream if the test is true. The function then places either the ⟨true code⟩ or ⟨false code⟩ in the input stream (as appropriate to the result of the test).

\peek_charcode:NTF \peek_charcode:NTF ⟨test token⟩ {⟨true code⟩} {⟨false code⟩}
Tests if the next ⟨token⟩ in the input stream has the same character code as the ⟨test token⟩ (as defined by the test \token_if_eq_charcode:NTF). Spaces are respected by the test and the ⟨token⟩ is left in the input stream after the ⟨true code⟩ or ⟨false code⟩ (as appropriate to the result of the test).

\peek_charcode_ignore_spaces:NTF \peek_charcode_ignore_spaces:NTF ⟨test token⟩ {⟨true code⟩} {⟨false code⟩}
Tests if the next non-space ⟨token⟩ in the input stream has the same character code as the ⟨test token⟩ (as defined by the test \token_if_eq_charcode:NTF). Explicit and implicit space tokens (with character code 32 and category code 10) are ignored and removed by the test and the ⟨token⟩ is left in the input stream after the ⟨true code⟩ or ⟨false code⟩ (as appropriate to the result of the test).
\peek_charcode_remove:NTF \peek_charcode_remove:NTF \test \token \true \false
Tests if the next \token in the input stream has the same character code as the \test \token (as defined by the test \token_if_eq_charcode:NNTF). Spaces are respected by the test and the \token is removed from the input stream if the test is true. The function then places either the \true or \false in the input stream (as appropriate to the result of the test).

\peek_charcode_remove_ignore_spaces:NTF \peek_charcode_remove_ignore_spaces:NTF \test \token \true \false
Tests if the next non-space \token in the input stream has the same character code as the \test \token (as defined by the test \token_if_eq_charcode:NNTF). Explicit and implicit space tokens (with character code 32 and category code 10) are ignored and removed by the test and the \token is removed from the input stream if the test is true. The function then places either the \true or \false in the input stream (as appropriate to the result of the test).

\peek_meaning:NTF \peek_meaning:NTF \test \token \true \false
Tests if the next \token in the input stream has the same meaning as the \test \token (as defined by the test \token_if_eq_meaning:NNTF). Spaces are respected by the test and the \token is left in the input stream after the \true or \false (as appropriate to the result of the test).

\peek_meaning_remove:NTF \peek_meaning_remove:NTF \test \token \true \false
Tests if the next \token in the input stream has the same meaning as the \test \token (as defined by the test \token_if_eq_meaning:NNTF). Spaces are respected by the test and the \token is removed from the input stream if the test is true. The function then places either the \true or \false in the input stream (as appropriate to the result of the test).

\peek_meaning_remove_ignore_spaces:NTF \peek_meaning_remove_ignore_spaces:NTF \test \token \true \false
Tests if the next non-space \token in the input stream has the same meaning as the \test \token (as defined by the test \token_if_eq_meaning:NNTF). Explicit and implicit space tokens (with character code 32 and category code 10) are ignored and removed by the test and the \token is removed from the input stream if the test is true. The function then places either the \true or \false in the input stream (as appropriate to the result of the test).
\peek_N_type:TF \{\text{true code}\} \{\text{false code}\}\\
Tests if the next \langle token\rangle in the input stream can be safely grabbed as an N-type argument. The test is \langle false\rangle if the next \langle token\rangle is either an explicit or implicit begin-group or end-group token (with any character code), or an explicit or implicit space character (with character code 32 and category code 10), or an outer token (never used in \LaTeX) and \langle true\rangle in all other cases. Note that a \langle true\rangle result ensures that the next \langle token\rangle is a valid N-type argument. However, if the next \langle token\rangle is for instance \c_space_token, the test takes the \langle false\rangle branch, even though the next \langle token\rangle is in fact a valid N-type argument. The \langle token\rangle is left in the input stream after the \langle true code\rangle or \langle false code\rangle (as appropriate to the result of the test).

\peek_analysis_map_inline:n \peek_analysis_map_inline:n \{\langle inline function\rangle\}\\
Repeatedly removes one \langle token\rangle from the input stream and applies the \langle inline function\rangle to it, until \peek_analysis_map_break: is called. The \langle inline function\rangle receives three arguments for each \langle token\rangle in the input stream:

- \langle tokens\rangle, which both o-expand and x-expand to the \langle token\rangle. The detailed form of \langle tokens\rangle may change in later releases.
- \langle char code\rangle, a decimal representation of the character code of the \langle token\rangle, \text{-}1 if it is a control sequence.
- \langle catcode\rangle, a capital hexadecimal digit which denotes the category code of the \langle token\rangle (0: control sequence, 1: begin-group, 2: end-group, 3: math shift, 4: alignment tab, 6: parameter, 7: superscript, 8: subscript, A: space, B: letter, C: other, D: active). This can be converted to an integer by writing \text{"\langle catcode\rangle"}.

These arguments are the same as for \l_analysis_map_inline:n defined in \l_analysis. The \langle char code\rangle and \langle catcode\rangle do not take the meaning of a control sequence or active character into account: for instance, upon encountering the token \c_group_begin_token in the input stream, \peek_analysis_map_inline:n calls the \langle inline function\rangle with \#1 being \exp_not:n { \c_group_begin_token } (with the current implementation), \#2 being \text{-}1, and \#3 being 0, as for any other control sequence. In contrast, upon encountering an explicit begin-group token \{, the \langle inline function\rangle is called with arguments \exp_after:wN { \if_false: } \fi: 123 and 1.

The mapping is done at the current group level, \textit{i.e.} any local assignments made by the \langle inline function\rangle remain in effect after the loop. Within the code, \l_peek_token is set equal (as a token, not a token list) to the token under consideration.

\peek_analysis_map_break: \peek_analysis_map_break:n \{ ... \peek_analysis_map_break:n \{\langle code\rangle\} \}\\
Stops the \peek_analysis_map_inline:n loop from seeking more tokens, and inserts \langle code\rangle in the input stream (empty for \peek_analysis_map_break:).
\textbf{\texttt{\textbackslash peek\_regex:NTF}} \texttt{\{\langle regex\rangle\} \{\langle true code\rangle\} \{\langle false code\rangle\}}

Tests if the \langle tokens \rangle that follow in the input stream match the \langle regular expression \rangle. Any \langle tokens \rangle that have been read are left in the input stream after the \langle true code \rangle or \langle false code \rangle (as appropriate to the result of the test). See \texttt{l3regex} for documentation of the syntax of regular expressions. The \langle regular expression \rangle is implicitly anchored at the start, so for instance \texttt{\peek\_regex:NTF \{ a \}} is essentially equivalent to \texttt{\peek\_charcode:NTF a}.

\textbf{\texttt{\textbackslash peek\_regex\_remove\_once:NTF}} \texttt{\{\langle regex\rangle\} \{\langle true code\rangle\} \{\langle false code\rangle\}}

Tests if the \langle tokens \rangle that follow in the input stream match the \langle regex \rangle. If the test is true, the \langle tokens \rangle are removed from the input stream and the \langle true code \rangle is inserted, while if the test is false, the \langle false code \rangle is inserted followed by the \langle tokens \rangle that were originally in the input stream. See \texttt{l3regex} for documentation of the syntax of regular expressions. The \langle regular expression \rangle is implicitly anchored at the start, so for instance \texttt{\peek\_regex\_remove\_once:NTF \{ a \}} is essentially equivalent to \texttt{\peek\_charcode\_remove:NTF a}.

\textbf{\texttt{\textbackslash tex\_hackers\ text}}: Implicit character tokens are correctly considered by \texttt{\peek\_regex:NTF} as control sequences, while functions that inspect individual tokens (for instance \texttt{\peek\_charcode:NTF}) only take into account their meaning.

The \texttt{\peek\_regex\_remove\_once:NTF} function only inspects as few tokens as necessary to determine whether the regular expression matches. For instance \texttt{\peek\_regex\_remove\_once:NTF \{ abc | [a-z] \} \{ \} \{ \} abc} will only inspect the first token \texttt{a} even though the first branch \texttt{abc} of the alternative is preferred in functions such as \texttt{\peek\_regex\_remove\_once:NTF}. This may have an effect on tokenization if the input stream has not yet been tokenized and category codes are changed.
If the \textit{tokens} that follow in the input stream match the \textit{regex}, replaces them according to the \textit{replacement} as for \texttt{\regex_replace_once:nn}, and leaves the result in the input stream, after the \textit{true code}. Otherwise, leaves \textit{false code} followed by the \textit{tokens} that were originally in the input stream, with no modifications. See \texttt{\l3regex} for documentation of the syntax of regular expressions and of the \textit{replacement}: for instance \texttt{\0} in the \textit{replacement} is replaced by the tokens that were matched in the input stream. The \textit{regular expression} is implicitly anchored at the start. In contrast to \texttt{\regex_replace_once:nn}, no error arises if the \textit{replacement} leads to an unbalanced token list: the tokens are inserted into the input stream without issue.

\textbf{\TeX hacklers note:} Implicit character tokens are correctly considered by \texttt{\peek_regex_replace_once:nn\TF} as control sequences, while functions that inspect individual tokens (for instance \texttt{\peek_charcode:NTF}) only take into account their meaning.

### 23.7 Description of all possible tokens

Let us end by reviewing every case that a given token can fall into. This section is quite technical and some details are only meant for completeness. We distinguish the meaning of the token, which controls the expansion of the token and its effect on \TeX’s state, and its shape, which is used when comparing token lists such as for delimited arguments. Two tokens of the same shape must have the same meaning, but the converse does not hold.

A token has one of the following shapes.

- A control sequence, characterized by the sequence of characters that constitute its name: for instance, \texttt{\use:n} is a five-letter control sequence.

- An active character token, characterized by its character code (between 0 and 1114111 for Lua\TeX and X\TeX and less for other engines) and category code 13.

- A character token, characterized by its character code and category code (one of 1, 2, 3, 4, 6, 7, 8, 10, 11 or 12 whose meaning is described below).

There are also a few internal tokens. The following list may be incomplete in some engines.

- Expanding \texttt{\the\font} results in a token that looks identical to the command that was used to select the current font (such as \texttt{\tenrm}) but it differs from it in shape.

- A “frozen” \texttt{\relax}, which differs from the primitive in shape (but has the same meaning), is inserted when the closing \texttt{\fi} of a conditional is encountered before the conditional is evaluated.

- Expanding \texttt{\noexpand\{token\}} (when the \texttt{\{token\}} is expandable) results in an internal token, displayed (temporarily) as \texttt{\notexpanded:\{token\}}, whose shape coincides with the \texttt{\{token\}} and whose meaning differs from \texttt{\relax}. 

195
• An `\outer` endtemplate: can be encountered when peeking ahead at the next token; this expands to another internal token, end of alignment template.

• Tricky programming might access a frozen `\endwrite`.

• Some frozen tokens can only be accessed in interactive sessions: `\cr`, `\right`, `\endgroup`, `\fi`, `\inaccessible`.

• In LuaTEX, there is also the strange case of “bytes” `\text{1100}` where `x, y` are any two lowercase hexadecimal digits, so that the hexadecimal number ranges from `\text{110000}=1114112` to `\text{1100ff}=1114367`. These are used to output individual bytes to files, rather than UTF-8. For the purposes of token comparisons they behave like non-expandable primitive control sequences (not characters) whose `\meaning` is the character followed by the given byte. If this byte is in the range `80–ff` this gives an “invalid utf-8 sequence” error: applying `\token_to_str:N` or `\token_to_meaning:N` to these tokens is unsafe. Unfortunately, they don’t seem to be detectable safely by any means except perhaps Lua code.

The meaning of a (non-active) character token is fixed by its category code (and character code) and cannot be changed. We call these tokens explicit character tokens. Category codes that a character token can have are listed below by giving a sample output of the \TeX primitive `\meaning`, together with their \LaTeX3 names and most common example:

1 begin-group character (`group_begin`, often `{`),

2 end-group character (`group_end`, often `}`),

3 math shift character (`math_toggle`, often `$`),

4 alignment tab character (`alignment`, often `&`),

6 macro parameter character (`parameter`, often `#`),

7 superscript character (`math_superscript`, often `^`),

8 subscript character (`math_subscript`, often `_`),

10 blank space (`space`, often character code 32),

11 the letter (`letter`, such as `A`),

12 the character (`other`, such as `0`).

Category code 13 (active) is discussed below. Input characters can also have several other category codes which do not lead to character tokens for later processing: 0 (escape), 5 (end_line), 9 (ignore), 14 (comment), and 15 (invalid).

The meaning of a control sequence or active character can be identical to that of any character token listed above (with any character code), and we call such tokens implicit character tokens. The meaning is otherwise in the following list:

• a macro, used in \LaTeX3 for most functions and some variables (`tl`, `fp`, `seq`, ...),

• a primitive such as `\def` or `\topmark`, used in \LaTeX3 for some functions,

• a register such as `\count123`, used in \LaTeX3 for the implementation of some variables (`int`, `dim`, ...),
• a constant integer such as \texttt{char"56} or \texttt{mathchar"121},

• a font selection command,

• undefined.

Macros can be \texttt{\protect} or not, \texttt{\long} or not (the opposite of what \LaTeX{} calls \texttt{nopar}), and \texttt{\outer} or not (unused in \LaTeX{}). Their \texttt{\meaning} takes the form

\begin{verbatim}
  \{prefix\} macro:\langle\text{argument}\rangle\rightarrow\langle\text{replacement}\rangle
\end{verbatim}

where \texttt{(prefix)} is among \texttt{\protect}\texttt{\long}\texttt{\outer}, \texttt{\langle\text{argument}\rangle}\ describes parameters that the macro expects, such as \texttt{#1#2#3}, and \texttt{\langle\text{replacement}\rangle}\ describes how the parameters are manipulated, such as \texttt{\int_eval:n\{#2+#1\#3\}}.

Now is perhaps a good time to mention some subtleties relating to tokens with category code 10 (space). Any input character with this category code (normally, space and tab characters) becomes a normal space, with character code 32 and category code 10.

When a macro takes an undelimited argument, explicit space characters (with character code 32 and category code 10) are ignored. If the following token is an explicit character token with category code 1 (begin-group) and an arbitrary character code, then \TeX{} scans ahead to obtain an equal number of explicit character tokens with category code 1 (begin-group) and 2 (end-group), and the resulting list of tokens (with outer braces removed) becomes the argument. Otherwise, a single token is taken as the argument for the macro: we call such single tokens "\texttt{N-type}", as they are suitable to be used as an argument for a function with the signature \texttt{:\textt{N}}.

When a macro takes a delimited argument \TeX{} scans ahead until finding the delimiter (outside any pairs of begin-group/end-group explicit characters), and the resulting list of tokens (with outer braces removed) becomes the argument. Note that explicit space characters at the start of the argument are \texttt{not} ignored in this case (and they prevent brace-stripping).
Chapter 24

The \texttt{l3prop} package

Property lists

\LMPXe implements a “property list” data type, which contain an unordered list of entries each of which consists of a \langle key\rangle and an associated \langle value\rangle. The \langle key\rangle and \langle value\rangle may both be any \langle balanced text\rangle, the \langle key\rangle is processed using \texttt{tl_to_str:n}, meaning that category codes are ignored. It is possible to map functions to property lists such that the function is applied to every key–value pair within the list.

Each entry in a property list must have a unique \langle key\rangle: if an entry is added to a property list which already contains the \langle key\rangle then the new entry overwrites the existing one. The \langle keys\rangle are compared on a string basis, using the same method as \texttt{str_if_eq:nn}.

Property lists are intended for storing key-based information for use within code. This is in contrast to key–value lists, which are a form of input parsed by the \texttt{l3keys} module.

24.1 Creating and initialising property lists

\begin{verbatim}
\prop_new:N \prop_new:c
\prop_clear:N \prop_clear:c \prop_gclear:N \prop_gclear:c
\end{verbatim}

\prop_new:N \langle property list\rangle

Creates a new \langle property list\rangle or raises an error if the name is already taken. The declaration is global. The \langle property list\rangle initially contains no entries.

\prop_clear:N \langle property list\rangle

Clears all entries from the \langle property list\rangle.

\prop_gclear:N \langle property list\rangle

Ensures that the \langle property list\rangle exists globally by applying \prop_new:N if necessary, then applies \prop_(g)clear:N to leave the list empty.
\prop_set_eq:NN \prop_set_eq: \langle property list \rangle \langle property list \rangle

Sets the content of \langle property list \rangle equal to that of \langle property list \rangle.

\prop_set_eq:NN \prop_set_eq: (cN|Nc|cc)
\prop_gset_eq:NN \prop_gset_eq: (cN|Nc|cc)

\prop_set_from_keyval:Nn \prop_set_from_keyval: \langle prop var \rangle \{ \langle key1 \rangle = \langle value1 \rangle, \langle key2 \rangle = \langle value2 \rangle, \ldots \}

Sets \langle prop var \rangle to contain key–value pairs given in the second argument. If duplicate keys appear only the last of the values is kept.

Spaces are trimmed around every \langle key \rangle and every \langle value \rangle, and if the result of trimming spaces consists of a single brace group then a set of outer braces is removed. This enables both the \langle key \rangle and the \langle value \rangle to contain spaces, commas or equal signs. The \langle key \rangle is then processed by \tl_to_str:n. This function correctly detects the = and , signs provided they have the standard category code 12 or they are active.

\prop_set_from_keyval:Nn \prop_set_from_keyval:cn
\prop_gset_from_keyval:Nn \prop_gset_from_keyval:cn

New: 2017-11-28
Updated: 2021-11-07

\prop_const_from_keyval:Nn \prop_const_from_keyval: \langle prop var \rangle \{ \langle key1 \rangle = \langle value1 \rangle, \langle key2 \rangle = \langle value2 \rangle, \ldots \}

Creates a new constant \langle prop var \rangle or raises an error if the name is already taken. The \langle prop var \rangle is set globally to contain key–value pairs given in the second argument, processed in the way described for \prop_set_from_keyval:Nn. If duplicate keys appear only the last of the values is kept. This function correctly detects the = and , signs provided they have the standard category code 12 or they are active.

\prop_const_from_keyval:Nn \prop_const_from_keyval:cn

New: 2017-11-28
Updated: 2021-11-07

24.2 Adding and updating property list entries

\prop_put:Nnn \prop_put: \langle property list \rangle \{ \langle key \rangle \} \{ \langle value \rangle \}
\prop_gput:Nnn \prop_gput: \langle property list \rangle \{ \langle key \rangle \} \{ \langle value \rangle \}

Adds an entry to the \langle property list \rangle which may be accessed using the \langle key \rangle and which has \langle value \rangle. If the \langle key \rangle is already present in the \langle property list \rangle, the existing entry is overwritten by the new \langle value \rangle. Both the \langle key \rangle and \langle value \rangle may contain any \langle balanced text \rangle. The \langle key \rangle is stored after processing with \tl_to_str:n, meaning that category codes are ignored.

\prop_put_if_new:Nnn \prop_put_if_new: \langle property list \rangle \{ \langle key \rangle \} \{ \langle value \rangle \}
\prop_gput_if_new:Nnn \prop_gput_if_new:cnn

If the \langle key \rangle is present in the \langle property list \rangle then no action is taken. Otherwise, a new entry is added as described for \prop_put:Nnn.
\prop_get:Nn \prop_get:NnN \prop_gget:NnN \prop_put_from_keyval:Nn \prop_gput_from_keyval:Nn \prop_pop:NnN \prop_gpop:NnN \prop_gconcat:NNN \prop_concat:NNN \prop_gput_from_keyval:cn \prop_gput_from_keyval:ccc

Updated: 2021-08-18

Updated: 2011-08-18

Updated: 2021-11-07

Updated: 2021-05-16

\textbf{24.3 Recovering values from property lists}

\prop_gget:NnN \prop_gget:Nn \(\text{\langle}\text{property list\text{\rangle}}\) \{\text{\langle}key\text{\rangle}\} \langle tl\text{\ var} \rangle

Recoversthe \(\langle value\rangle\) stored with \(\langle key\rangle\) from the \(\langle property\ list\rangle\), and places this in the \(\langle\text{token\ list\ variable}\rangle\). If the \(\langle key\rangle\) is not found in the \(\langle property\ list\rangle\) then the \(\langle\text{token\ list\ variable}\rangle\) is set to the special marker \texttt{\q_no_value}. The \(\langle\text{token\ list\ variable}\rangle\) is set within the current \TeX group. See also \prop_get:NnNTF.

\prop_gput:Nn \prop_gput:NnN \prop_gput_from_keyval:Nn \prop_gput_from_keyval:ncN \prop_gput_from_keyval:ccN

Updated: 2021-06-16

Updated: 2021-05-16

Updated: 2021-11-07

\prop_gpop:Nn \prop_gpop:NnN \prop_gpop:Nn \prop_gpop:NnN \prop_gpop:NnN

Updated: 2021-08-18

Updated: 2021-08-18

Recoversthe \(\langle value\rangle\) stored with \(\langle key\rangle\) from the \(\langle property\ list\rangle\), and places this in the \(\langle\text{token\ list\ variable}\rangle\). If the \(\langle key\rangle\) is not found in the \(\langle property\ list\rangle\) then the \(\langle\text{token\ list\ variable}\rangle\) is set to the special marker \texttt{\q_no_value}. The \(\langle key\rangle\) and \(\langle value\rangle\) are then deleted from the property list. Both assignments are local. See also \prop_gpop:NnNTF.
\prop_item:Nn \prop_item:cn

Expands to the \langle value \rangle corresponding to the \langle key \rangle in the \langle property list \rangle. If the \langle key \rangle is missing, this has an empty expansion.

\TeXhackers note: This function is slower than the non-expandable analogue \prop_get:Nn. The result is returned within the \unexpanded primitive (\exp_not:n), which means that the \langle value \rangle does not expand further when appearing in an x-type argument expansion.

\prop_count:N \prop_count:c

Leaves the number of key–value pairs in the \langle property list \rangle in the input stream as an \langle integer denotation \rangle.

24.4 Modifying property lists

\prop_remove:Nn \prop_remove:(NV|cn|cV)
\prop_gremove:Nn \prop_gremove:(NV|cn|cV)

Removes the entry listed under \langle key \rangle from the \langle property list \rangle. If the \langle key \rangle is not found in the \langle property list \rangle no change occurs, i.e there is no need to test for the existence of a key before deleting it.

24.5 Property list conditionals

\prop_if_exist_p:N \prop_if_exist_p:c \prop_if_exist_p:NTF \prop_if_exist_p:cTF
\prop_if_empty_p:N \prop_if_empty_p:c \prop_if_empty_p:NTF \prop_if_empty_p:cTF
\prop_if_in_p:Nn \prop_if_in:p:(NV|No|cn|cV|co) \prop_if_in:NN \prop_if_in:NTF \prop_if_in:NNTF

Tests if the \langle key \rangle is present in the \langle property list \rangle, making the comparison using the method described by \str_if_eq:nnTF.

\TeXhackers note: This function iterates through every key–value pair in the \langle property list \rangle and is therefore slower than using the non-expandable \prop_get:NnTF.

Updated: 2011-09-15
24.6 Recovering values from property lists with branching

The functions in this section combine tests for the presence of a key in a property list with recovery of the associated value. This makes them useful for cases where different cases follow dependent on the presence or absence of a key in a property list. They offer increased readability and performance over separate testing and recovery phases.

\[ \text{prop_get:NnNTF} \]

\[ \text{prop_get:(NVN|NvN|cnN|cvN|coN)TF} \]

Updated: 2012-05-19

If the \( \text{key} \) is not present in the \( \text{property list} \), leaves the \( \text{false code} \) in the input stream. The value of the \( \text{token list variable} \) is not defined in this case and should not be relied upon. If the \( \text{key} \) is present in the \( \text{property list} \), stores the corresponding \( \text{value} \) in the \( \text{token list variable} \) without removing it from the \( \text{property list} \), then leaves the \( \text{true code} \) in the input stream. The \( \text{token list variable} \) is assigned locally.

\[ \text{prop_pop:NnNTF} \]

\[ \text{prop_pop:(NVN|NvN|cnN|cvN|coN)TF} \]

Updated: 2012-05-19

If the \( \text{key} \) is not present in the \( \text{property list} \), leaves the \( \text{false code} \) in the input stream. The value of the \( \text{token list variable} \) is not defined in this case and should not be relied upon. If the \( \text{key} \) is present in the \( \text{property list} \), pops the corresponding \( \text{value} \) in the \( \text{token list variable} \), i.e. removes the item from the \( \text{property list} \). Both the \( \text{property list} \) and the \( \text{token list variable} \) are assigned locally.

\[ \text{prop_gpop:NnNTF} \]

\[ \text{prop_gpop:(NVN|NvN|cnN|cvN|coN)TF} \]

Updated: 2012-05-19

If the \( \text{key} \) is not present in the \( \text{property list} \), leaves the \( \text{false code} \) in the input stream. The value of the \( \text{token list variable} \) is not defined in this case and should not be relied upon. If the \( \text{key} \) is present in the \( \text{property list} \), pops the corresponding \( \text{value} \) in the \( \text{token list variable} \), i.e. removes the item from the \( \text{property list} \). The \( \text{property list} \) is modified globally, while the \( \text{token list variable} \) is assigned locally.

24.7 Mapping over property lists

All mappings are done at the current group level, i.e. any local assignments made by the \( \text{function} \) or \( \text{code} \) discussed below remain in effect after the loop.

\[ \text{prop_map_function:NN} \]

\[ \text{prop_map_function:CN} \]

Updated: 2013-01-08

Applies \( \text{function} \) to every \( \text{entry} \) stored in the \( \text{property list} \). The \( \text{function} \) receives two arguments for each iteration: the \( \text{key} \) and associated \( \text{value} \). The order in which \( \text{entries} \) are returned is not defined and should not be relied upon. To pass further arguments to the \( \text{function} \), see \( \text{prop_map_tokens:Nn} \).
\prop_map_inline:Nn \prop_map_inline:cn

Applies \emph{(inline function)} to every \emph{(entry)} stored within the \emph{(property list)}. The \emph{(inline function)} should consist of code which receives the \emph{(key)} as \#1 and the \emph{(value)} as \#2. The order in which \emph{(entries)} are returned is not defined and should not be relied upon.

\prop_map_tokens:Nn \prop_map_tokens:cn

Analogue of \prop_map_function:NN which maps several tokens instead of a single function. The \emph{(code)} receives each key–value pair in the \emph{(property list)} as two trailing brace groups. For instance,

\prop_map_tokens:Nn \l_my_prop \{ \str_if_eq:nnT { mykey } \}

expands to the value corresponding to \texttt{mykey}: for each pair in \l_my_prop the function \str_if_eq:nnT receives \texttt{mykey}, the \emph{(key)} and the \emph{(value)} as its three arguments. For that specific task, \prop_item:Nn is faster.

\prop_map_break: \prop_map_break:

Used to terminate a \prop_map_... function before all entries in the \emph{(property list)} have been processed. This normally takes place within a conditional statement, for example

\prop_map_inline:Nn \l_my_prop
{\str_if_eq:nnTF { #1 } { bingo } { \prop_map_break: }
{ % Do something useful }\}

Use outside of a \prop_map_... scenario leads to low level \TeX errors.

\textbf{\TeX hackers note:} When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.
\prop_map_break:n {\langle code\rangle}

Used to terminate a \prop_map... function before all entries in the \langle property list \rangle have been processed, inserting the \langle code \rangle after the mapping has ended. This normally takes place within a conditional statement, for example

\prop_map_inline:Nn \l_my_prop
{\str_if_eq:nnTF { #1 } { bingo }
{ \prop_map_break:n { <code> } }
{ % Do something useful
}
}

Use outside of a \prop_map... scenario leads to low level \TeX{} errors.

\TeX{}hackers note: When the mapping is broken, additional tokens may be inserted before the \langle code \rangle is inserted into the input stream. This depends on the design of the mapping function.

24.8 Viewing property lists

\prop_show:N \prop_show:c

Displays the entries in the \langle property list \rangle in the terminal.

\prop_log:N \prop_log:c

Writes the entries in the \langle property list \rangle in the log file.

24.9 Scratch property lists

\l_tmpa_prop \l_tmpb_prop

Scratch property lists for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX{}-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_prop \g_tmpb_prop

Scratch property lists for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX{}-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
24.10 Constants

\c_empty_prop A permanently-empty property list used for internal comparisons.
Chapter 25

The l3skip package
Dimensions and skips

\LaTeX3 provides two general length variables: \texttt{dim} and \texttt{skip}. Lengths stored as \texttt{dim} variables have a fixed length, whereas \texttt{skip} lengths have a rubber (stretch/shrink) component. In addition, the \texttt{muskip} type is available for use in math mode: this is a special form of \texttt{skip} where the lengths involved are determined by the current math font (in \texttt{mu}). There are common features in the creation and setting of length variables, but for clarity the functions are grouped by variable type.

25.1 Creating and initialising \texttt{dim} variables

\begin{itemize}
\item \texttt{\dim_new:N} \langle \texttt{dimension} \rangle
\item \texttt{\dim_new:c} \langle \texttt{dimension} \rangle
\item \texttt{\dim_const:Nn} \langle \texttt{dimension} \rangle \left\langle \texttt{dimension expression} \right\rangle
\item \texttt{\dim_const:cn}
\item \texttt{\dim_zero:N} \langle \texttt{dimension} \rangle
\item \texttt{\dim_zero:c}
\item \texttt{\dim_gzero:N}
\item \texttt{\dim_gzero:c}
\item \texttt{\dim_zero_new:N} \langle \texttt{dimension} \rangle
\item \texttt{\dim_zero_new:c}
\item \texttt{\dim_gzero_new:N}
\item \texttt{\dim_gzero_new:c}
\end{itemize}

- \texttt{\dim_new:N} \langle \texttt{dimension} \rangle
  
  Creates a new \langle \texttt{dimension} \rangle or raises an error if the name is already taken. The declaration is global. The \langle \texttt{dimension} \rangle is initially equal to 0 pt.

- \texttt{\dim_new:c} \langle \texttt{dimension} \rangle

  Creates a new constant \langle \texttt{dimension} \rangle or raises an error if the name is already taken. The value of the \langle \texttt{dimension} \rangle is set globally to the \langle \texttt{dimension expression} \rangle.

- \texttt{\dim_zero:N} \langle \texttt{dimension} \rangle

  Sets \langle \texttt{dimension} \rangle to 0 pt.

- \texttt{\dim_zero:c}

- \texttt{\dim_gzero:N}

- \texttt{\dim_gzero:c}

- \texttt{\dim_zero_new:N} \langle \texttt{dimension} \rangle

  Ensures that the \langle \texttt{dimension} \rangle exists globally by applying \texttt{\dim_new:N} if necessary, then applies \texttt{\dim (g)zero:N} to leave the \langle \texttt{dimension} \rangle set to zero.
Tests whether the \textit{dimension} is currently defined. This does not check that the \textit{dimension} really is a dimension variable.

### 25.2 Setting \texttt{dim} variables

\texttt{\dim_add:N} \texttt{(dimension)} \{(dimension expression)}

Adds the result of the \textit{dimension expression} to the current content of the \textit{dimension}.

\texttt{\dim_set:Nn} \texttt{(dimension)} \{(dimension expression)}

Sets \textit{dimension} to the value of \textit{dimension expression}, which must evaluate to a length with units.

\texttt{\dim_set_eq:NN} \texttt{(dimension\_1) (dimension\_2)}

Sets the content of \textit{dimension\_1} equal to that of \textit{dimension\_2}.

\texttt{\dim_sub:Nn} \texttt{(dimension)} \{(dimension expression)}

Subtracts the result of the \textit{dimension expression} from the current content of the \textit{dimension}.

### 25.3 Utilities for dimension calculations

\texttt{\dim_abs:n} \texttt{(dimexpr)}

Converts the \textit{dimexpr} to its absolute value, leaving the result in the input stream as a \textit{dimension denotation}.

\texttt{\dim_max:nn} \texttt{\dim_min:nn}

Evaluates the two \textit{dimension expressions} and leaves either the maximum or minimum value in the input stream as appropriate, as a \textit{dimension denotation}.
\dim_ratio:nn \dim_ratio:nn {(dimexpr_1)} {(dimexpr_2)}

Parses the two \textit{dimension expressions} and converts the ratio of the two to a form suitable for use inside a \textit{dimension expression}. This ratio is then left in the input stream, allowing syntax such as

\begin{verbatim}
\dim_set:Nn \l_my_dim
 { 10 pt * \dim_ratio:nn { 5 pt } { 10 pt } }
\end{verbatim}

The output of \texttt{\dim_ratio:nn} on full expansion is a ratio expression between two integers, with all distances converted to scaled points. Thus

\begin{verbatim}
\tl_set:Nx \l_my_tl { \dim_ratio:nn { 5 pt } { 10 pt } }
\tl_show:N \l_my_tl
\end{verbatim}

displays 327680/655360 on the terminal.

\section*{25.4 Dimension expression conditionals}

\begin{verbatim}
\dim_compare_p:nNn \dim_compare_p:nNn {(dimexpr_1)} {relation} {(dimexpr_2)}
\end{verbatim}

\begin{verbatim}
\dim_compare:nNnTF \dim_compare:nNnTF {(dimexpr_1)} {relation} {(dimexpr_2)}
{true code} {false code}
\end{verbatim}

This function first evaluates each of the \textit{dimension expressions} as described for \texttt{\dim_eval:n}. The two results are then compared using the \textit{relation}:

\begin{verbatim}
Equal  =
Greater than  >
Less than    <
\end{verbatim}

This function is less flexible than \texttt{\dim_compare:nTF} but around 5 times faster.
\def\dim_compare_p:n *
\def\dim_compare:nTF *
\verbatimverbatim
\begin{verbatim}
  \dim_compare:nTF
  { 
    \dimexpr 1 \relax \dimexpr \dimexpr N - 1 \relax \dimexpr N \relax \dimexpr N + 1 \relax 
    \endverbatim
  }
\end{verbatim}

This function evaluates the \emph{dimension expressions} as described for \texttt{\textbackslash dim_eval:n} and compares consecutive results using the corresponding \emph{relations}, namely it compares \texttt{\dimexpr 1 \relax} and \texttt{\dimexpr 2 \relax} using the \texttt{\relation 1 \relax}, then \texttt{\dimexpr 2 \relax} and \texttt{\dimexpr 3 \relax} using the \texttt{\relation 2 \relax}, until finally comparing \texttt{\dimexpr N \relax} and \texttt{\dimexpr N + 1 \relax} using the \texttt{\relation N \relax}. The test yields \texttt{true} if all comparisons are \texttt{true}. Each \emph{dimension expression} is evaluated only once, and the evaluation is lazy, in the sense that if one comparison is \texttt{false}, then no other \emph{dimension expression} is evaluated and no other comparison is performed. The \emph{relations} can be any of the following:

\begin{itemize}
  \item Equal \texttt{=} or \texttt{==}
  \item Greater than or equal to \texttt{>=}
  \item Greater than \texttt{>}
  \item Less than or equal to \texttt{<=}
  \item Less than \texttt{<}
  \item Not equal \texttt{!=}
\end{itemize}

This function is more flexible than \texttt{\dim_compare:nNnTF} but around 5 times slower.
This function evaluates the \textit{test dimension expression} and compares this in turn to each of the \textit{dimension expression cases}. If the two are equal then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{false code} is inserted. The function \texttt{\textbackslash dim\_case:nn}, which does nothing if there is no match, is also available. For example

\begin{verbatim}
\dim_set:Nn \l_tmpa_dim { 5 pt }
\dim_case:nnF { 2 \l_tmpa_dim }
{ 5 pt } { Small }
{ 4 pt + 6 pt } { Medium }
{ - 10 pt } { Negative }
{ No idea! }
\end{verbatim}

leaves “Medium” in the input stream.

\subsection*{25.5 Dimension expression loops}

\begin{verbatim}
\dim_do_until:Nnn \{ \text{dimexpr}_1 \} \{ \text{relation} \} \{ \text{dimexpr}_2 \} \{ \text{code} \}
\end{verbatim}

\texttt{\textbackslash dim\_do\_until:Nnn} \texttt{\textbackslash dim\_do\_until:nNnn} places the \textit{code} in the input stream for \texttt{\TeX} to process, and then evaluates the relationship between the two \textit{dimension expressions} as described for \texttt{\textbackslash dim\_compare:nNnTF}. If the test is \texttt{false} then the \textit{code} is inserted into the input stream again and a loop occurs until the \textit{relation} is \texttt{true}.

\begin{verbatim}
\dim_do_while:Nnn \{ \text{dimexpr}_1 \} \{ \text{relation} \} \{ \text{dimexpr}_2 \} \{ \text{code} \}
\end{verbatim}

\texttt{\textbackslash dim\_do\_while:Nnn} \texttt{\textbackslash dim\_do\_while:nNnn} places the \textit{code} in the input stream for \texttt{\TeX} to process, and then evaluates the relationship between the two \textit{dimension expressions} as described for \texttt{\textbackslash dim\_compare:nNnTF}. If the test is \texttt{true} then the \textit{code} is inserted into the input stream again and a loop occurs until the \textit{relation} is \texttt{false}.

\begin{verbatim}
\dim_until_do:Nnn \{ \text{dimexpr}_1 \} \{ \text{relation} \} \{ \text{dimexpr}_2 \} \{ \text{code} \}
\end{verbatim}

\texttt{\textbackslash dim\_until\_do:Nnn} \texttt{\textbackslash dim\_until\_do:nNnn} evaluates the relationship between the two \textit{dimension expressions} as described for \texttt{\textbackslash dim\_compare:nNnTF}, and then places the \textit{code} in the input stream if the \textit{relation} is \texttt{false}. After the \textit{code} has been processed by \texttt{\TeX} the test is repeated, and a loop occurs until the test is \texttt{true}.
\texttt{\dim_step:nnn} \texttt{(\textit{initial value}) \texttt{(step)} \texttt{(final value)} \texttt{(function)}}

This function first evaluates the \textit{(initial value)}, \textit{(step)} and \textit{(final value)}, all of which should be dimension expressions. The \textit{(function)} is then placed in front of each \textit{(value)} from the \textit{(initial value)} to the \textit{(final value)} in turn (using \textit{(step)} between each \textit{(value)}). The \textit{(step)} must be non-zero. If the \textit{(step)} is positive, the loop stops when the \textit{(value)} becomes larger than the \textit{(final value)}. If the \textit{(step)} is negative, the loop stops when the \textit{(value)} becomes smaller than the \textit{(final value)}. The \textit{(function)} should absorb one argument.

\texttt{\dim_step_inline:nnn} \texttt{(\textit{initial value}) \texttt{(step)} \texttt{(final value)} \texttt{(code)}}

This function first evaluates the \textit{(initial value)}, \textit{(step)} and \textit{(final value)}, all of which should be dimension expressions. Then for each \textit{(value)} from the \textit{(initial value)} to the \textit{(final value)} in turn (using \textit{(step)} between each \textit{(value)}), the \textit{(code)} is inserted into the input stream with \#1 replaced by the current \textit{(value)}. Thus the \textit{(code)} should define a function of one argument (\#1).

25.6 Dimension step functions
This function first evaluates the \langle initial value \rangle, \langle step \rangle and \langle final value \rangle, all of which should be dimension expressions. Then for each \langle value \rangle from the \langle initial value \rangle to the \langle final value \rangle in turn (using \langle step \rangle between each \langle value \rangle), the \langle code \rangle is inserted into the input stream, with the \langle tl var \rangle defined as the current \langle value \rangle. Thus the \langle code \rangle should make use of the \langle tl var \rangle.

25.7 Using dim expressions and variables

\dim_eval:n \{\langle dimension expression \rangle\}

Evaluates the \langle dimension expression \rangle, expanding any dimensions and token list variables within the \langle expression \rangle to their content (without requiring \dim_use:N/\tl_use:N) and applying the standard mathematical rules. The result of the calculation is left in the input stream as a \langle dimension denotation \rangle after two expansions. This is expressed in points (pt), and requires suitable termination if used in a \TeX-style assignment as it is not an \langle internal dimension \rangle.

\dim_sign:n \{\langle dimexpr \rangle\}

Evaluates the \langle dimexpr \rangle then leaves 1 or 0 or −1 in the input stream according to the sign of the result.

\dim_use:N \{\langle dimension \rangle\}

Recoveres the content of a \langle dimension \rangle and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Can be omitted in places where a \langle dimension \rangle is required (such as in the argument of \dim_eval:n).

\TeXhacksnote: \dim_use:N is the \TeX primitive \the: this is one of several \TeXthree names for this primitive.

\dim_to_decimal:n \{\langle dimexpr \rangle\}

Evaluates the \langle dimension expression \rangle, and leaves the result, expressed in points (pt) in the input stream, with no units. The result is rounded by \TeX to four or five decimal places. If the decimal part of the result is zero, it is omitted, together with the decimal marker.

For example

\dim_to_decimal:n \{1bp\}

leaves 1.00374 in the input stream, i.e. the magnitude of one “big point” when converted to \langle \TeX \rangle points.
\texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_bp:n}} * \texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_bp:n}} {\langle \texttt{dimexpr} \rangle} \New: 2014-07-15

Evaluates the \textit{(dimension expression)}, and leaves the result, expressed in big points (bp) in the input stream, with \textit{no units}. The result is rounded by \TeX to four or five decimal places. If the decimal part of the result is zero, it is omitted, together with the decimal marker.

For example

\texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_bp:n}} \{1\text{pt}\}

leaves $0.99628$ in the input stream, \textit{i.e.} the magnitude of one (\TeX) point when converted to big points.

\texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_sp:n}} * \texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_sp:n}} {\langle \texttt{dimexpr} \rangle} \New: 2015-05-18

Evaluates the \textit{(dimension expression)}, and leaves the result, expressed in scaled points (sp) in the input stream, with \textit{no units}. The result is necessarily an integer.

\texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_unit:nn}} * \texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_unit:nn}} {\langle \texttt{dimexpr}_1 \rangle} {\langle \texttt{dimexpr}_2 \rangle} \New: 2014-07-15

Evaluates the \textit{(dimension expressions)}, and leaves the value of $\langle \texttt{dimexpr}_1 \rangle$, expressed in a unit given by $\langle \texttt{dimexpr}_2 \rangle$, in the input stream. The result is a decimal number, rounded by \TeX to four or five decimal places. If the decimal part of the result is zero, it is omitted, together with the decimal marker.

For example

\texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_unit:nn}} \{1bp\} \{1mm\}

leaves $0.35277$ in the input stream, \textit{i.e.} the magnitude of one big point when converted to millimetres.

Note that this function is not optimised for any particular output and as such may give different results to \texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_bp:n}} or \texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_sp:n}}. In particular, the latter is able to take a wider range of input values as it is not limited by the ability to calculate a ratio using \texttt{\textbackslash \varepsilon\textbackslash \textbackslash \texttt{\textbackslash \texttt{TeX}}} primitives, which is required internally by \texttt{\textbackslash \texttt{dim\_to\_decimal\_in\_unit:nn}}.

\texttt{\textbackslash \texttt{dim\_to\_fp:n}} * \texttt{\textbackslash \texttt{dim\_to\_fp:n}} {\langle \texttt{dimexpr} \rangle} \New: 2012-05-08

Expands to an internal floating point number equal to the value of the \textit{(dimexpr)} in pt. Since dimension expressions are evaluated much faster than their floating point equivalent, \texttt{\textbackslash \texttt{dim\_to\_fp:n}} can be used to speed up parts of a computation where a low precision and a smaller range are acceptable.

\section{25.8 Viewing \texttt{dim} variables}

\texttt{\textbackslash \texttt{dim\_show:N}} \texttt{\textbackslash \texttt{dim\_show:c}} \texttt{\textbackslash \texttt{dim\_show:N}} \{\texttt{dimension}\}

Displays the value of the \textit{(dimension)} on the terminal.
25.9 Constant dimensions

\c_max_dim The maximum value that can be stored as a dimension. This can also be used as a component of a skip.

\c_zero_dim A zero length as a dimension. This can also be used as a component of a skip.

25.10 Scratch dimensions

\l_tmpa_dim \l_tmpb_dim Scratch dimension for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_dim \g_tmpb_dim Scratch dimension for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

25.11 Creating and initialising \texttt{skip} variables

\skip_new:N \skip_new:c Creates a new \texttt{skip} or raises an error if the name is already taken. The declaration is global. The \texttt{skip} is initially equal to 0pt.
\skip_const:Nn \skip_const:cn
\skip_zero:N \skip_zero:c \skip_gzero:N \skip_gzero:c
\skip_zero_new:N \skip_zero_new:c \skip_gzero_new:N \skip_gzero_new:c
\skip_if_exist_p:N \skip_if_exist_p:c \skip_if_exist:N \skip_if_exist:c
\skip_add:Nn \skip_add:cn \skip_gadd:Nn \skip_gadd:cn
\skip_set:Nn \skip_set:cn \skip_gset:Nn \skip_gset:cn
\skip_set_eq:NN \skip_set_eq:(cN|Nc|cc) \skip_gset_eq:NN \skip_gset_eq:(ccN|Nc|cc)
\skip_sub:Nn \skip_sub:cn \skip_gsub:Nn \skip_gsub:cn

\skip_const:Nn \skip \{\textit{skip expression}\}

Creates a new constant \textit{skip} or raises an error if the name is already taken. The value of the \textit{skip} is set globally to the \textit{skip expression}.

\skip_zero:N \skip
Sets \textit{skip} to 0 pt.

\skip_zero_new:N \skip
Ensures that the \textit{skip} exists globally by applying \skip_new:N if necessary, then applies \skip_(g)zero:N to leave the \textit{skip} set to zero.

\skip_if_exist_p:N \skip \textit{skip} \skip_if_exist:NTF \skip \{\textit{true code}\} \{\textit{false code}\}

Tests whether the \textit{skip} is currently defined. This does not check that the \textit{skip} really is a skip variable.

25.12 Setting skip variables

\skip_add:Nn \skip \{\textit{skip expression}\}

Adds the result of the \textit{skip expression} to the current content of the \textit{skip}.

\skip_set:Nn \skip \{\textit{skip expression}\}

Sets \textit{skip} to the value of \textit{skip expression}, which must evaluate to a length with units and may include a rubber component (for example 1 cm plus 0.5 cm).

\skip_set_eq:NN \skip_1 \skip \skip_2

Sets the content of \textit{skip}_1 equal to that of \textit{skip}_2.

\skip_sub:Nn \skip \{\textit{skip expression}\}

Subtracts the result of the \textit{skip expression} from the current content of the \textit{skip}.
25.13 Skip expression conditionals

\skip_if_eq:n\nn \{\skipexpr_1\}\{\skipexpr_2\}
\skip_if_eq:nnTF \{\skipexpr_1\}\{\skipexpr_2\}\{\true\}\{\false\}

This function first evaluates each of the \{\skip expressions\} as described for \skip_eval:n. The two results are then compared for exact equality, i.e. both the fixed and rubber components must be the same for the test to be true.

\skip_if_finite:n\nn \{\skipexpr\}
\skip_if_finite:nTF \{\skipexpr\}\{\true\}\{\false\}

Evaluates the \{\skip expression\} as described for \skip_eval:n, and then tests if all of its components are finite.

25.14 Using skip expressions and variables

\skip_eval:n \{\skip expression\}

Evaluates the \{\skip expression\}, expanding any skips and token list variables within the \{\expression\} to their content (without requiring \skip_use:N/\tl_use:N) and applying the standard mathematical rules. The result of the calculation is left in the input stream as a \{\glue\} after two expansions. This is expressed in points (pt), and requires suitable termination if used in a \TeX-style assignment as it is not an \{\internal\}.

\skip_use:N \{\skip\}
\skip_use:c \{\skip\}

Recovers the content of a \{\skip\} and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Can be omitted in places where a \{\dimension\} or \{\skip\} is required (such as in the argument of \skip_eval:n).

\TeXhackers note: \skip_use:N is the \TeX primitive \the: this is one of several \TeX3 names for this primitive.

25.15 Viewing skip variables

\skip_show:N \{\skip\}
\skip_show:n \{\skip expression\}

Displays the value of the \{\skip\} on the terminal.

Displays the result of evaluating the \{\skip expression\} on the terminal.
\skip_log:N \skip_log:c

New: 2014-08-22
Updated: 2015-08-03

\skip_log:n

New: 2014-08-22
Updated: 2015-08-07

Wrote the value of the \langle \skip \rangle in the log file.

Wrote the result of evaluating the \langle \skip \ expression \rangle in the log file.

25.16 Constant skips

The maximum value that can be stored as a skip (equal to \c_max_dim in length), with no stretch nor shrink component.

A zero length as a skip, with no stretch nor shrink component.

25.17 Scratch skips

\l_tmpa_skip \l_tmpb_skip

Scratch skip for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_skip \g_tmpb_skip

Scratch skip for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

25.18 Inserting skips into the output

\skip_horizontal:N \skip_horizontal:c \skip_horizontal:n

Updated: 2011-10-22

\skip_horizontal:N \langle \skip \rangle
\skip_horizontal:n \{\langle \skip \ expression \rangle\}

Inserts a horizontal \langle \skip \rangle into the current list. The argument can also be a \langle \dim \rangle.

\TeXhackers note: \skip_horizontal:N is the \TeX primitive \hskip renamed.
Inserts a vertical \( \langle \text{skip} \rangle \) into the current list. The argument can also be a \( \langle \text{dim} \rangle \).

\textbf{\texttt{\LaTeX} hackers note:} \texttt{\skip_vertical:N} is the \LaTeX\ primitive \texttt{\vskip} renamed.

## 25.19 Creating and initialising \texttt{muskip} variables

\texttt{\muskip_new:N} \texttt{\muskip_new:c}

Creates a new \( \langle \text{muskip} \rangle \) or raises an error if the name is already taken. The declaration is global. The \( \langle \text{muskip} \rangle \) is initially equal to 0 mu.

\texttt{\muskip_const:Nn} \texttt{\muskip_const:cn}

Creates a new constant \( \langle \text{muskip} \rangle \) or raises an error if the name is already taken. The value of the \( \langle \text{muskip} \rangle \) is set globally to the \( \langle \text{muskip expression} \rangle \).

\texttt{\skip_zero:N} \texttt{\skip_zero:c} \texttt{\skip_gzero:N} \texttt{\skip_gzero:c}

Sets \( \langle \text{muskip} \rangle \) to 0 mu.

\texttt{\muskip_zero_new:N} \texttt{\muskip_zero_new:c} \texttt{\muskip_gzero_new:N} \texttt{\muskip_gzero_new:c}

Ensures that the \( \langle \text{muskip} \rangle \) exists globally by applying \texttt{\muskip_new:N} if necessary, then applies \texttt{\muskip\_\(g\)zero:N} to leave the \( \langle \text{muskip} \rangle \) set to zero.

\texttt{\muskip_if_exist_p:N} \texttt{\muskip_if_exist:NTF} \texttt{\muskip_if_exist:TF}

Tests whether the \( \langle \text{muskip} \rangle \) is currently defined. This does not check that the \( \langle \text{muskip} \rangle \) really is a muskip variable.

## 25.20 Setting \texttt{muskip} variables

\texttt{\muskip_add:Nn} \texttt{\muskip_add:cn} \texttt{\muskip_gadd:Nn} \texttt{\muskip_gadd:cn}

Adds the result of the \( \langle \text{muskip expression} \rangle \) to the current content of the \( \langle \text{muskip} \rangle \).
\muskip_set:Nn \muskip_set:cn \muskip_gset:Nn \muskip_gset:cn Updated: 2011-10-22

\muskip_set_eq:NN \muskip_set_eq:cn \muskip_gset_eq:NN\muskip_gset_eq:cn
\muskip_sub:Nn \muskip_sub:cn \muskip_gsub:Nn \muskip_gsub:cn Updated: 2011-10-22

\muskip_eval:n \muskip_eval:cn
\muskip_use:N \muskip_use:c
\muskip_show:N \muskip_show:c Updated: 2015-08-03

\textbf{25.21 Using muskip expressions and variables}

\textbf{25.22 Viewing muskip variables}
\muskip_show:n \muskip_show:n \{\(\text{muskip expression}\)\}
Displays the result of evaluating the \(\text{muskip expression}\) on the terminal.
\muskip_log:N \muskip_log:N \n \muskip_log:c \muskip_log:n \{\(\text{muskip}\)\}
Writes the value of the \(\text{muskip}\) in the log file.
\muskip_log:n \muskip_log:n \{\(\text{muskip expression}\)\}
Writes the result of evaluating the \(\text{muskip expression}\) in the log file.

\section{25.23 Constant muskips}

\c_max_muskip
The maximum value that can be stored as a muskip, with no stretch nor shrink component.
\c_zero_muskip
A zero length as a muskip, with no stretch nor shrink component.

\section{25.24 Scratch muskips}

\l_tmpa_muskip \l_tmpb_muskip
Scratch muskip for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
\g_tmpa_muskip \g_tmpb_muskip
Scratch muskip for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\section{25.25 Primitive conditional}

\if_dim:w \* \if_dim:w \(\text{dimen}_1\) \(\text{relation}\) \(\text{dimen}_2\)
\{\(\text{true code}\)\}
\else:\n\{\(\text{false}\)\}
\fi:
Compare two dimensions. The \(\text{relation}\) is one of \(<\), \(=\) or \(>\) with category code 12.

\TeXhackers note: This is the \TeX primitive \texttt{ifdim}. 

220
Chapter 26

The \texttt{l3keys} package

Key–value interfaces

The key–value method is a popular system for creating large numbers of settings for controlling function or package behaviour. The system normally results in input of the form

\begin{verbatim}
\MyModuleSetup{
  key-one = value one,
  key-two = value two
}
\end{verbatim}

or

\begin{verbatim}
\MyModuleMacro[{
  key-one = value one,
  key-two = value two
}]{argument}
\end{verbatim}

for the user.

The high level functions here are intended as a method to create key–value controls. Keys are themselves created using a key–value interface, minimising the number of functions and arguments required. Each key is created by setting one or more \texttt{properties} of the key:

\begin{verbatim}
\keys_define:nn { mymodule }
\{
  key-one .code:n = code including parameter #1,
  key-two .tl_set:N = \l_mymodule_store_tl
\}
\end{verbatim}

These values can then be set as with other key–value approaches:

\begin{verbatim}
\keys_set:nn { mymodule }
\{
  key-one = value one,
  key-two = value two
\}
\end{verbatim}
At a document level, \texttt{\keys_set:nn} is used within a document function, for example

\begin{verbatim}
\DeclareDocumentCommand \MyModuleSetup { m } { \keys_set:nn { mymodule } { #1 } }
\DeclareDocumentCommand \MyModuleMacro { o m } { \group_begin: \keys_set:nn { mymodule } { #1 } \% Main code for \MyModuleMacro \group_end: }
\end{verbatim}

Key names may contain any tokens, as they are handled internally using \texttt{\tl_to_-str:n}. As discussed in section 26.2, it is suggested that the character \texttt{/} is reserved for sub-division of keys into logical groups. Functions and variables are not expanded when creating key names, and so

\begin{verbatim}
\tl_set:Nn \l_mymodule_tmp_tl { key }
\keys_define:nn { mymodule } { \l_mymodule_tmp_tl .code:n = code }
\end{verbatim}

creates a key called \texttt{\l_mymodule_tmp_tl}, and not one called \texttt{key}.

### 26.1 Creating keys

\begin{verbatim}
\keys_define:nn { \langle module \rangle } { \langle keyval list \rangle }
\end{verbatim}

Parses the \texttt{\langle keyval list \rangle} and defines the keys listed there for \texttt{\langle module \rangle}. The \texttt{\langle module \rangle} name is treated as a string. In practice the \texttt{\langle module \rangle} should be chosen to be unique to the module in question (unless deliberately adding keys to an existing module).

The \texttt{\langle keyval list \rangle} should consist of one or more key names along with an associated key property. The properties of a key determine how it acts. The individual properties are described in the following text; a typical use of \texttt{\keys_define:nn} might read

\begin{verbatim}
\keys_define:nn { mymodule } { keyname .code:n = Some-code-using-#1, keyname .value_required:n = true }
\end{verbatim}

where the properties of the key begin from the \texttt{\texttt{.}} after the key name.

The various properties available take either no arguments at all, or require one or more arguments. This is indicated in the name of the property using an argument specification. In the following discussion, each property is illustrated attached to an arbitrary \texttt{\langle key \rangle}, which when used may be supplied with a \texttt{\langle value \rangle}. All key definitions are local.

Key properties are applied in the reading order and so the ordering is significant. Key properties which define “actions”, such as \texttt{.code:n}, \texttt{.tl_set:N}, etc., override one another. Some other properties are mutually exclusive, notably \texttt{.value_required:n} and
.value_forbidden:n, and so they replace one another. However, properties covering non-exclusive behaviours may be given in any order. Thus for example the following definitions are equivalent.

\keys_define:nn { mymodule }
\{
    keyname .code:n = Some-code-using-\#1,
    keyname .value_required:n = true
\}
\keys_define:nn { mymodule }
\{
    keyname .value_required:n = true,
    keyname .code:n = Some-code-using-\#1
\}

Note that with the exception of the special .undefine: property, all key properties define the key within the current \TeX scope.

\bool_set:N \bool_set:c \bool_gset:N \bool_gset:c
Updated: 2013-07-08

\bool_set_inverse:N \bool_set_inverse:c \bool_gset_inverse:N \bool_gset_inverse:c
New: 2011-08-28
Updated: 2013-07-08

\choice:
\choice:

\choices:nn \choices:(\{V\}on|x) \choices:(\{\}on|x)
New: 2011-08-21
Updated: 2013-07-10

\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
New: 2011-09-11

\value:nn { mymodule }
\{
    keyname .value_required:n = true
\}
\value:nn { mymodule }
\{
    keyname .value_required:n = true
\}

\keys_define:nn { mymodule }
\{
    keyname .code:n = Some-code-using-\#1,
    keyname .value_required:n = true
\}
\keys_define:nn { mymodule }
\{
    keyname .value_required:n = true,
    keyname .code:n = Some-code-using-\#1
\}

Note that with the exception of the special .undefine: property, all key properties define the key within the current \TeX scope.

\bool_set:N \bool_set:c \bool_gset:N \bool_gset:c
Updated: 2013-07-08

\bool_set_inverse:N \bool_set_inverse:c \bool_gset_inverse:N \bool_gset_inverse:c
New: 2011-08-28
Updated: 2013-07-08

\choice:
\choice:

\choices:nn \choices:(\{V\}on|x) \choices:(\{\}on|x)
New: 2011-08-21
Updated: 2013-07-10

\clist_set:N \clist_set:c \clist_gset:N \clist_gset:c
New: 2011-09-11
\[\text{code:n} = \{\text{code}\}\]

Stores the \(\text{code}\) for execution when \(\text{key}\) is used. The \(\text{code}\) can include one parameter \((#1)\), which will be the \(\text{value}\) given for the \(\text{key}\).

\[\text{key} \cdot \text{cs_set:Np} = \{\text{control sequence}\} \{\text{arg. spec.}\}\]

Defines \(\text{key}\) to set \(\text{control sequence}\) to have \(\text{arg. spec.}\) and replacement text \(\text{value}\).

\[\text{key} \cdot \text{default:n} = \{\text{default}\}\]

Creates a \(\text{default}\) value for \(\text{key}\), which is used if no value is given. This will be used if only the key name is given, but not if a blank \(\text{value}\) is given:

\[
\begin{align*}
\keys_define:nn \{ \text{mymodule} \} \{ \\
\text{key} \cdot \text{code:n} & = \text{Hello-#1}, \\
\text{key} \cdot \text{default:n} & = \text{World} \\
\end{align*}
\]

\[
\begin{align*}
\keys_set:nn \{ \text{mymodule} \} \{ \\
\text{key} = \text{Fred}, & \% \text{Prints ‘Hello Fred’} \\
\text{key}, & \% \text{Prints ‘Hello World’} \\
\text{key} = , & \% \text{Prints ‘Hello ‘} \\
\end{align*}
\]

The default does not affect keys where values are required or forbidden. Thus a required value cannot be supplied by a default value, and giving a default value for a key which cannot take a value does not trigger an error.

\[\text{key} \cdot \text{dim_set:N} = \{\text{dimension}\}\]

Defines \(\text{key}\) to set \(\text{dimension}\) to \(\text{value}\) (which must a dimension expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

\[\text{key} \cdot \text{fp_set:N} = \{\text{floating point}\}\]

Defines \(\text{key}\) to set \(\text{floating point}\) to \(\text{value}\) (which must a floating point expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

\[\text{key} \cdot \text{groups:n} = \{\text{groups}\}\]

Defines \(\text{key}\) as belonging to the \(\text{groups}\) declared. Groups provide a “secondary axis” for selectively setting keys, and are described in Section 26.6.
\texttt{\textbackslash keys\_define:nn \{ foo \} \{ test .code:n = \textbackslash tl\_show:n \{#1\} \} \textbackslash keys\_define:nn \{} \{} \text{bar .inherit:n = foo} \}

setting

\texttt{\textbackslash keys\_set:nn \{ bar \} \{ test = a \} }

will be equivalent to

\texttt{\textbackslash keys\_set:nn \{ foo \} \{ test = a \} }

\texttt{\textbackslash keys\_define:nn \{ key \} \{ value \} }

Initialises the \texttt{key} with the \texttt{value}, equivalent to

\texttt{\textbackslash keys\_set:nn \{ module \} \{ \langle key \rangle = \langle value \rangle \} }

\texttt{\textbackslash keys\_set:nn \{ key \} \{ \langle integer \rangle \} }

Defines \texttt{key} to set \texttt{\langle integer \rangle} to \texttt{\langle value \rangle} (which must be an integer expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

\texttt{\textbackslash keys\_define:nn \{ key \} \{ \langle keyval list \rangle \} }

Makes \texttt{key} a meta-key, which will set \texttt{\langle keyval list \rangle} in one go. The \texttt{\langle keyval list \rangle} can refer as \texttt{\#1} to the value given at the time the \texttt{key} is used (or, if no value is given, the \texttt{key}’s default value).

\texttt{\textbackslash keys\_set:nn \{ key \} \{ \langle path \rangle \} \{ \langle keyval list \rangle \} }

Makes \texttt{key} a meta-key, which will set \texttt{\langle keyval list \rangle} in one go using the \texttt{\langle path \rangle} in place of the current one. The \texttt{\langle keyval list \rangle} can refer as \texttt{\#1} to the value given at the time the \texttt{key} is used (or, if no value is given, the \texttt{key}’s default value).

\texttt{\textbackslash keys\_define:nn \{ key \} \{ \langle choices \rangle \} \{ \langle code \rangle \} }

Sets \texttt{key} to act as a multiple choice key, and defines a series \texttt{\langle choices \rangle} which are implemented using the \texttt{\langle code \rangle}. Inside \texttt{\langle code \rangle}, \texttt{\text{l\_keys\_choice_tl} will be the name of the choice made, and \texttt{l\_keys\_choice_int} will be the position of the choice in the list of \texttt{\langle choices \rangle} (indexed from 1). Choices are discussed in detail in section 26.3.
(key) .muskip_set:N = (muskip)
Defines (key) to set (muskip) to (value) (which must be a muskip expression). If the
variable does not exist, it is created globally at the point that the key is set up. The key
will require a value at point-of-use unless a default is set.

(key) .prop_put:N = (property list)
Defines (key) to put the (value) onto the (property list) stored under the (key). If the
variable does not exist, it is created globally at the point that the key is set up.

(key) .skip_set:N = (skip)
Defines (key) to set (skip) to (value) (which must be a skip expression). If the variable
does not exist, it is created globally at the point that the key is set up. The key will
require a value at point-of-use unless a default is set.

(key) .str_set:N = (string variable)
Defines (key) to set (string variable) to (value). If the variable does not exist, it is created
globally at the point that the key is set up.

(key) .str_set_x:N = (string variable)
Defines (key) to set (string variable) to (value), which will be subjected to an x-type
expansion (i.e. using \str_set:Nx). If the variable does not exist, it is created globally
at the point that the key is set up.

(key) .tl_set:N = (token list variable)
Defines (key) to set (token list variable) to (value). If the variable does not exist, it is
created globally at the point that the key is set up.

(key) .tl_set_x:N = (token list variable)
Defines (key) to set (token list variable) to (value), which will be subjected to an x-type
expansion (i.e. using \tl_set:Nx). If the variable does not exist, it is created globally
at the point that the key is set up.

(key) .undefine:
Removes the definition of the (key) within the current scope.

(key) .value_forbidden:n = true|false
Specifies that (key) cannot receive a (value) when used. If a (value) is given then an error
will be issued. Setting the property “false” cancels the restriction.
26.2 Sub-dividing keys

When creating large numbers of keys, it may be desirable to divide them into several sub-groups for a given module. This can be achieved either by adding a sub-division to the module name:

\keys_define:nn { mymodule / subgroup }
{ key .code:n = code }

or to the key name:

\keys_define:nn { mymodule }
{ subgroup / key .code:n = code }

As illustrated, the best choice of token for sub-dividing keys in this way is `/`. This is because of the method that is used to represent keys internally. Both of the above code fragments set the same key, which has full name `mymodule/subgroup/key`.

As illustrated in the next section, this subdivision is particularly relevant to making multiple choices.

26.3 Choice and multiple choice keys

The \texttt{l3keys} system supports two types of choice key, in which a series of pre-defined input values are linked to varying implementations. Choice keys are usually created so that the various values are mutually-exclusive: only one can apply at any one time. “Multiple” choice keys are also supported: these allow a selection of values to be chosen at the same time.

Mutually-exclusive choices are created by setting the \texttt{.choice:} property:

\keys_define:nn { mymodule }
{ key .choice: }

For keys which are set up as choices, the valid choices are generated by creating sub-keys of the choice key. This can be carried out in two ways.

In many cases, choices execute similar code which is dependant only on the name of the choice or the position of the choice in the list of all possibilities. Here, the keys can share the same code, and can be rapidly created using the \texttt{.choices:nn} property.

\keys_define:nn { mymodule }
{ key .choices:nn =
{ choice-a, choice-b, choice-c }
{ You-gave-choice-`\tl_use:N \l_keys_choice_tl',-
which-is-in-position-`\int_use:N \l_keys_choice_int `\c_space_tl
in-the-list. }
}
The index \texttt{\_keys_choice_int} in the list of choices starts at 1.

Inside the code block for a choice generated using \texttt{.choices:nn}, the variables \texttt{\_keys_choice_tl} and \texttt{\_keys_choice_int} are available to indicate the name of the current choice, and its position in the comma list. The position is indexed from 1. Note that, as with standard key code generated using \texttt{.code:n}, the value passed to the key (i.e. the choice name) is also available as \texttt{#1}.

On the other hand, it is sometimes useful to create choices which use entirely different code from one another. This can be achieved by setting the \texttt{.choice:} property of a key, then manually defining sub-keys.

\begin{verbatim}
\keys_define:nn { mymodule }
{
  key .choice:,
  key / choice-a .code:n = code-a,
  key / choice-b .code:n = code-b,
  key / choice-c .code:n = code-c,
}
\end{verbatim}

It is possible to mix the two methods, but manually-created choices should \textit{not} use \texttt{\_keys_choice_tl} or \texttt{\_keys_choice_int}. These variables do not have defined behaviour when used outside of code created using \texttt{.choices:nn} (i.e. anything might happen).

It is possible to allow choice keys to take values which have not previously been defined by adding code for the special \texttt{unknown} choice. The general behavior of the \texttt{unknown} key is described in Section 26.5. A typical example in the case of a choice would be to issue a custom error message:

\begin{verbatim}
\keys_define:nn { mymodule }
{
  key .choice:,
  key / choice-a .code:n = code-a,
  key / choice-b .code:n = code-b,
  key / choice-c .code:n = code-c,
  key / unknown .code:n = \msg_error:nnxxx { mymodule } { unknown-choice }
    { key } % Name of choice key
    { choice-a , choice-b , choice-c } % Valid choices
    { \exp_not:n {#1} } % Invalid choice given
  
}
\end{verbatim}

Multiple choices are created in a very similar manner to mutually-exclusive choices, using the properties \texttt{.multichoice:} and \texttt{.multichoice:nn}. As with mutually exclusive choices, multiple choices are define as sub-keys. Thus both

\begin{verbatim}
\keys_define:nn { mymodule }
{
  key .multichoice:nn =
    { choice-a , choice-b , choice-c } 
}
\end{verbatim}
You gave choice- `\tl_use:N \l_keys_choice_tl',-
which is in position-
`\int_use:N \l_keys_choice_int \c_space_tl
in the list.
}

and

\keys_define:nn { mymodule }
{
  key .multichoice:,
  key / choice-a .code:n = code-a,
  key / choice-b .code:n = code-b,
  key / choice-c .code:n = code-c,
}

are valid.

When a multiple choice key is set

\keys_set:nn { mymodule }
{
  key = { a , b , c } % 'key' defined as a multiple choice
}

each choice is applied in turn, equivalent to a clist mapping or to applying each value
individually:

\keys_set:nn { mymodule }
{
  key = a ,
  key = b ,
  key = c ,
}

Thus each separate choice will have passed to it the \l_keys_choice_tl and \l_keys_choice_int in exactly the same way as described for .choices:nn.

### 26.4 Setting keys

\keys_set:nn ⟨module⟩ ⟨⟨keyval list⟩⟩

 Parses the ⟨keyval list⟩, and sets those keys which are defined for ⟨module⟩. The behaviour on finding an unknown key can be set by defining a special unknown key: this is illustrated later.
For each key processed, information of the full path of the key, the name of the key and the value of the key is available within three token list variables. These may be used within the code of the key.

The value is everything after the =, which may be empty if no value was given. This is stored in \l\_keys\_value\_tl, and is not processed in any way by \keys\_set:nn.

The path of the key is a “full” description of the key, and is unique for each key. It consists of the module and full key name, thus for example

\keys\_set:nn { mymodule } { key-a = some-value }

has path mymodule/key-a while

\keys\_set:nn { mymodule } { subset / key-a = some-value }

has path mymodule/subset/key-a. This information is stored in \l\_keys\_path\_str.

The name of the key is the part of the path after the last /, and thus is not unique. In the preceding examples, both keys have name key-a despite having different paths. This information is stored in \l\_keys\_key\_str.

### 26.5 Handling of unknown keys

If a key has not previously been defined (is unknown), \keys\_set:nn looks for a special unknown key for the same module, and if this is not defined raises an error indicating that the key name was unknown. This mechanism can be used for example to issue custom error texts.

\keys\_define:nn { mymodule } {
  unknown .code:n =
  You\_tried\_to\_set\_key-'\l\_keys\_key\_str'-to-,'#1'.
}

These functions set keys which are known for the ⟨module⟩, and simply ignore other keys. The \keys\_set\_known:nn function parses the ⟨keyval list⟩, and sets those keys which are defined for ⟨module⟩. Any keys which are unknown are not processed further by the parser. In addition, \keys\_set\_known:nnN stores the key–value pairs in the ⟨tl⟩ in comma-separated form (i.e. an edited version of the ⟨keyval list⟩). When a ⟨root⟩ is given (\keys\_set\_known:nnN), the key–value entries are returned relative to this point in the key tree. When it is absent, only the key name and value are provided. The correct list is returned by nested calls.
26.6 Selective key setting

In some cases it may be useful to be able to select only some keys for setting, even though these keys have the same path. For example, with a set of keys defined using

\keys define:nn { mymodule }
{
  key-one .code:n = { \my_func:n {#1} } ,
  key-two .tl_set:N = \l_my_a_tl ,
  key-three .tl_set:N = \l_my_b_tl ,
  key-four .fp_set:N = \l_my_a_fp ,
}

the use of \keys_set:nn attempts to set all four keys. However, in some contexts it may only be sensible to set some keys, or to control the order of setting. To do this, keys may be assigned to \textit{groups}: arbitrary sets which are independent of the key tree. Thus modifying the example to read

\keys define:nn { mymodule }
{
  key-one .code:n = { \my_func:n {#1} } ,
  key-one .groups:n = { first } ,
  key-two .tl_set:N = \l_my_a_tl ,
  key-two .groups:n = { first } ,
  key-three .tl_set:N = \l_my_b_tl ,
  key-three .groups:n = { second } ,
  key-four .fp_set:N = \l_my_a_fp ,
}

assigns key-one and key-two to group first, key-three to group second, while key-four is not assigned to a group.

Selective key setting may be achieved either by selecting one or more groups to be made “active”, or by marking one or more groups to be ignored in key setting.

\keys_set_filter:nnn
\keys_set_filter:nnv
\keys_set_filter:nno
\keys_set_filter:nnN
\keys_set_filter:nnvN
\keys_set_filter:nnoN
\keys_set_filter:nnVnN
\keys_set_filter:nnvnN
\keys_set_filter:nnonN

Activates key filtering in an “opt-out” sense: keys assigned to any of the \textit{groups} specified are ignored. The \textit{groups} are given as a comma-separated list. Unknown keys are not assigned to any group and are thus always set. The key–value pairs for each key which is filtered out are stored in the \textit{tl} in a comma-separated form (\textit{i.e.} an edited version of the \textit{keyval list}). The \keys_set_filter:nn version skips this stage.

Use of \keys_set_filter:nnN can be nested, with the correct residual \textit{keyval list} returned at each stage. In the version which takes a \textit{root} argument, the key list is returned relative to that point in the key tree. In the cases without a \textit{root} argument, only the key names and values are returned.
Activates key filtering in an “opt-in” sense: only keys assigned to one or more of the ⟨groups⟩ specified are set. The ⟨groups⟩ are given as a comma-separated list. Unknown keys are not assigned to any group and are thus never set.

26.7 Utility functions for keys

Tests if the ⟨key⟩ exists for ⟨module⟩, i.e. if any code has been defined for ⟨key⟩.

Tests if the ⟨choice⟩ is defined for the ⟨key⟩ within the ⟨module⟩, i.e. if any code has been defined for ⟨key⟩/⟨choice⟩. The test is false if the ⟨key⟩ itself is not defined.

Displays in the terminal the information associated to the ⟨key⟩ for a ⟨module⟩, including the function which is used to actually implement it.

Writes in the log file the information associated to the ⟨key⟩ for a ⟨module⟩. See also \keys_show:nn which displays the result in the terminal.

26.8 Low-level interface for parsing key–val lists

To re-cap from earlier, a key–value list is input of the form

\KeyOne = \ValueOne ,
\KeyTwo = \ValueTwo ,
\KeyThree

where each key–value pair is separated by a comma from the rest of the list, and each key–value pair does not necessarily contain an equals sign or a value! Processing this type of input correctly requires a number of careful steps, to correctly account for braces, spaces and the category codes of separators.

While the functions described earlier are used as a high-level interface for processing such input, in special circumstances you may wish to use a lower-level approach. The low-level parsing system converts a ⟨key–value list⟩ into ⟨keys⟩ and associated ⟨values⟩.
After the parsing phase is completed, the resulting keys and values (or keys alone) are available for further processing. This processing is not carried out by the low-level parser itself, and so the parser requires the names of two functions along with the key–value list. One function is needed to process key–value pairs (it receives two arguments), and a second function is required for keys given without any value (it is called with a single argument).

The parser does not double # tokens or expand any input. Active tokens = and , appearing at the outer level of braces are converted to category “other” (12) so that the parser does not “miss” any due to category code changes. Spaces are removed from the ends of the keys and values. Keys and values which are given in braces have exactly one set removed (after space trimming), thus

key = {value here},

and

key = value here,

are treated identically.

\keyval_parse:nnn \keyval_parse:nnn \{\langle code_1\rangle\} \{\langle code_2\rangle\} \{\langle key–value\ list\rangle\}

Parses the \langle key–value\ list\rangle into a series of \langle keys\rangle and associated \langle values\rangle, or keys alone (if no \langle value\rangle was given). \langle code_1\rangle receives each \langle key\rangle (with no \langle value\rangle) as a trailing brace group, whereas \langle code_2\rangle is appended by two brace groups, the \langle key\rangle and \langle value\rangle. The order of the \langle keys\rangle in the \langle key–value\ list\rangle is preserved. Thus

\keyval_parse:nnn
\{ \use_none:nn { code 1 } \}
\{ \use_none:nn { code 2 } \}
\{ key1 = value1 , key2 = value2 , key3 = , key4 \}

is converted into an input stream

\use_none:nn { code 2 } { key1 } { value1 }
\use_none:nn { code 2 } { key2 } { value2 }
\use_none:nn { code 2 } { key3 } { }
\use_none:nn { code 1 } { key4 }

Note that there is a difference between an empty value (an equals sign followed by nothing) and a missing value (no equals sign at all). Spaces are trimmed from the ends of the \langle key\rangle and \langle value\rangle, then one outer set of braces is removed from the \langle key\rangle and \langle value\rangle as part of the processing. If you need exactly the output shown above, you’ll need to either x-type or e-type expand the function.

\TeX hackers note: The result of each list element is returned within \exp_not:n, which means that the converted input stream does not expand further when appearing in an x-type or e-type argument expansion.
\keyval_parse:NNn \keyval_parse:NNn \{function\_1\} \{function\_2\} \{(key-value list)\}

Parses the (key-value list) into a series of (keys) and associated (values), or keys alone (if no (value) was given). (function\_1) should take one argument, while (function\_2) should absorb two arguments. After \keyval_parse:NNn has parsed the (key-value list), (function\_1) is used to process keys given with no value and (function\_2) is used to process keys given with a value. The order of the (keys) in the (key-value list) is preserved. Thus

\keyval_parse:NNn \function:n \function:nn
{ key\_1 = value\_1, key\_2 = value\_2, key\_3 = , key\_4 } 

is converted into an input stream

\function:nn \{ key\_1 \} \{ value\_1 \}
\function:nn \{ key\_2 \} \{ value\_2 \}
\function:nn \{ key\_3 \} \{ \}
\function:n \{ key\_4 \}

Note that there is a difference between an empty value (an equals sign followed by nothing) and a missing value (no equals sign at all). Spaces are trimmed from the ends of the (key) and (value), then one outer set of braces is removed from the (key) and (value) as part of the processing.

This shares the implementation of \keyval_parse:nnn, the difference is only semantically.

\TeXhackers note: The result is returned within \exp_not:n, which means that the converted input stream does not expand further when appearing in an x-type or e-type argument expansion.
Chapter 27

The \texttt{l3intarray} package: fast global integer arrays

27.1 \texttt{l3intarray} documentation

For applications requiring heavy use of integers, this module provides arrays which can be accessed in constant time (contrast \texttt{l3seq}, where access time is linear). These arrays have several important features

- The size of the array is fixed and must be given at point of initialisation
- The absolute value of each entry has maximum $2^{30} - 1$ (i.e. one power lower than the usual $\texttt{c\_max\_int}$ ceiling of $2^{31} - 1$)

The use of \texttt{intarray} data is therefore recommended for cases where the need for fast access is of paramount importance.

\begin{verbatim}
\intarray_new:Nn \langle intarray var \rangle \{ \langle size \rangle \}
\end{verbatim}

Evaluates the integer expression $\langle size \rangle$ and allocates an \texttt{⟨integer array variable⟩} with that number of (zero) entries. The variable name should start with $\texttt{g}$ because assignments are always global.

\begin{verbatim}
\intarray_count:N \langle intarray var \rangle
\end{verbatim}

Expands to the number of entries in the \texttt{ ⟨integer array variable⟩}. Contrarily to $\texttt{\seq\_count:N}$ this is performed in constant time.

\begin{verbatim}
\intarray_gset:Nnn \langle intarray var \rangle \{ \langle position \rangle \} \{ \langle value \rangle \}
\end{verbatim}

Stores the result of evaluating the integer expression $\langle value \rangle$ into the \texttt{ ⟨integer array variable⟩} at the (integer expression) \texttt{ ⟨position⟩}. If the \texttt{ ⟨position⟩} is not between 1 and the $\texttt{\intarray_count:N}$, or the \texttt{ ⟨value⟩}’s absolute value is bigger than $2^{30} - 1$, an error occurs. Assignments are always global.
\intarray_const_from_clist:N \intarray_const_from_clist:cn
\intarray_const_from_clist:Nn \intarray_const_from_clist:cn

New: 2018-05-04

Creates a new constant \textit{(integer array variable)} or raises an error if the name is already taken. The \textit{(integer array variable)} is set (globally) to contain as its items the results of evaluating each \textit{(integer expression)} in the \textit{(comma list)}.

\intarray_gzero:N \intarray_gzero:c
\intarray_gzero:N \intarray_gzero:c

New: 2018-05-04

Sets all entries of the \textit{(integer array variable)} to zero. Assignments are always global.

\intarray_item:Nn \intarray_item:cn
\intarray_item:Nn \intarray_item:cn
\intarray_item:Nn \intarray_item:cn
\intarray_item:Nn \intarray_item:cn
\intarray_item:Nn \intarray_item:cn

New: 2018-03-29

Expands to the integer entry stored at the (integer expression) \textit{(position)} in the \textit{(integer array variable)}. If the \textit{(position)} is not between 1 and the \intarray_count:N, an error occurs.

\intarray_rand_item:N \intarray_rand_item:c
\intarray_rand_item:N \intarray_rand_item:c

New: 2018-05-05

Selects a pseudo-random item of the \textit{(integer array)}. If the \textit{(integer array)} is empty, produce an error.

\intarray_show:N \intarray_show:c \intarray_log:N \intarray_log:c
\intarray_show:N \intarray_show:c \intarray_log:N \intarray_log:c

New: 2018-05-04

Displays the items in the \textit{(integer array variable)} in the terminal or writes them in the log file.

27.1.1 Implementation notes

It is a wrapper around the \texttt{fontdimen} primitive, used to store arrays of integers (with a restricted range: absolute value at most $2^{30} - 1$). In contrast to \texttt{l3seq} sequences the access to individual entries is done in constant time rather than linear time, but only integers can be stored. More precisely, the primitive \texttt{fontdimen} stores dimensions but the \texttt{l3intarray} package transparently converts these from/to integers. Assignments are always global.

While Lua\TeX{}'s memory is extensible, other engines can “only” deal with a bit less than $4 \times 10^6$ entries in all \texttt{fontdimen} arrays combined (with default \TeX{} Live settings).
Chapter 28

The l3fp package: Floating points

A decimal floating point number is one which is stored as a significand and a separate exponent. The module implements expandably a wide set of arithmetic, trigonometric, and other operations on decimal floating point numbers, to be used within floating point expressions. Floating point expressions support the following operations with their usual precedence.

- Basic arithmetic: addition $x + y$, subtraction $x - y$, multiplication $x \times y$, division $x/y$, square root $\sqrt{x}$, and parentheses.
- Comparison operators: $x < y$, $x \leq y$, $x > y$, $x ! = y$ etc.
- Boolean logic: sign $sign \ x$, negation $! \ x$, conjunction $x && \ y$, disjunction $x || \ y$, ternary operator $x ? \ y : z$.
- Exponentials: $exp \ x$, $ln \ x$, $x^y$, $logb \ x$.
- Integer factorial: $fact \ x$.
- Trigonometry: $\sin \ x$, $\cos \ x$, $\tan \ x$, $\cot \ x$, $sec \ x$, $csc \ x$ expecting their arguments in radians, and $sind \ x$, $cosd \ x$, $tand \ x$, $cotd \ x$, $secd \ x$, $csced \ x$ expecting their arguments in degrees.
- Inverse trigonometric functions: $\arcsin \ x$, $\arccos \ x$, $\arctan \ x$, $\arccot \ x$, $\arcsec \ x$, $\arccsc \ x$ giving a result in radians, and $asind \ x$, $acosd \ x$, $atand \ x$, $acotd \ x$, $asecd \ x$, $acscd \ x$ giving a result in degrees.

(not yet) Hyperbolic functions and their inverse functions: $\sinh \ x$, $\cosh \ x$, $\tanh \ x$, $\coth \ x$, $\sech \ x$, $\csch \ x$, and $\arcsinh \ x$, $\arccosh \ x$, $\arctanh \ x$, $\arccoth \ x$, $\arcsech \ x$, $\arccsch \ x$.

- Extrema: $\max(x_1, x_2, \ldots)$, $\min(x_1, x_2, \ldots)$, $\abs(x)$.
- Rounding functions, controlled by two optional values, $n$ (number of places, 0 by default) and $t$ (behavior on a tie, NaN by default):
  - $\text{trunc}(x, n)$ rounds towards zero,
  - $\text{floor}(x, n)$ rounds towards $-\infty$,
- \( \text{ceil}(x,n) \) rounds towards \( +\infty \),
- \( \text{round}(x,n,t) \) rounds to the closest value, with ties rounded to an even value by default, towards zero if \( t = 0 \), towards \( +\infty \) if \( t > 0 \) and towards \( -\infty \) if \( t < 0 \).

And (not yet) modulo, and “quantize”.

- Random numbers: \( \text{rand}() \), \( \text{randint}(m,n) \).
- Constants: \( \pi \), \( \text{deg} \) (one degree in radians).
- Dimensions, automatically expressed in points, e.g., \( \text{pc} \) is 12.
- Automatic conversion (no need for \( \text{\langle type\rangle\_use:N} \)) of integer, dimension, and skip variables to floating point numbers, expressing dimensions in points and ignoring the stretch and shrink components of skips.
- Tuples: \((x_1,\ldots,x_n)\) that can be stored in variables, added together, multiplied or divided by a floating point number, and nested.

Floating point numbers can be given either explicitly (in a form such as \( 1.234\times10^{-34} \), or \( -0.0001 \)), or as a stored floating point variable, which is automatically replaced by its current value. A “floating point” is a floating point number or a tuple thereof. See section 28.9.1 for a description of what a floating point is, section 28.9.2 for details about how an expression is parsed, and section 28.9.3 to know what the various operations do. Some operations may raise exceptions (error messages), described in section 28.7.

An example of use could be the following.

\LaTeX{} can now compute: $ \frac{\sin(3.5)}{2} + 2\cdot10^{-3} = \ExplSyntaxOn \fp_to_decimal:n {\sin(3.5)/2 + 2e-3} $.

The operation \text{round} can be used to limit the result’s precision. Adding \(+0\) avoids the possibly undesirable output \(-0\), replacing it by \(+0\). However, the \text{l3fp} module is mostly meant as an underlying tool for higher-level commands. For example, one could provide a function to typeset nicely the result of floating point computations.

\documentclass{article}
\usepackage{xparse, siunitx}
\ExplSyntaxOn
\NewDocumentCommand { \calcnum } { m } { \num { \fp_to_scientific:n {#1} } }
\ExplSyntaxOff
\begin{document}
\calcnum { 2 \pi * \sin ( 2.3 ^ 5 ) }
\end{document}

See the documentation of \text{siunitx} for various options of \text{\num}.
28.1 Creating and initialising floating point variables

\texttt{\textbackslash fp\_new:N} \texttt{(fp\ var)}

Creates a new \texttt{⟨fp var⟩} or raises an error if the name is already taken. The declaration is global. The \texttt{⟨fp var⟩} is initially +0.

\texttt{\textbackslash fp\_const:Nn} \texttt{(fp\ var) \{⟨floating point expression⟩\}}

Creates a new constant \texttt{⟨fp var⟩} or raises an error if the name is already taken. The \texttt{⟨fp var⟩} is set globally equal to the result of evaluating the \texttt{⟨floating point expression⟩}.

\texttt{\textbackslash fp\_zero:N} \texttt{(fp\ var)}

Sets the \texttt{⟨fp var⟩} to +0.

\texttt{\textbackslash fp\_zero_new:N} \texttt{(fp\ var)}

Ensures that the \texttt{⟨fp var⟩} exists globally by applying \texttt{\textbackslash fp\_new:N} if necessary, then applies \texttt{\textbackslash fp\_(g)zero:N} to leave the \texttt{⟨fp var⟩} set to +0.

28.2 Setting floating point variables

\texttt{\textbackslash fp\_set:Nn} \texttt{(fp\ var) \{⟨floating point expression⟩\}}

Sets \texttt{⟨fp var⟩} equal to the result of computing the \texttt{⟨floating point expression⟩}.

\texttt{\textbackslash fp\_set_eq:NN} \texttt{(fp\ var\_1) (fp\ var\_2)}

Sets the floating point variable \texttt{⟨fp var\_1⟩} equal to the current value of \texttt{⟨fp var\_2⟩}.

\texttt{\textbackslash fp\_add:Nn} \texttt{(fp\ var) \{⟨floating point expression⟩\}}

Adds the result of computing the \texttt{⟨floating point expression⟩} to the \texttt{⟨fp var⟩}. This also applies if \texttt{⟨fp var⟩} and \texttt{⟨floating point expression⟩} evaluate to tuples of the same size.
\fp_sub:Nn \fp_var \{\textit{floating point expression}\}
Subtracts the result of computing the \textit{floating point expression} from the \textit{fp var}. This also applies if \textit{fp var} and \textit{floating point expression} evaluate to tuples of the same size.

\section*{28.3 Using floating points}

\fp_eval:n \{\textit{floating point expression}\}
Evaluates the \textit{floating point expression} and expresses the result as a decimal number with no exponent. Leading or trailing zeros may be inserted to compensate for the exponent. Non-significant trailing zeros are trimmed, and integers are expressed without a decimal separator. The values $\pm\infty$ and NaN trigger an “invalid operation” exception. For a tuple, each item is converted using \fp_eval:n and they are combined as $(\{fp_1\}, \{fp_2\}, \ldots, \{fp_n\})$ if $n > 1$ and $(\{fp_1\})$ or {} for fewer items. This function is identical to \fp_to_decimal:n.

\fp_sign:n \{\textit{fpexpr}\}
Evaluates the \textit{fpexpr} and leaves its sign in the input stream using \fp_eval:n \{\texttt{sign(}\{\textit{result}\}\texttt{)}\}: $+1$ for positive numbers and for $+\infty$, $-1$ for negative numbers and for $-\infty$, $\pm 0$ for $\pm 0$. If the operand is a tuple or is NaN, then “invalid operation” occurs and the result is 0.

\fp_to_decimal:N \fp_var
\fp_to_decimal:n \{\textit{floating point expression}\}
Evaluates the \textit{floating point expression} and expresses the result as a decimal number with no exponent. Leading or trailing zeros may be inserted to compensate for the exponent. Non-significant trailing zeros are trimmed, and integers are expressed without a decimal separator. The values $\pm\infty$ and NaN trigger an “invalid operation” exception. For a tuple, each item is converted using \fp_to_decimal:n and they are combined as $(\{fp_1\}, \{fp_2\}, \ldots, \{fp_n\})$ if $n > 1$ and $(\{fp_1\})$ or {} for fewer items.

\fp_to_int:N \fp_var
\fp_to_int:n \{\textit{floating point expression}\}
Evaluates the \textit{floating point expression}, and rounds the result to the closest integer, rounding exact ties to an even integer. The result may be outside the range $[-2^{31} + 1, 2^{31} - 1]$ of valid \TeX\ dimensions, leading to overflow errors if used in an integer expression. Tuples, as well as the values $\pm\infty$ and NaN, trigger an “invalid operation” exception.
\texttt{fp\_to\_scientific:N}\star
\texttt{fp\_to\_scientific:c}\star
\texttt{fp\_to\_scientific:n}\star

\texttt{fp\_to\_tl:N}\star
\texttt{fp\_to\_tl:c}\star
\texttt{fp\_to\_tl:n}\star

\texttt{fp\_use:N}\star
\texttt{fp\_use:c}\star

\texttt{fp\_if\_exist\_p:N}\star
\texttt{fp\_if\_exist\_p:c}\star
\texttt{fp\_if\_exist:NTF}\star
\texttt{fp\_if\_exist:CF}\star

28.4 Floating point conditionals
\fp_compare_p:nNn  \fp_compare_p:nNn \{\textit{fpexpr}_1\} \{\textit{relation}\} \{\textit{fpexpr}_2\}
\fp_compare:nNnTF \{\textit{fpexpr}_1\} \{\textit{relation}\} \{\textit{fpexpr}_2\} \{(true\ code)\} \{(false\ code)\}

Compares the \textit{fpexpr}_1\ and the \textit{fpexpr}_2, and returns \texttt{true} if the \textit{relation} is obeyed. Two floating points \textit{x} and \textit{y} may obey four mutually exclusive relations: \textit{x} < \textit{y}, \textit{x} = \textit{y}, \textit{x} > \textit{y}, or \textit{x}?\textit{y} ("not ordered"). The last case occurs exactly if one or both operands is NaN or is a tuple, unless they are equal tuples. Note that a NaN is distinct from any value, even another NaN, hence \textit{x} = \textit{x} is not true for a NaN. To test if a value is NaN, compare it to an arbitrary number with the "not ordered" relation.

\begin{verbatim}
\fp_compare:nNnTF { <value> } ? { 0 }
{ } % <value> is nan
{ } % <value> is not nan
\end{verbatim}

Tuples are equal if they have the same number of items and items compare equal (in particular there must be no NaN). At present any other comparison with tuples yields \? (not ordered). This is experimental.

This function is less flexible than \fp_compare:nTF but slightly faster. It is provided for consistency with \int_compare:nNnTF and \dim_compare:nNnTF.
\fp_compare_p:n \fp_compare:nTF
\fp_compare_p:n
{  \(\textit{fpexpr}_1\) \(\textit{relation}_1\)
  \ldots
  \(\textit{fpexpr}_N\) \(\textit{relation}_N\)
  \(\textit{fpexpr}_{N+1}\)
}
\fp_compare:nTF
{  \(\textit{fpexpr}_1\) \(\textit{relation}_1\)
  \ldots
  \(\textit{fpexpr}_N\) \(\textit{relation}_N\)
  \(\textit{fpexpr}_{N+1}\)
} {
  \{\{\textit{true code}\}\}\{\{\textit{false code}\}\}
}

Evaluates the \textit{(floating point expressions)} as described for \texttt{\fp_eval:n} and compares consecutive result using the corresponding \textit{(relation)}, namely it compares \(\textit{fpexpr}_1\) and \(\textit{fpexpr}_2\) using the \(\textit{relation}_1\), then \(\textit{fpexpr}_2\) and \(\textit{fpexpr}_3\) using the \(\textit{relation}_2\), until finally comparing \(\textit{fpexpr}_N\) and \(\textit{fpexpr}_{N+1}\) using the \(\textit{relation}_N\). The test yields \texttt{true} if all comparisons are \texttt{true}. Each \textit{(floating point expression)} is evaluated only once.

Contrarily to \texttt{\int_compare:nTF}, all \textit{(floating point expressions)} are computed, even if one comparison is \texttt{false}. Two floating points \(x\) and \(y\) may obey four mutually exclusive relations: \(x < y\), \(x = y\), \(x > y\), or \(x\neq y\) ("not ordered"). The last case occurs exactly if one or both operands is NaN or is a tuple, unless they are equal tuples. Each \textit{(relation)} can be any (non-empty) combination of \(<\), \(=\), \(>\), and \(?\), plus an optional leading \(!\) (which negates the \textit{(relation)}), with the restriction that the \textit{(relation)} may not start with \(?\), as this symbol has a different meaning (in combination with : ) within floating point expressions. The comparison \(x \textit{(relation)} y\) is then \texttt{true} if the \textit{(relation)} does not start with \(!\) and the actual relation (\(<\), \(=\), \(>\), or \(?\)) between \(x\) and \(y\) appears within the \textit{(relation)}, or on the contrary if the \textit{(relation)} starts with \(!\) and the relation between \(x\) and \(y\) does not appear within the \textit{(relation)}. Common choices of \textit{(relation)} include \(\geq\) (greater or equal), \(!=\) (not equal), \(!?\) or \(<=>\) (comparable).

This function is more flexible than \texttt{\fp_compare:nNnTF} and only slightly slower.

\section{Floating point expression loops}

\fp_do_until:nNnn \fp_do_until:nNnn
\fp_do_until:nNnn \{\textit{fpexpr}_1\} \{\textit{relation}\} \{\textit{fpexpr}_2\} \{\langle\textit{code}\rangle\}

Places the \textit{(code)} in the input stream for \TeX{} to process, and then evaluates the relationship between the two \textit{(floating point expressions)} as described for \texttt{\fp_compare:nNnTF}. If the test is \texttt{false} then the \textit{(code)} is inserted into the input stream again and a loop occurs until the \textit{(relation)} is \texttt{true}.

\fp_do_while:nNnn \fp_do_while:nNnn
\fp_do_while:nNnn \{\textit{fpexpr}_1\} \{\textit{relation}\} \{\textit{fpexpr}_2\} \{\langle\textit{code}\rangle\}

Places the \textit{(code)} in the input stream for \TeX{} to process, and then evaluates the relationship between the two \textit{(floating point expressions)} as described for \texttt{\fp_compare:nNnTF}. If the test is \texttt{true} then the \textit{(code)} is inserted into the input stream again and a loop occurs until the \textit{(relation)} is \texttt{false}.
\texttt{\textbackslash fp\_until\_do:nNnn} \texttt{\textdagger} \texttt{\{fexpr1\} \{relation\} \{fexpr2\} \{\texttt{code}\}}

Evaluates the relationship between the two \textit{floating point expressions} as described for \texttt{\textbackslash fp\_compare:nNnTF}, and then places the \texttt{\{code\}} in the input stream if the \texttt{\{relation\}} is \texttt{false}. After the \texttt{\{code\}} has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{true}.

\texttt{\textbackslash fp\_while\_do:nNnn} \texttt{\textdagger} \texttt{\{fexpr1\} \{relation\} \{fexpr2\} \{\texttt{code}\}}

Evaluates the relationship between the two \textit{floating point expressions} as described for \texttt{\textbackslash fp\_compare:nNnTF}, and then places the \texttt{\{code\}} in the input stream if the \texttt{\{relation\}} is \texttt{true}. After the \texttt{\{code\}} has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{false}.

\texttt{\textbackslash fp\_do\_until:nn} \texttt{\textdagger} \texttt{\{fexpr1\} \{relation\} \{fexpr2\} \{\texttt{code}\}}

Places the \texttt{\{code\}} in the input stream for \TeX to process, and then evaluates the relationship between the two \textit{floating point expressions} as described for \texttt{\textbackslash fp\_compare:nNnTF}. If the test is \texttt{false} then the \texttt{\{code\}} is inserted into the input stream again and a loop occurs until the \texttt{\{relation\}} is \texttt{true}.

\texttt{\textbackslash fp\_do\_while:nn} \texttt{\textdagger} \texttt{\{fexpr1\} \{relation\} \{fexpr2\} \{\texttt{code}\}}

Places the \texttt{\{code\}} in the input stream for \TeX to process, and then evaluates the relationship between the two \textit{floating point expressions} as described for \texttt{\textbackslash fp\_compare:nNnTF}. If the test is \texttt{true} then the \texttt{\{code\}} is inserted into the input stream again and a loop occurs until the \texttt{\{relation\}} is \texttt{false}.

\texttt{\textbackslash fp\_until\_do:nn} \texttt{\textdagger} \texttt{\{fexpr1\} \{relation\} \{fexpr2\} \{\texttt{code}\}}

Evaluates the relationship between the two \textit{floating point expressions} as described for \texttt{\textbackslash fp\_compare:nNnTF}, and then places the \texttt{\{code\}} in the input stream if the \texttt{\{relation\}} is \texttt{false}. After the \texttt{\{code\}} has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{true}.

\texttt{\textbackslash fp\_while\_do:nn} \texttt{\textdagger} \texttt{\{fexpr1\} \{relation\} \{fexpr2\} \{\texttt{code}\}}

Evaluates the relationship between the two \textit{floating point expressions} as described for \texttt{\textbackslash fp\_compare:nNnTF}, and then places the \texttt{\{code\}} in the input stream if the \texttt{\{relation\}} is \texttt{true}. After the \texttt{\{code\}} has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{false}.  

244
This function first evaluates the ⟨initial value⟩, ⟨step⟩ and ⟨final value⟩, each of which should be a floating point expression evaluating to a floating point number, not a tuple. The ⟨function⟩ is then placed in front of each ⟨value⟩ from the ⟨initial value⟩ to the ⟨final value⟩ in turn (using ⟨step⟩ between each ⟨value⟩). The ⟨step⟩ must be non-zero. If the ⟨step⟩ is positive, the loop stops when the ⟨value⟩ becomes larger than the ⟨final value⟩. If the ⟨step⟩ is negative, the loop stops when the ⟨value⟩ becomes smaller than the ⟨final value⟩. The ⟨function⟩ should absorb one numerical argument. For example

\cs_set:Npn \my_func:n #1 { [I saw #1] \quad }
\fp_step_function:nnnN { 1.0 } { 0.1 } { 1.5 } \my_func:n

would print

[I saw 1.0] [I saw 1.1] [I saw 1.2] [I saw 1.3] [I saw 1.4] [I saw 1.5]

\textbf{\TeX} hackers note: Due to rounding, it may happen that adding the ⟨step⟩ to the ⟨value⟩ does not change the ⟨value⟩; such cases give an error, as they would otherwise lead to an infinite loop.

This function first evaluates the ⟨initial value⟩, ⟨step⟩ and ⟨final value⟩, all of which should be floating point expressions evaluating to a floating point number, not a tuple. Then for each ⟨value⟩ from the ⟨initial value⟩ to the ⟨final value⟩ in turn (using ⟨step⟩ between each ⟨value⟩), the ⟨code⟩ is inserted into the input stream with #1 replaced by the current ⟨value⟩. Thus the ⟨code⟩ should define a function of one argument (#1).

This function first evaluates the ⟨initial value⟩, ⟨step⟩ and ⟨final value⟩, all of which should be floating point expressions evaluating to a floating point number, not a tuple. Then for each ⟨value⟩ from the ⟨initial value⟩ to the ⟨final value⟩ in turn (using ⟨step⟩ between each ⟨value⟩), the ⟨code⟩ is inserted into the input stream, with the ⟨tl var⟩ defined as the current ⟨value⟩. Thus the ⟨code⟩ should make use of the ⟨tl var⟩.

### 28.6 Some useful constants, and scratch variables

\c_zero_fp
\c_minus_zero_fp

Zero, with either sign.

\c_one_fp

One as an \textit{fp}: useful for comparisons in some places.
Infinity, with either sign. These can be input directly in a floating point expression as `inf` and `-inf`.

The value of the base of the natural logarithm, $e = \exp(1)$.

The value of $\pi$. This can be input directly in a floating point expression as `pi`.

The value of $1^\circ$ in radians. Multiply an angle given in degrees by this value to obtain a result in radians. Note that trigonometric functions expecting an argument in radians or in degrees are both available. Within floating point expressions, this can be accessed as `deg`.

Scratch floating points for local assignment. These are never used by the kernel code, and so are safe for use with any \LTTEX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

Scratch floating points for global assignment. These are never used by the kernel code, and so are safe for use with any \LTTEX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

28.7 Floating point exceptions

The functions defined in this section are experimental, and their functionality may be altered or removed altogether.

“Exceptions” may occur when performing some floating point operations, such as $0 / 0$, or $10 ^ {1e9999}$. The relevant IEEE standard defines 5 types of exceptions, of which we implement 4.

- **Overflow** occurs whenever the result of an operation is too large to be represented as a normal floating point number. This results in $\pm \infty$.

- **Underflow** occurs whenever the result of an operation is too close to 0 to be represented as a normal floating point number. This results in $\pm 0$.

- **Invalid operation** occurs for operations with no defined outcome, for instance $0 / 0$ or $\sin(\infty)$, and results in a NaN. It also occurs for conversion functions whose target type does not have the appropriate infinite or NaN value (e.g., `\fp_to_dim:n`).

- **Division by zero** occurs when dividing a non-zero number by 0, or when evaluating functions at poles, e.g., $\ln(0)$ or $\cot(0)$. This results in $\pm \infty$. 
(not yet) *Inexact* occurs whenever the result of a computation is not exact, in other words, almost always. At the moment, this exception is entirely ignored in \( \LaTeX \).

To each exception we associate a “flag”: \texttt{fp\_overflow}, \texttt{fp\_underflow}, \texttt{fp\_invalid\_operation} and \texttt{fp\_division\_by\_zero}. The state of these flags can be tested and modified with commands from \texttt{l3flag}.

By default, the “invalid operation” exception triggers an (expandable) error, and raises the corresponding flag. Other exceptions raise the corresponding flag but do not trigger an error. The behaviour when an exception occurs can be modified (using \texttt{\fp\_trap:nn}) to either produce an error and raise the flag, or only raise the flag, or do nothing at all.

\[
\texttt{\fp\_trap:nn \{(exception)\} \{(trap type)\}}
\]

All occurrences of the \{(exception)\} (overflow, underflow, invalid\_operation or division\_by\_zero) within the current group are treated as \{(trap type)\}, which can be

- **none**: the \{(exception)\} will be entirely ignored, and leave no trace;
- **flag**: the \{(exception)\} will turn the corresponding flag on when it occurs;
- **error**: additionally, the \{(exception)\} will halt the \( \TeX \) run and display some information about the current operation in the terminal.

*This function is experimental, and may be altered or removed.*

---

Flags denoting the occurrence of various floating-point exceptions.

- \texttt{flag\_fp\_overflow}
- \texttt{flag\_fp\_underflow}
- \texttt{flag\_fp\_invalid\_operation}
- \texttt{flag\_fp\_division\_by\_zero}

---

### 28.8 Viewing floating points

\[
\texttt{\fp\_show:N \{fp\ var\}}
\]

Evaluates the \{(floating point expression)\} and displays the result in the terminal.

\[
\texttt{\fp\_log:N \{fp\ var\}}
\]

Evaluates the \{(floating point expression)\} and writes the result in the log file.
28.9 Floating point expressions

28.9.1 Input of floating point numbers

We support four types of floating point numbers:

- $\pm m \cdot 10^n$, a floating point number, with integer $1 \leq m \leq 10^{16}$, and $-10000 \leq n \leq 10000$;
- $\pm 0$, zero, with a given sign;
- $\pm \infty$, infinity, with a given sign;
- NaN, is “not a number”, and can be either quiet or signalling (not yet: this distinction is currently unsupported);

Normal floating point numbers are stored in base 10, with up to 16 significant figures.

On input, a normal floating point number consists of:

- $\langle \text{sign} \rangle$: a possibly empty string of + and − characters;
- $\langle \text{significand} \rangle$: a non-empty string of digits together with zero or one dot;
- $\langle \text{exponent} \rangle$ optionally: the character $e$ or $E$, followed by a possibly empty string of + and − tokens, and a non-empty string of digits.

The sign of the resulting number is + if $\langle \text{sign} \rangle$ contains an even number of −, and − otherwise, hence, an empty $\langle \text{sign} \rangle$ denotes a non-negative input. The stored significand is obtained from $\langle \text{significand} \rangle$ by omitting the decimal separator and leading zeros, and rounding to 16 significant digits, filling with trailing zeros if necessary. In particular, the value stored is exact if the input $\langle \text{significand} \rangle$ has at most 16 digits. The stored $\langle \text{exponent} \rangle$ is obtained by combining the input $\langle \text{exponent} \rangle$ (0 if absent) with a shift depending on the position of the significand and the number of leading zeros.

A special case arises if the resulting $\langle \text{exponent} \rangle$ is either too large or too small for the floating point number to be represented. This results either in an overflow (the number is then replaced by $\pm \infty$), or an underflow (resulting in $\pm 0$).

The result is thus $\pm 0$ if and only if $\langle \text{significand} \rangle$ contains no non-zero digit (i.e., consists only in characters 0, and an optional period), or if there is an underflow. Note that a single dot is currently a valid floating point number, equal to $+0$, but that is not guaranteed to remain true.

The $\langle \text{significand} \rangle$ must be non-empty, so e1 and e-1 are not valid floating point numbers. Note that the latter could be mistaken with the difference of “e” and 1. To avoid confusions, the base of natural logarithms cannot be input as $e$ and should be input as $\exp(1)$ or $\mathtt{c_e_fp}$ (which is faster).

Special numbers are input as follows:

- $\mathtt{inf}$ represents $+\infty$, and can be preceded by any $\langle \text{sign} \rangle$, yielding $\pm \infty$ as appropriate.
- $\mathtt{nan}$ represents a (quiet) non-number. It can be preceded by any sign, but that sign is ignored.
- Any unrecognizable string triggers an error, and produces a NaN.
- Note that commands such as $\mathtt{\text{inf\text{ty}}}$, $\mathtt{\text{pi}}$, or $\mathtt{\text{sin}}$ do not work in floating point expressions. They may silently be interpreted as completely unexpected numbers, because integer constants (allowed in expressions) are commonly stored as mathematical characters.
28.9.2 Precedence of operators

We list here all the operations supported in floating point expressions, in order of decreasing precedence: operations listed earlier bind more tightly than operations listed below them.

- Function calls ($\sin$, $\ln$, etc).
- Binary ** and ~ (right associative).
- Unary +, −, !.
- Implicit multiplication by juxtaposition ($2\pi$) when neither factor is in parentheses.
- Binary * and /, implicit multiplication by juxtaposition with parentheses (for instance $3(4+5)$).
- Binary + and −.
- Comparisons $\geq$, $\neq$, $<$, $>$, etc.
- Logical and, denoted by $\&\&$.
- Logical or, denoted by $||$.
- Ternary operator $?:$ (right associative).
- Comma (to build tuples).

The precedence of operations can be overridden using parentheses. In particular, the precedence of juxtaposition implies that

\[
\frac{1}{2\pi} = \frac{1}{(2\pi)}, \\
\frac{1}{2\pi}(\pi + \pi) = (2\pi)^{-1}(\pi + \pi) \approx 1, \\
\sin 2\pi = \sin(2\pi) \neq 0, \\
2^{2\max(3,5)} = 2^2 \max(3,5) = 20, \\
1\text{in}/1\text{cm} = (1\text{in})/(1\text{cm}) = 2.54.
\]

Functions are called on the value of their argument, contrarily to \TeX macros.

28.9.3 Operations

We now present the various operations allowed in floating point expressions, from the lowest precedence to the highest. When used as a truth value, a floating point expression is false if it is ±0, and true otherwise, including when it is NaN or a tuple such as (0, 0). Tuples are only supported to some extent by operations that work with truth values ($?:$, $||$, $\&\&$, !), by comparisons ($!<$, $<=$, $>$), and by +, −, *, /.

Unless otherwise specified, providing a tuple as an argument of any other operation yields the “invalid operation” exception and a NaN result.
The ternary operator `?:` results in `⟨operand2⟩` if `⟨operand1⟩` is true (not ±0), and `⟨operand3⟩` if `⟨operand1⟩` is false (±0). All three `⟨operands⟩` are evaluated in all cases; they may be tuples. The operator is right associative, hence

\[ \text{\texttt{\textbackslash fp\_eval:n \{ (operand1) ? (operand2) : (operand3) \}}} \]

first tests whether `1 + 3 > 4`; since this isn’t true, the branch following `:` is taken, and `2 + 4 > 5` is compared; since this is true, the branch before `:` is taken, and everything else is (evaluated then) ignored. That allows testing for various cases in a concise manner, with the drawback that all computations are made in all cases.

\[ \text{\texttt{\textbackslash fp\_eval:n \{ (operand1) || (operand2) \}}} \]

If `⟨operand1⟩` is true (not ±0), use that value, otherwise the value of `⟨operand2⟩`. Both `⟨operands⟩` are evaluated in all cases; they may be tuples. In `⟨operand1⟩ || ⟨operand2⟩ || ... || ⟨operandsn⟩`, the first true (nonzero) `⟨operand⟩` is used and if all are zero the last one (±0) is used.

\[ \text{\texttt{\textbackslash fp\_eval:n \{ (operand1) && (operand2) \}}} \]

If `⟨operand1⟩` is false (equal to ±0), use that value, otherwise the value of `⟨operand2⟩`. Both `⟨operands⟩` are evaluated in all cases; they may be tuples. In `⟨operand1⟩ && ⟨operand2⟩ && ... && ⟨operandsn⟩`, the first false (±0) `⟨operand⟩` is used and if none is zero the last one is used.

Each `⟨relation⟩` consists of a non-empty string of `<`, `=`, `>`, and `?`, optionally preceded by `, and may not start with `. This evaluates to `1` if all comparisons `⟨operandi⟩ (relationi)` `⟨operandsi+1⟩` are true, and `0` otherwise. All `⟨operands⟩` are evaluated (once) in all cases. See \texttt{\textbackslash fp\_compare:nTF} for details.

\[ \text{\texttt{\textbackslash fp\_eval:n \{ (operand1) + (operand2) \}}} \]

\[ \text{\texttt{\textbackslash fp\_eval:n \{ (operand1) - (operand2) \}}} \]

Computes the sum or the difference of its two `⟨operands⟩`. The “invalid operation” exception occurs for `∞ - ∞`. “Underflow” and “overflow” occur when appropriate. These operations support the itemwise addition or subtraction of two tuples, but if they have a different number of items the “invalid operation” exception occurs and the result is NaN.
\* \fp_eval:n { \langle \text{operand} \rangle \ast \langle \text{operand} \rangle } \\
/ \fp_eval:n { \langle \text{operand} \rangle / \langle \text{operand} \rangle }

Computes the product or the ratio of its two \textit{(operands)}. The “invalid operation” exception occurs for $\infty / \infty$, $0/0$, or $0 \ast \infty$. “Division by zero” occurs when dividing a finite non-zero number by $\pm 0$. “Underflow” and “overflow” occur when appropriate. When $\langle \text{operand} \rangle_1$ is a tuple and $\langle \text{operand} \rangle_2$ is a floating point number, each item of $\langle \text{operand} \rangle_1$ is multiplied or divided by $\langle \text{operand} \rangle_2$. Other combinations yield an “invalid operation” exception and a NaN result.

\+ \fp_eval:n { \langle \text{operand} \rangle + } \\
- \fp_eval:n { \langle \text{operand} \rangle - } \\
! \fp_eval:n { \langle \text{operand} \rangle }

The unary + does nothing, the unary – changes the sign of the $\langle \text{operand} \rangle$ (for a tuple, of all its components), and ! $\langle \text{operand} \rangle$ evaluates to 1 if $\langle \text{operand} \rangle$ is false (is $\pm 0$) and 0 otherwise (this is the \texttt{not} boolean function). Those operations never raise exceptions.

\** \fp_eval:n { \langle \text{operand} \rangle \ast\ast \langle \text{operand} \rangle } \\
\^ \fp_eval:n { \langle \text{operand} \rangle \ast\^ \langle \text{operand} \rangle }

Raises $\langle \text{operand} \rangle_1$ to the power $\langle \text{operand} \rangle_2$. This operation is right associative, hence $2 \ast\ast 2 \ast\ast 3 = 256$. If $\langle \text{operand} \rangle_1$ is negative or $-0$ then: the result’s sign is $+$ if the $\langle \text{operand} \rangle_2$ is infinite and $(-1)^p$ if the $\langle \text{operand} \rangle_2$ is $p/q$ with $p, q$ integers; the result is $+0$ if $\text{abs}(\langle \text{operand} \rangle_1)^\langle \text{operand} \rangle_2$ evaluates to zero; in other cases the “invalid operation” exception occurs because the sign cannot be determined. “Division by zero” occurs when raising $\pm 0$ to a finite strictly negative power. “Underflow” and “overflow” occur when appropriate. If either operand is a tuple, “invalid operation” occurs.

abs \fp_eval:n { \text{abs}( \langle \text{fexpr} \rangle ) }

Computes the absolute value of the $\langle \text{fexpr} \rangle$. If the operand is a tuple, “invalid operation” occurs. This operation does not raise exceptions in other cases. See also $\text{\texttt{\textbackslash{fp\_abs:n}}}$.

\exp \fp_eval:n { \exp( \langle \text{fexpr} \rangle ) }

Computes the exponential of the $\langle \text{fexpr} \rangle$. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

\fact \fp_eval:n { \text{fact}( \langle \text{fexpr} \rangle ) }

Computes the factorial of the $\langle \text{fexpr} \rangle$. If the $\langle \text{fexpr} \rangle$ is an integer between $-0$ and $3248$ included, the result is finite and correctly rounded. Larger positive integers give $+\infty$ with “overflow”, while $\text{\texttt{\textbackslash{fact}(}+\infty) = +\infty$ and $\text{\texttt{fact(nan)}} = \text{\texttt{nan}}$ with no exception. All other inputs give NaN with the “invalid operation” exception.

\ln \fp_eval:n { \ln( \langle \text{fexpr} \rangle ) }

Computes the natural logarithm of the $\langle \text{fexpr} \rangle$. Negative numbers have no (real) logarithm, hence the “invalid operation” is raised in that case, including for $\text{\texttt{ln(-0)} = -\infty$. “Division by zero” occurs when evaluating $\ln(+0) = -\infty$. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.
**logb** \* 
\fp_eval:n { logb( \langle fpexpr \rangle ) } 

Determines the exponent of the \langle fpexpr \rangle, namely the floor of the base-10 logarithm of its absolute value. “Division by zero” occurs when evaluating \text{logb}(\pm 0) = -\infty. Other special values are \text{logb}(\pm \infty) = +\infty and \text{logb}(\text{NaN}) = \text{NaN}. If the operand is a tuple or is \text{NaN}, then “invalid operation” occurs and the result is \text{NaN}.

**max** 
\fp_eval:n { max( \langle fpexpr1 \rangle, \langle fpexpr2 \rangle, \ldots ) } 
\fp_eval:n { min( \langle fpexpr1 \rangle, \langle fpexpr2 \rangle, \ldots ) } 

Evaluates each \langle fpexpr \rangle and computes the largest (smallest) of those. If any of the \langle fpexpr \rangle is a \text{NaN} or tuple, the result is \text{NaN}. If any operand is a tuple, “invalid operation” occurs; these operations do not raise exceptions in other cases.

**round** 
\fp_eval:n { round ( \langle fpexpr \rangle ) } 
\fp_eval:n { round ( \langle fpexpr1 \rangle, \langle fpexpr2 \rangle ) } 
\fp_eval:n { round ( \langle fpexpr1 \rangle, \langle fpexpr2 \rangle, \langle fpexpr3 \rangle ) } 

Only \texttt{round} accepts a third argument. Evaluates \langle fpexpr1 \rangle = x and \langle fpexpr2 \rangle = n and \langle fpexpr3 \rangle = t then rounds \(x\) to \(n\) places. If \(n\) is an integer, this rounds \(x\) to a multiple of \(10^{-n}\); if \(n = +\infty\), this always yields \(x\); if \(n = -\infty\), this yields one of \(\pm 0\), \(\pm \infty\), or \(\text{NaN}\); if \(n = \text{NaN}\), this yields \(\text{NaN}\); if \(n\) is neither \(\pm \infty\) nor an integer, then an “invalid operation” exception is raised. When \langle fpexpr2 \rangle is omitted, \(n = 0\), i.e., \langle fpexpr1 \rangle is rounded to an integer. The rounding direction depends on the function.

- \texttt{round} yields the multiple of \(10^{-n}\) closest to \(x\), with ties (\(x\) half-way between two such multiples) rounded as follows. If \(t\) is \text{nan} (or not given) the even multiple is chosen (“ties to even”), if \(t = \pm 0\) the multiple closest to 0 is chosen (“ties to zero”), if \(t\) is positive/negative the multiple closest to \(\infty/-\infty\) is chosen (“ties towards positive/negative infinity”).
- \texttt{floor} yields the largest multiple of \(10^{-n}\) smaller or equal to \(x\) (“round towards negative infinity”);
- \texttt{ceil} yields the smallest multiple of \(10^{-n}\) greater or equal to \(x\) (“round towards positive infinity”);
- \texttt{trunc} yields a multiple of \(10^{-n}\) with the same sign as \(x\) and with the largest absolute value less than that of \(x\) (“round towards zero”).

“Overflow” occurs if \(x\) is finite and the result is infinite (this can only happen if \langle fpexpr2 \rangle < -9984). If any operand is a tuple, “invalid operation” occurs.

**sign** 
\fp_eval:n { sign( \langle fpexpr \rangle ) } 

Evaluates the \langle fpexpr \rangle and determines its sign: \(+1\) for positive numbers and for \(+\infty\), \(-1\) for negative numbers and for \(-\infty\), \(\pm 0\) for \(\pm 0\), and \text{NaN} for \text{NaN}. If the operand is a tuple, “invalid operation” occurs. This operation does not raise exceptions in other cases.
Computes the sine, cosine, tangent, cotangent, cosecant, or secant of the \( \langle \text{fpexpr} \rangle \) given in radians. For arguments given in degrees, see \text{sind}, \text{cosd}, etc. Note that since \( \pi \) is irrational, \( \sin(8\pi) \) is not quite zero, while its analogue \( \text{sind}(8 \times 180) \) is exactly zero. The trigonometric functions are undefined for an argument of \( \pm \infty \), leading to the “invalid operation” exception. Additionally, evaluating tangent, cotangent, cosecant, or secant at one of their poles leads to a “division by zero” exception. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

Updated: 2013-11-17
\texttt{atan} \begin{verbatim} \fp_eval:n { atan( \fpexpr ) } \end{verbatim} \texttt{acot} \begin{verbatim} \fp_eval:n { acot( \fpexpr ) } \end{verbatim} \texttt{atand} \begin{verbatim} \fp_eval:n { atand( \fpexpr ) } \end{verbatim} \texttt{acotd} \begin{verbatim} \fp_eval:n { acotd( \fpexpr ) } \end{verbatim} \texttt{sqrt} \begin{verbatim} \fp_eval:n { sqrt( \fpexpr ) } \end{verbatim}

Those functions yield an angle in radians: \texttt{atand} and \texttt{acotd} are their analogs in degrees. The one-argument versions compute the arctangent or arccotangent of the \texttt{fpexpr}: arctangent takes values in the range \([-\pi/2, \pi/2]\), and arccotangent in the range \([0, \pi]\). The two-argument arctangent computes the angle in polar coordinates of the point with Cartesian coordinates \((\fpexpr_2, \fpexpr_1)\): this is the arctangent of \(\fpexpr_1/\fpexpr_2\), possibly shifted by \(\pi\) depending on the signs of \(\fpexpr_1\) and \(\fpexpr_2\). The two-argument arccotangent computes the angle in polar coordinates of the point \((\fpexpr_1, \fpexpr_2)\), equal to the arccotangent of \(\fpexpr_1/\fpexpr_2\), possibly shifted by \(\pi\). Both two-argument functions take values in the wider range \([-\pi, \pi]\). The ratio \(\fpexpr_1/\fpexpr_2\) need not be defined for the two-argument arctangent: when both expressions yield \(\pm 0\), or when both yield \(\pm \infty\), the resulting angle is one of \(\{\pm \pi/4, \pm 3\pi/4\}\) depending on signs. The “underflow” exception can occur. If any operand is a tuple, “invalid operation” occurs.

\texttt{atan} \begin{verbatim} \fp_eval:n { atan( \fpexpr ) } \end{verbatim} \texttt{acot} \begin{verbatim} \fp_eval:n { acot( \fpexpr ) } \end{verbatim} \texttt{atand} \begin{verbatim} \fp_eval:n { atand( \fpexpr ) } \end{verbatim} \texttt{acotd} \begin{verbatim} \fp_eval:n { acotd( \fpexpr ) } \end{verbatim}

Those functions yield an angle in degrees: \texttt{atand} and \texttt{acotd} are their analogs in radians. The one-argument versions compute the arctangent or arccotangent of the \texttt{fpexpr}: arctangent takes values in the range \([-90, 90]\), and arccotangent in the range \([0, 180]\). The two-argument arctangent computes the angle in polar coordinates of the point with Cartesian coordinates \((\fpexpr_2, \fpexpr_1)\): this is the arctangent of \(\fpexpr_1/\fpexpr_2\), possibly shifted by 180 depending on the signs of \(\fpexpr_1\) and \(\fpexpr_2\). The two-argument arccotangent computes the angle in polar coordinates of the point \((\fpexpr_1, \fpexpr_2)\), equal to the arccotangent of \(\fpexpr_1/\fpexpr_2\), possibly shifted by 180. Both two-argument functions take values in the wider range \([-180, 180]\]. The ratio \(\fpexpr_1/\fpexpr_2\) need not be defined for the two-argument arctangent: when both expressions yield \(\pm 0\), or when both yield \(\pm \infty\), the resulting angle is one of \(\{\pm 45, \pm 135\}\) depending on signs. The “underflow” exception can occur. If any operand is a tuple, “invalid operation” occurs.

\texttt{sqrt} \begin{verbatim} \fp_eval:n { sqrt( \fpexpr ) } \end{verbatim}

Computes the square root of the \texttt{fpexpr}. The “invalid operation” is raised when the \texttt{fpexpr} is negative or is a tuple; no other exception can occur. Special values yield \(\sqrt{-0} = -0\), \(\sqrt{+0} = +0\), \(\sqrt{-\infty} = +\infty\) and \(\sqrt{NaN} = NaN\).
\fp_eval:n \{ rand() \}

New: 2016-12-05

Produces a pseudo-random floating-point number (multiple of $10^{-16}$) between 0 included and 1 excluded. This is not available in older versions of Xe\TeX. The random seed can be queried using \sys_rand_seed: and set using \sys_gset_rand_seed:n.

**\TeXhackers note:** This is based on pseudo-random numbers provided by the engine’s primitive \pdfuniformdeviate in pdf\TeX, \pdf\TeX, \up\TeX and \uniformdeviate in Lua\TeX and Xe\TeX. The underlying code is based on Metapost, which follows an additive scheme recommended in Section 3.6 of “The Art of Computer Programming, Volume 2”.

While we are more careful than \uniformdeviate to preserve uniformity of the underlying stream of 28-bit pseudo-random integers, these pseudo-random numbers should of course not be relied upon for serious numerical computations nor cryptography.

\fp_eval:n \{ randint(⟨fpexpr⟩) \}
\fp_eval:n \{ randint(⟨fpexpr⟩₁, ⟨fpexpr⟩₂) \}

New: 2016-12-05

Produces a pseudo-random integer between 1 and ⟨fpexpr⟩ or between ⟨fpexpr₁⟩ and ⟨fpexpr₂⟩ inclusive. The bounds must be integers in the range $(-10^{16}, 10^{16})$ and the first must be smaller or equal to the second. See rand for important comments on how these pseudo-random numbers are generated.

\c_{inf_fp}
\c_{minus_inf_fp}
\c_{nan_fp}

The special values $+∞$, $-∞$, and NaN are represented as inf, -inf and nan (see \c_{inf_fp}, \c_{minus_inf_fp} and \c_{nan_fp}).

\c_{pi_fp}

The value of π (see \c_{pi_fp}).

\c_{one_degree_fp}

The value of 1° in radians (see \c_{one_degree_fp}).
Those units of measurement are equal to their values in pt, namely

\begin{align*}
1 \text{ in} &= 72.27 \text{ pt} \\
1 \text{ pt} &= 1 \text{ pt} \\
1 \text{ pc} &= 12 \text{ pt} \\
1 \text{ cm} &= \frac{1}{25.4} \text{ in} = 84.5275590551181 \text{ pt} \\
1 \text{ mm} &= \frac{1}{25.4} \text{ in} = 2.845275590551181 \text{ pt} \\
1 \text{ dd} &= 0.376065 \text{ mm} = 1.07000856496063 \text{ pt} \\
1 \text{ cc} &= 12 \text{ dd} = 12.84010277952756 \text{ pt} \\
1 \text{ nd} &= 0.375 \text{ mm} = 1.066978346456693 \text{ pt} \\
1 \text{ nc} &= 12 \text{ nd} = 12.80374015748031 \text{ pt} \\
1 \text{ bp} &= \frac{1}{72} \text{ in} = 1.00375 \text{ pt} \\
1 \text{ sp} &= 2^{-16} \text{ pt} = 1.52587890625 \times 10^{-5} \text{ pt}.
\end{align*}

The values of the (font-dependent) units \texttt{em} and \texttt{ex} are gathered from \TeX when the surrounding floating point expression is evaluated.

Other names for 1 and \texttt{+0}.

\begin{itemize}
\item \texttt{\fp_abs:n} {\langle floating point expression\rangle}
\end{itemize}

Evaluates the \texttt{\langle floating point expression\rangle} as described for \texttt{\fp_eval:n} and leaves the absolute value of the result in the input stream. If the argument is $\pm\infty$, NaN or a tuple, “invalid operation” occurs. Within floating point expressions, \texttt{abs()} can be used; it accepts $\pm\infty$ and NaN as arguments.

\begin{itemize}
\item \texttt{\fp_max:nn} {\langle fp expression 1\rangle} {\langle fp expression 2\rangle}
\end{itemize}

Evaluates the \texttt{\langle floating point expressions\rangle} as described for \texttt{\fp_eval:n} and leaves the resulting larger (\texttt{max}) or smaller (\texttt{min}) value in the input stream. If the argument is a tuple, “invalid operation” occurs, but no other case raises exceptions. Within floating point expressions, \texttt{max()} and \texttt{min()} can be used.

\section{Disclaimer and roadmap}

The package may break down if the escape character is among 0123456789+, or if it receives a \TeX primitive conditional affected by \texttt{\exp_not:N}.

The following need to be done. I’ll try to time-order the items.

\begin{itemize}
\item Function to count items in a tuple (and to determine if something is a tuple).
\item Decide what exponent range to consider.
\end{itemize}
• Support signalling \texttt{nan}.
• Modulo and remainder, and rounding function \texttt{quantize} (and its friends analogous to \texttt{trunc, ceil, floor}).
• \texttt{\_fp\_format:nn \{\_fpexpr\}\{\_format\}}, but what should \{\_format\} be? More general pretty printing?
• Add \texttt{and}, \texttt{or}, \texttt{xor}? Perhaps under the names \texttt{all, any, xor}?
• Add $\log(x, b)$ for logarithm of $x$ in base $b$.
• \texttt{hypot} (Euclidean length). Cartesian-to-polar transform.
• Hyperbolic functions $\cosh, \sinh, \tanh$.
• Inverse hyperbolics.
• Base conversion, input such as \texttt{0xAB.CDEF}.
• Factorial (not with \texttt{!}), gamma function.
• Improve coefficients of the $\sin$ and $\tan$ series.
• Treat upper and lower case letters identically in identifiers, and ignore underscores.
• Add an \texttt{array(1,2,3)} and \texttt{i=complex(0,1)}.
• Provide an experimental \texttt{map} function? Perhaps easier to implement if it is a single character, \texttt{@sin(1,2)}?
• Provide an \texttt{isnan} function analogue of \texttt{\_fp\_if\_nan:nTF}?
• Support keyword arguments?

\texttt{Pgfmath} also provides box-measurements (depth, height, width), but boxes are not possible expandably.

Bugs, and tests to add.
• Check that functions are monotonic when they should.
• Add exceptions to \texttt{?:, !<=>?, \&\&, ||, and !}.
• Logarithms of numbers very close to 1 are inaccurate.
• When rounding towards $-\infty$, \texttt{\dim\_to\_fp:n \{0pt\}} should return $-0$, not $+0$.
• The result of $(\pm 0) + (\pm 0)$, of $x + (-x)$, and of $(-x) + x$ should depend on the rounding mode.
• \texttt{0e9999999999} gives a TeX “number too large” error.
• Subnormals are not implemented.

Possible optimizations/improvements.
• Document that \texttt{l3trial/l3fp-types} introduces tools for adding new types.
• In subsection 28.9.1, write a grammar.
• It would be nice if the \texttt{parse} auxiliaries for each operation were set up in the corresponding module, rather than centralizing in \texttt{l3fp-parse}.

• Some functions should get an \texttt{\_o} ending to indicate that they expand after their result.

• More care should be given to distinguish expandable/restricted expandable (auxiliary and internal) functions.

• The code for the \texttt{ternary} set of functions is ugly.

• There are many \texttt{-} missing in the doc to avoid bad line-breaks.

• The algorithm for computing the logarithm of the significand could be made to use a 5 terms Taylor series instead of 10 terms by taking $c = 2000/([200x] + 1) \in [10, 95]$ instead of $c \in [1, 10]$. Also, it would then be possible to simplify the computation of $t$. However, we would then have to hard-code the logarithms of 44 small integers instead of 9.

• Improve notations in the explanations of the division algorithm (\texttt{l3fp-basics}).

• Understand and document \texttt{\_\_fp_basics_pack_weird_low:NNNNw} and \texttt{\_\_fp_- basics_pack_weird_high:NNNNNNNNw} better. Move the other \texttt{basics_pack} auxiliaries to \texttt{l3fp-aux} under a better name.

• Find out if underflow can really occur for trigonometric functions, and redoc as appropriate.

• Add bibliography. Some of Kahan’s articles, some previous T\texttt{E}X fp packages, the international standards,…

• Also take into account the “inexact” exception?

• Support multi-character prefix operators (\texttt{e.g., @} or whatever)?
Chapter 29

The l3fparray package: fast global floating point arrays

29.1 l3fparray documentation

For applications requiring heavy use of floating points, this module provides arrays which
can be accessed in constant time (contrast l3seq, where access time is linear). The
interface is very close to that of l3intarray. The size of the array is fixed and must be
given at point of initialisation

\fparray_new:Nn \fparray_new:Nn \langle fparray \ var \rangle \{\langle size\rangle\}

Evaluates the integer expression \langle size\rangle and allocates an \langle floating point array variable\rangle
with that number of (zero) entries. The variable name should start with \g_ because
assignments are always global.

\fparray_count:N \fparray_count:N \star \fparray_count:N \star

Expands to the number of entries in the \langle floating point array variable\rangle. This is performed
in constant time.

\fparray_gset:Nnn \fparray_gset:Nnn \langle fparray \ var \rangle \{\langle position\rangle\} \{\langle value\rangle\}

Stores the result of evaluating the floating point expression \langle value\rangle into the \langle floating point
array variable\rangle at the (integer expression) \langle position\rangle. If the \langle position\rangle is not between 1
and the \fparray_count:N, an error occurs. Assignments are always global.

\fparray_gzero:N \fparray_gzero:N \fparray_gzero:N \star \fparray_gzero:N \star

Sets all entries of the \langle floating point array variable\rangle to +0. Assignments are always global.

\fparray_item:Nn \fparray_item:Nn \star \fparray_item_to_tl:Nn \star \fparray_item_to_tl:Nn \star

Applies \fp_use:N or \fp_to_tl:N (respectively) to the floating point entry stored at
the (integer expression) \langle position\rangle in the \langle floating point array variable\rangle. If the \langle position\rangle
is not between 1 and the \fparray_count:N, an error occurs.
Chapter 30

The l3cctab package
Category code tables

A category code table enables rapid switching of all category codes in one operation. For LuaLaTeX, this is possible over the entire Unicode range. For other engines, only the 8-bit range (0–255) is covered by such tables.

30.1 Creating and initialising category code tables

\cctab_new:N \cctab_new:c \cctab_const:Nn \cctab_const:cn \cctab_gset:Nn \cctab_gset:cn

\cctab_new:N \cctab_new:c
\cctab_const:Nn \cctab_const:cn
\cctab_gset:Nn \cctab_gset:cn

Creates a new \emph{category code table} variable or raises an error if the name is already taken. The declaration is global. The \emph{category code table} is initialised with the codes as used by iniTeX.

Creates a new \emph{category code table}, applies (in a group) the \emph{category code set up} on top of iniTeX settings, then saves them globally as a constant table. The \emph{category code set up} can include a call to \cctab_select:N.

Starting from the iniTeX category codes, applies (in a group) the \emph{category code set up}, then saves them globally in the \emph{category code table}. The \emph{category code set up} can include a call to \cctab_select:N.

30.2 Using category code tables

\cctab_begin:N \cctab_begin:c
\cctab_end:

Switches locally the category codes in force to those stored in the \emph{category code table}. The prevailing codes before the function is called are added to a stack, for use with \cctab_end:. This function does not start a \TeX{} group.
\end{verbatim}

\begin{verbatim}
\ctab_end:
\end{verbatim}

\texttt{\ctab_end:}

Ends the scope of a \emph{category code table} started using \texttt{\ctab_begin:N}, returning the codes to those in force before the matching \texttt{\ctab_begin:N} was used. This must be used within the same \TeX{} group (and at the same \TeX{} group level) as the matching \texttt{\ctab_begin:N}.

\begin{verbatim}
\ctab_select:N \ctab_select:c
\end{verbatim}

\texttt{\ctab_select:N} \texttt{\ctab_select:c}

Selects the \emph{category code table} for the scope of the current group. This is in particular useful in the \emph{setup} arguments of \texttt{\tl_set_rescan:Nnn}, \texttt{\tl_rescan:nn}, \texttt{\ctab_const:Nn}, and \texttt{\ctab_gset:Nn}.

\begin{verbatim}
\ctab_item:Nn \ctab_item:cn
\end{verbatim}

\texttt{\ctab_item:Nn} \texttt{\ctab_item:cn}

Determines the \emph{character} with character code given by the \emph{(integer expression)} and expands to its category code specified by the \emph{category code table}.

\subsection*{30.3 Category code table conditionals}

\begin{verbatim}
\ctab_if_exist_p:N \ctab_if_exist_p:c \ctab_if_exist:NTF \ctab_if_exist:cTF
\end{verbatim}

\texttt{\ctab_if_exist_p:N} \texttt{\ctab_if_exist_p:c} \texttt{\ctab_if_exist:NTF} \texttt{\ctab_if_exist:cTF}

Tests whether the \emph{category code table} is currently defined. This does not check that the \emph{category code table} really is a category code table.

\subsection*{30.4 Constant category code tables}

\begin{verbatim}
\c_code_cctab \c_document_cctab \c_initex_cctab \c_other_cctab \c_str_cctab
\end{verbatim}

\texttt{\c_code_cctab} \texttt{\c_document_cctab} \texttt{\c_initex_cctab} \texttt{\c_other_cctab} \texttt{\c_str_cctab}

Category code table for the expl3 code environment; this does not include $\emptyset$, which is retained as an “other” character.

Category code table for a standard \LaTeX{} document, as set by the \LaTeX{} kernel. In particular, the upper-half of the 8-bit range will be set to “active” with pdf\TeX{} only. No babel shorthands will be activated.

Category code table as set up by ini\TeX{}.

Category code table where all characters have category code 12 (other).

Category code table where all characters have category code 12 (other) with the exception of spaces, which have category code 10 (space).
Part V
Text manipulation
Chapter 31

The l3unicode package: Unicode support functions

This module provides Unicode-specific functions along with loading data from a range of Unicode Consortium files. At present, it provides no public functions.
Chapter 32

The \texttt{l3text} package: text processing

This module deals with manipulation of (formatted) text; such material is comprised of a restricted set of token list content. The functions provided here concern conversion of textual content for example in case changing, generation of bookmarks and extraction to tags. All of the major functions operate by expansion. Begin-group and end-group tokens in the \langle \text {text} \rangle are normalized and become \{ and \}, respectively.

32.1 Expanding text

\texttt{\texttt{text\_expand:n}} \star \texttt{\texttt{text\_expand:n} \{(text)\}}

Takes user input \langle \text {text} \rangle and expands the content. Protected commands (typically formatting) are left in place, and no processing takes place of math mode material (as delimited by pairs given in \langle \texttt{l\_text\_math\_delims\_tl} or as the argument to commands listed in \langle \texttt{l\_text\_math\_arg\_tl}). Commands which are neither engine- nor \LaTeXX protected are expanded exhaustively. Any commands listed in \langle \texttt{l\_text\_expand\_exclude\_tl}, \langle \texttt{l\_text\_accents\_tl} and \langle \texttt{l\_text\_letterlike\_tl} are excluded from expansion.

\texttt{\texttt{text\_declare\_expand\_equivalent:Nn}} \texttt{\texttt{text\_declare\_expand\_equivalent:cn}}

\texttt{\texttt{text\_declare\_expand\_equivalent:Nn}} \langle \texttt{cmd} \rangle \{\langle \texttt{replacement} \rangle \}

\texttt{\texttt{text\_declare\_expand\_equivalent:cn}}

Declares that the \langle \texttt{replacement} \rangle tokens should be used whenever the \langle \texttt{cmd} \rangle (a single token) is encountered. The \langle \texttt{replacement} \rangle tokens should be expandable.
32.2 Case changing

\text_lowercase:n \text_uppercase:n \text_uppercase:nn \text_titlecase:n \text_titlecase_first:n \text_lowercase:nn

Takes user input \textit{(text)} first applies \texttt{text\_expand}, then transforms the case of character tokens as specified by the function name. The category code of letters are not changed by this process (at least where they can be represented by the engine as a single token: 8-bit engines may require active characters).

Upper- and lowercase have the obvious meanings. Titlecasing may be regarded informally as converting the first character of the \textit{(tokens)} to uppercase and the rest to lowercase. However, the process is more complex than this as there are some situations where a single lowercase character maps to a special form, for example \texttt{ij} in Dutch which becomes \texttt{IJ}. The \texttt{titlecase\_first} variant does not attempt any case changing at all after the first letter has been processed.

Importantly, notice that these functions are intended for working with user \textit{text\_for\_typesetting}. For case changing programmatic data see the \texttt{l3str} module and discussion there of \texttt{str\_lowercase:n}, \texttt{str\_uppercase:n} and \texttt{str\_foldcase:n}.

Case changing does not take place within math mode material so for example

\texttt{\text_uppercase:n \{ Some-text-$y = mx + c$-with-\{Braces\} \}}

becomes

\texttt{SOME TEXT $y = mx + c$ WITH \{BRACES\}}

The arguments of commands listed in \texttt{l\_text\_case\_exclude\_arg\_tl} are excluded from case changing; the latter are entirely non-textual content (such as labels).

As is generally true for expl3, these functions are designed to work with Unicode input only. As such, UTF-8 input is assumed for all engines. When used with Xe\TeX\ or Lua\TeX\ a full range of Unicode transformations are enabled. Specifically, the standard mappings here follow those defined by the Unicode Consortium in UnicodeData.txt and SpecialCasing.txt. In the case of 8-bit engines, mappings are provided for characters which can be represented in output typeset using the T1, T2 and LGR font encodings. Thus for example \texttt{ä} can be case-changed using pdf\TeX. For \texttt{p\TeX} only the ASCII range is covered as the engine treats input outside of this range as east Asian.

Language-sensitive conversions are enabled using the \texttt{(language)} argument, and follow Unicode Consortium guidelines. Currently, the languages recognised for special handling are as follows.

- Azeri and Turkish (az and tr). The case pairs I/i-dotless and I-dot/i are activated for these languages. The combining dot mark is removed when lowercasing I-dot and introduced when upper casing i-dotless.

- German (de-alt). An alternative mapping for German in which the lowercase \textit{Eszett} maps to a \textit{großes Eszett}. Since there is a T1 slot for the \textit{großes Eszett} in T1, this tailoring \textit{is} available with pdf\TeX as well as in the Unicode \TeX engines.
• Greek (el). Removes accents from Greek letters when uppercasing; titlecasing leaves accents in place. (At present this is implemented only for Unicode engines.)

• Lithuanian (lt). The lowercase letters i and j should retain a dot above when the accents grave, acute or tilde are present. This is implemented for lowercasing of the relevant uppercase letters both when input as single Unicode codepoints and when using combining accents. The combining dot is removed when uppercasing in these cases. Note that only the accents used in Lithuanian are covered: the behaviour of other accents are not modified.

• Dutch (nl). Capitalisation of ij at the beginning of titlecased input produces IJ rather than Ij. The output retains two separate letters, thus this transformation is available using pdftex.

For titlecasing, note that there are two functions available. The function \text_titlecase:n applies (broadly) uppercasing to the first letter of the input, then lowercasing to the remainder. In contrast, \text_titlecase_first:n only carries out the uppercasing operation, and leaves the balance of the input unchanged. Determining whether non-letter characters at the start of text should switch from upper- to lowercasing is controllable. When \l_text_titlecase_check_letter_bool is true, characters which are not letters (category code 11) are left unchanged and “skipped”: the first letter is uppercased. (With 8-bit engines, this is extended to active characters which form part of a multi-byte letter codepoint.) When \l_text_titlecase_check_letter_bool is false, the first character is uppercased, and the rest lowercased, irrespective of the nature of the character.

32.3 Removing formatting from text

\text_purify:n \{⟨text⟩\}

Takes user input ⟨text⟩ and expands as described for \text_expand:n, then removes all functions from the resulting text. Math mode material (as delimited by pairs given in \l_text_math_delims_tl or as the argument to commands listed in \l_text_math_arg_tl) is left contained in a pair of $ delimiters. Non-expandable functions present in the ⟨text⟩ must either have a defined equivalent (see \text_declare_purify_equivalent:Nn) or will be removed from the result. Implicit tokens are converted to their explicit equivalent.

\text_declare_purify_equivalent:Nn \{⟨cmd⟩\} \{⟨replacement⟩\}
\text_declare_purify_equivalent:Nx

Declares that the ⟨replacement⟩ tokens should be used whenever the ⟨cmd⟩ (a single token) is encountered. The ⟨replacement⟩ tokens should be expandable.

32.4 Control variables

\l_text_accents_tl  Lists commands which represent accents, and which are left unchanged by expansion. (Defined only for the \LaTeX \texttt{2e} package.)
\ltext_letterlike_tl \textbf{ Lists commands which represent letters; these are left unchanged by expansion. (Defined only for the \LaTeX\ package.) }

\ltext_math_arg_tl \textbf{ Lists commands present in the \texttt{⟨text⟩} where the argument of the command should be treated as math mode material. The treatment here is similar to \ltext_math_delims_tl but for a command rather than paired delimiters. }

\ltext_math_delims_tl \textbf{ Lists pairs of tokens which delimit (in-line) math mode content; such content \textit{may} be excluded from processing. }

\ltext_case_exclude_arg_tl \textbf{ Lists commands which are excluded from case changing. }

\ltext_expand_exclude_tl \textbf{ Lists commands which are excluded from expansion. }

\ltext_titlecase_check_letter_bool \textbf{ Controls how the start of titlecasing is handled: when \texttt{true}, the first \textit{letter} in text is considered. The standard setting is \texttt{true}. }
Part VI
Typesetting
Chapter 33

The l3box package

Boxes

Box variables contain typeset material that can be inserted on the page or in other boxes. Their contents cannot be converted back to lists of tokens. There are three kinds of box operations: horizontal mode denoted with prefix \hbox_, vertical mode with prefix \vbox_, and the generic operations working in both modes with prefix \box_. For instance, a new box variable containing the words “Hello, world!” (in a horizontal box) can be obtained by the following code.

\box_new:N \l_hello_box
\hbox_set:Nn \l_hello_box { Hello, ~ world! }

The argument is typeset inside a TEX group so that any variables assigned during the construction of this box restores its value afterwards.

Box variables from l3box are compatible with those of \LaTeX{} 2ε and plain \TeX{} and can be used interchangeably. The l3box commands to construct boxes, such as \hbox:n or \hbox_set:Nn, are “color-safe”, meaning that

\hbox:n { \color_select:n { blue } Hello, } ~ world!

will result in “Hello,” taking the color blue, but “world!” remaining with the prevailing color outside the box.

### 33.1 Creating and initialising boxes

\begin{Verbatim}
\box_new:N \box_new:c
\box_clear:N \box_clear:c \box_gclear:N \box_gclear:c
\end{Verbatim}

\begin{itemize}
  \item \texttt{\box_new:N} ⟨\texttt{box}⟩
  Creates a new ⟨\texttt{box}⟩ or raises an error if the name is already taken. The declaration is global. The ⟨\texttt{box}⟩ is initially void.
  \item \texttt{\box_clear:N} ⟨\texttt{box}⟩
  Clears the content of the ⟨\texttt{box}⟩ by setting the box equal to \texttt{\c_empty_box}.  
\end{itemize}
\box_clear_new:N \box_clear_new:c
\box_gclear_new:N \box_gclear_new:c

\box_set_eq:NN \box_gset_eq:NN
\box_if_exist_p:N \box_if_exist:NTF \box_if_exist:c TF

\box_use:N \box_use:c

\box_move_right:nn \box_move_right:nn
\box_move_left:nn \box_move_left:nn
\box_move_up:nn \box_move_up:nn
\box_move_down:nn \box_move_down:nn

Ensures that the \texttt{(box)} exists globally by applying \texttt{\box_new:N} if necessary, then applies \texttt{\box_(g)clear:N} to leave the \texttt{(box)} empty.

\box_set_eq:NN \box_gset_eq:NN
Sets the content of \texttt{(box1)} equal to that of \texttt{(box2)}.

\box_if_exist_p:N \box_if_exist:NTF \box_if_exist:c TF
Tests whether the \texttt{(box)} is currently defined. This does not check that the \texttt{(box)} really is a box.

33.2 Using boxes

\box_use:N \box_use:c
Inserts the current content of the \texttt{(box)} onto the current list for typesetting. An error is raised if the variable does not exist or if it is invalid.

\TeXhackers note: This is the \TeX primitive \texttt{\copy}.

\box_move_right:nn \box_move_right:nn
\box_move_left:nn \box_move_left:nn
\box_move_up:nn \box_move_up:nn
\box_move_down:nn \box_move_down:nn
This function operates in vertical mode, and inserts the material specified by the \texttt{(box function)} such that its reference point is displaced horizontally by the given \texttt{(dimexpr)} from the reference point for typesetting, to the right or left as appropriate. The \texttt{(box function)} should be a box operation such as \texttt{\box_use:N \box} or a “raw” box specification such as \texttt{\vbox:n \{ xyz \}}.

\box_move_up:nn \box_move_up:nn
\box_move_down:nn \box_move_down:nn
This function operates in horizontal mode, and inserts the material specified by the \texttt{(box function)} such that its reference point is displaced vertically by the given \texttt{(dimexpr)} from the reference point for typesetting, up or down as appropriate. The \texttt{(box function)} should be a box operation such as \texttt{\box_use:N \box} or a “raw” box specification such as \texttt{\vbox:n \{ xyz \}}.
### 33.3 Measuring and setting box dimensions

\texttt{\textbackslash box\_dp:N} \texttt{\textbackslash box\_dp:c} \texttt{\textbackslash box\_dp:N \{box\}}

Calculates the depth (below the baseline) of the \texttt{\{box\}} in a form suitable for use in a \textit{dimension expression}.

\textbf{\TeX{}hackers note:} This is the \TeX{} primitive \texttt{\dp{}}.

\texttt{\textbackslash box\_ht:N} \texttt{\textbackslash box\_ht:c} \texttt{\textbackslash box\_ht:N \{box\}}

Calculates the height (above the baseline) of the \texttt{\{box\}} in a form suitable for use in a \textit{dimension expression}.

\textbf{\TeX{}hackers note:} This is the \TeX{} primitive \texttt{\ht{}}.

\texttt{\textbackslash box\_wd:N} \texttt{\textbackslash box\_wd:c} \texttt{\textbackslash box\_wd:N \{box\}}

Calculates the width of the \texttt{\{box\}} in a form suitable for use in a \textit{dimension expression}.

\textbf{\TeX{}hackers note:} This is the \TeX{} primitive \texttt{\wd{}}.

\texttt{\textbackslash box\_ht\_plus\_dp:N} \texttt{\textbackslash box\_ht\_plus\_dp:c} \texttt{\textbackslash box\_ht\_plus\_dp:N \{box\}}

Calculates the total vertical size (height plus depth) of the \texttt{\{box\}} in a form suitable for use in a \textit{dimension expression}.

\texttt{\textbackslash box\_set\_dp:Nn} \texttt{\textbackslash box\_set\_dp:cn}

\texttt{\textbackslash box\_gset\_dp:Nn} \texttt{\textbackslash box\_gset\_dp:cn}

\texttt{\textbackslash box\_set\_dp:N \{dimension expression\}}

Set the depth (below the baseline) of the \texttt{\{box\}} to the value of the \texttt{\{dimension expression\}}.

\texttt{\textbackslash box\_set\_ht:Nn} \texttt{\textbackslash box\_set\_ht:cn}

\texttt{\textbackslash box\_gset\_ht:Nn} \texttt{\textbackslash box\_gset\_ht:cn}

\texttt{\textbackslash box\_set\_ht:N \{dimension expression\}}

Set the height (above the baseline) of the \texttt{\{box\}} to the value of the \texttt{\{dimension expression\}}.

\texttt{\textbackslash box\_set\_wd:Nn} \texttt{\textbackslash box\_set\_wd:cn}

\texttt{\textbackslash box\_gset\_wd:Nn} \texttt{\textbackslash box\_gset\_wd:cn}

\texttt{\textbackslash box\_set\_wd:N \{dimension expression\}}

Set the width of the \texttt{\{box\}} to the value of the \texttt{\{dimension expression\}}.
33.4 Box conditionals

\box_if_empty_p:N \box_if_empty_p:c \box_if_empty:NTF \box_if_empty:cTF

\box_if_empty_p:N \box_if_empty_p:c \box_if_empty:NTF \box_if_empty:cTF

\box_if_horizontal_p:N \box_if_horizontal_p:c \box_if_horizontal:NTF \box_if_horizontal:cTF

\box_if_horizontal_p:N \box_if_horizontal_p:c \box_if_horizontal:NTF \box_if_horizontal:cTF

\box_if_vertical_p:N \box_if_vertical_p:c \box_if_vertical:NTF \box_if_vertical:cTF

\box_if_vertical_p:N \box_if_vertical_p:c \box_if_vertical:NTF \box_if_vertical:cTF

Tests if \box is a empty (equal to \c_empty_box).

Tests if \box is a horizontal box.

Tests if \box is a vertical box.

33.5 The last box inserted

\box_set_to_last:N \box_set_to_last:c \box_gset_to_last:N \box_gset_to_last:c

\box_set_to_last:N \box_set_to_last:c

Sets the \box equal to the last item (box) added to the current partial list, removing the item from the list at the same time. When applied to the main vertical list, the \box is always void as it is not possible to recover the last added item.

33.6 Constant boxes

\c_empty_box

This is a permanently empty box, which is neither set as horizontal nor vertical.

\TeXhackers note: At the \TeX level this is a void box.

33.7 Scratch boxes

\l_tmpa_box \l_tmpb_box

\l_tmpa_box \l_tmpb_box

Scratch boxes for local assignment. These are never used by the kernel code, and so are safe for use with any \EP\TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_box \g_tmpb_box

\g_tmpa_box \g_tmpb_box

Scratch boxes for global assignment. These are never used by the kernel code, and so are safe for use with any \EP\TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

Updated: 2012-11-04

Updated: 2012-11-04

272
33.8 Viewing box contents

\box_show:N \box_show:N (box)
Shows full details of the content of the (box) in the terminal.

\box_show:Nnn (box) \{\text{intexpr}_1\} \{\text{intexpr}_2\}
Display the contents of (box) in the terminal, showing the first \text{intexpr}_1 items of the box, and descending into \text{intexpr}_2 group levels.

\box_log:N \box_log:C
Writes full details of the content of the (box) to the log.

\box_log:Nnn (box) \{\text{intexpr}_1\} \{\text{intexpr}_2\}
Writes the contents of (box) to the log, showing the first \text{intexpr}_1 items of the box, and descending into \text{intexpr}_2 group levels.

33.9 Boxes and color

All \LaTeX3 boxes are “color safe”: a color set inside the box stops applying after the end of the box has occurred.

33.10 Horizontal mode boxes

\hbox:n \{\text{contents}\}
Typesets the \text{contents} into a horizontal box of natural width and then includes this box in the current list for typesetting.

\hbox_to_wd:nn \{\dimexpr\} \{\text{contents}\}
Typesets the \text{contents} into a horizontal box of width \dimexpr and then includes this box in the current list for typesetting.

\hbox_to_zero:n \{\text{contents}\}
Typesets the \text{contents} into a horizontal box of zero width and then includes this box in the current list for typesetting.

\hbox_set:Nn \hbox_set:cn \hbox_gset:Nn \hbox_gset:cn
Typesets the \text{contents} at natural width and then stores the result inside the (box).
\hbox_set_to_wd:Nnn \hbox_set_to_wd:cnn \hbox_gset_to_wd:Nnn \hbox_gset_to_wd:cnn

Updated: 2017-04-05

\hbox_overlap_center:n \{\textit{contents}\}

Typesets the \textit{contents} into a horizontal box of zero width such that material protrudes equally to both sides of the insertion point.

\hbox_overlap_right:n \{\textit{contents}\}

Updated: 2017-04-05

Typesets the \textit{contents} into a horizontal box of zero width such that material protrudes to the right of the insertion point.

\hbox_overlap_left:n \{\textit{contents}\}

Updated: 2017-04-05

Typesets the \textit{contents} into a horizontal box of zero width such that material protrudes to the left of the insertion point.

\hbox_set:Nw \hbox_set:cw \hbox_set_end:

Updated: 2017-04-05

\hbox_set_to_wd:Nnw \hbox_set_to_wd:cnw \hbox_gset_to_wd:Nnw \hbox_gset_to_wd:cnw

Updated: 2017-06-08

\hbox_unpack:N \hbox_unpack:c

Unpacks the content of the horizontal \textit{box}, retaining any stretching or shrinking applied when the \textit{box} was set.

\TeXhackers note: This is the \TeX primitive \unhcopy.

### 33.11 Vertical mode boxes

Vertical boxes inherit their baseline from their contents. The standard case is that the baseline of the box is at the same position as that of the last item added to the box. This means that the box has no depth unless the last item added to it had depth. As a result, most vertical boxes have a large height value and small or zero depth. The exception are
_top boxes, where the reference point is that of the first item added. These tend to have a large depth and small height, although the latter is typically non-zero.

\vbox:n \{(contents)\}
Typesets the \{(contents)\} into a vertical box of natural height and includes this box in the current list for typesetting.

\vbox_top:n \{(contents)\}
Typesets the \{(contents)\} into a vertical box of natural height and includes this box in the current list for typesetting. The baseline of the box is equal to that of the first item added to the box.

\vbox_to_ht:nn \{\dimexpr\} \{(contents)\}
Typesets the \{(contents)\} into a vertical box of height \dimexpr and then includes this box in the current list for typesetting.

\vbox_to_zero:n \{(contents)\}
Typesets the \{(contents)\} into a vertical box of zero height and then includes this box in the current list for typesetting.

\vbox_set:Nn \{box\} \{(contents)\}
Typesets the \{(contents)\} at natural height and then stores the result inside the \{box\}.

\vbox_set_top:Nn \{box\} \{(contents)\}
Typesets the \{(contents)\} at natural height and then stores the result inside the \{box\}. The baseline of the box is equal to that of the first item added to the box.

\vbox_set_to_ht:Nnn \{box\} \{(dimexpr)\} \{(contents)\}
Typesets the \{(contents)\} to the height given by the \{dimexpr\} and then stores the result inside the \{box\}.

\vbox_set:Nw \{box\} \{(contents)\} \vbox_set_end:
Typesets the \{(contents)\} at natural height and then stores the result inside the \{box\}. In contrast to \vbox_set:Nn this function does not absorb the argument when finding the \{content\}, and so can be used in circumstances where the \{content\} may not be a simple argument.
Typesets the \langle contents \rangle to the height given by the \langle dimexpr \rangle and then stores the result inside the \langle box \rangle. In contrast to \vbox_set_to_ht:Nnn this function does not absorb the argument when finding the \langle content \rangle, and so can be used in circumstances where the \langle content \rangle may not be a simple argument.

\vbox_set_split_to_ht:NNn \langle box1 \rangle \langle box2 \rangle \{ \langle dimexpr \rangle \}

Sets \langle box1 \rangle to contain material to the height given by the \langle dimexpr \rangle by removing content from the top of \langle box2 \rangle (which must be a vertical box).

\vbox_unpack:N \langle box \rangle
\vbox_unpack:c

Unpacks the content of the vertical \langle box \rangle, retaining any stretching or shrinking applied when the \langle box \rangle was set.

\TeXhacker{s note:} This is the \TeX primitive \texttt{unvcopy}.

### 33.12 Using boxes efficiently

The functions above for using box contents work in exactly the same way as for any other expl3 variable. However, for efficiency reasons, it is also useful to have functions which drop box contents on use. When a box is dropped, the box becomes empty at the group level where the box was originally set rather than necessarily at the current group level. For example, with

\hbox_set:Nn \l_tmpa_box { A }
\group_begin:
\hbox_set:Nn \l_tmpa_box { B }
\group_begin:
\box_use_drop:N \l_tmpa_box
\group_end:
\box_show:N \l_tmpa_box
\group_end:
\box_show:N \l_tmpa_box

the first use of \texttt{\box_show:N} will show an entirely cleared (void) box, and the second will show the letter \texttt{A} in the box.

These functions should be preferred when the content of the box is no longer required after use. Note that due to the unusual scoping behaviour of \texttt{drop} functions they may be applied to both local and global boxes: the latter will naturally be set and thus cleared at a global level.
\box_use_drop:N \box_use_drop:c

Inserts the current content of the \textit{box} onto the current list for typesetting then drops the box content. An error is raised if the variable does not exist or if it is invalid. This function may be applied to local or global boxes.

\textbf{\LaTeX{}hackers note:} This is the \texttt{\box} primitive.

\box_set_eq_drop:NN \box_set_eq_drop:(cN|cN|cc)

Sets the content of \texttt{\box} equal to that of \texttt{\box}, then drops \texttt{\box}.

\textbf{\LaTeX{}hackers note:} This is the \texttt{\box} primitive.

\box_gset_eq_drop:NN \box_gset_eq_drop:(cN|cN|cc)

Sets the content of \texttt{\box} globally equal to that of \texttt{\box}, then drops \texttt{\box}.

\textbf{\LaTeX{}hackers note:} This is the \texttt{\box} primitive.

\hbox_unpack_drop:N \hbox_unpack_drop:c

Unpacks the content of the horizontal \texttt{\box}, retaining any stretching or shrinking applied when the \texttt{\box} was set. The original \texttt{\box} is then dropped.

\textbf{\LaTeX{}hackers note:} This is the \texttt{\box} primitive.

\vbox_unpack_drop:N \vbox_unpack_drop:c

Unpacks the content of the vertical \texttt{\box}, retaining any stretching or shrinking applied when the \texttt{\box} was set. The original \texttt{\box} is then dropped.

\textbf{\LaTeX{}hackers note:} This is the \texttt{\box} primitive.

\subsection{Affine transformations}

Affine transformations are changes which (informally) preserve straight lines. Simple translations are affine transformations, but are better handled in \LaTeX{} by doing the translation first, then inserting an unmodified box. On the other hand, rotation and resizing of boxed material can best be handled by modifying boxes. These transformations are described here.
Resizes the \( \langle \text{box} \rangle \) to fit within the given \( \langle x\text{-size} \rangle \) (horizontally) and \( \langle y\text{-size} \rangle \) (vertically); both of the sizes are dimension expressions. The \( \langle y\text{-size} \rangle \) is the height only: it does not include any depth. The updated \( \langle \text{box} \rangle \) is an hbox, irrespective of the nature of the \( \langle \text{box} \rangle \) before the resizing is applied. The final size of the \( \langle \text{box} \rangle \) is the smaller of \( \{ \langle x\text{-size} \rangle \} \) and \( \{ \langle y\text{-size} \rangle \} \), i.e., the result fits within the dimensions specified. Negative sizes cause the material in the \( \langle \text{box} \rangle \) to be reversed in direction, but the reference point of the \( \langle \text{box} \rangle \) is unchanged. Thus a negative \( \langle y\text{-size} \rangle \) results in the \( \langle \text{box} \rangle \) having a depth dependent on the height of the original and vice versa.

Resizes the \( \langle \text{box} \rangle \) to fit within the given \( \langle x\text{-size} \rangle \) (horizontally) and \( \langle y\text{-size} \rangle \) (vertically); both of the sizes are dimension expressions. The \( \langle y\text{-size} \rangle \) is the total vertical size (height plus depth). The updated \( \langle \text{box} \rangle \) is an hbox, irrespective of the nature of the \( \langle \text{box} \rangle \) before the resizing is applied. The final size of the \( \langle \text{box} \rangle \) is the smaller of \( \{ \langle x\text{-size} \rangle \} \) and \( \{ \langle y\text{-size} \rangle \} \), i.e., the result fits within the dimensions specified. Negative sizes cause the material in the \( \langle \text{box} \rangle \) to be reversed in direction, but the reference point of the \( \langle \text{box} \rangle \) is unchanged. Thus a negative \( \langle y\text{-size} \rangle \) results in the \( \langle \text{box} \rangle \) having a depth dependent on the height of the original and vice versa.

Resizes the \( \langle \text{box} \rangle \) to \( \langle y\text{-size} \rangle \) (vertically), scaling the horizontal size by the same amount; \( \langle y\text{-size} \rangle \) is a dimension expression. The \( \langle y\text{-size} \rangle \) is the height only: it does not include any depth. The updated \( \langle \text{box} \rangle \) is an hbox, irrespective of the nature of the \( \langle \text{box} \rangle \) before the resizing is applied. A negative \( \langle y\text{-size} \rangle \) causes the material in the \( \langle \text{box} \rangle \) to be reversed in direction, but the reference point of the \( \langle \text{box} \rangle \) is unchanged. Thus a negative \( \langle y\text{-size} \rangle \) results in the \( \langle \text{box} \rangle \) having a depth dependent on the height of the original and vice versa.
Resizes the \( \langle \text{box} \rangle \) to \( \langle y\text{-size} \rangle \) (vertically), scaling the horizontal size by the same amount: \( \langle y\text{-size} \rangle \) is a dimension expression. The \( \langle y\text{-size} \rangle \) is the total vertical size (height plus depth). The updated \( \langle \text{box} \rangle \) is an \text{ hbox}, irrespective of the nature of the \( \langle \text{box} \rangle \) before the resizing is applied. A negative \( \langle y\text{-size} \rangle \) causes the material in the \( \langle \text{box} \rangle \) to be reversed in direction, but the reference point of the \( \langle \text{box} \rangle \) is unchanged. Thus a negative \( \langle y\text{-size} \rangle \) results in the \( \langle \text{box} \rangle \) having a depth dependent on the height of the original and \textit{vice versa}.

Resizes the \( \langle \text{box} \rangle \) to \( \langle x\text{-size} \rangle \) (horizontally), scaling the vertical size by the same amount: \( \langle x\text{-size} \rangle \) is a dimension expression. The updated \( \langle \text{box} \rangle \) is an \text{ hbox}, irrespective of the nature of the \( \langle \text{box} \rangle \) before the resizing is applied. A negative \( \langle x\text{-size} \rangle \) causes the material in the \( \langle \text{box} \rangle \) to be reversed in direction, but the reference point of the \( \langle \text{box} \rangle \) is unchanged. Thus a negative \( \langle x\text{-size} \rangle \) results in the \( \langle \text{box} \rangle \) having a depth dependent on the height of the original and \textit{vice versa}.

Resizes the \( \langle \text{box} \rangle \) to \( \langle x\text{-size} \rangle \) (horizontally) and \( \langle y\text{-size} \rangle \) (vertically): both of the sizes are dimension expressions. The \( \langle y\text{-size} \rangle \) is the height only and does not include any depth. The updated \( \langle \text{box} \rangle \) is an \text{ hbox}, irrespective of the nature of the \( \langle \text{box} \rangle \) before the resizing is applied. Negative sizes cause the material in the \( \langle \text{box} \rangle \) to be reversed in direction, but the reference point of the \( \langle \text{box} \rangle \) is unchanged. Thus a negative \( \langle y\text{-size} \rangle \) results in the \( \langle \text{box} \rangle \) having a depth dependent on the height of the original and \textit{vice versa}.

Resizes the \( \langle \text{box} \rangle \) to \( \langle x\text{-size} \rangle \) (horizontally) and \( \langle y\text{-size} \rangle \) (vertically): both of the sizes are dimension expressions. The \( \langle y\text{-size} \rangle \) is the total vertical size (height plus depth). The updated \( \langle \text{box} \rangle \) is an \text{ hbox}, irrespective of the nature of the \( \langle \text{box} \rangle \) before the resizing is applied. Negative sizes cause the material in the \( \langle \text{box} \rangle \) to be reversed in direction, but the reference point of the \( \langle \text{box} \rangle \) is unchanged. Thus a negative \( \langle y\text{-size} \rangle \) results in the \( \langle \text{box} \rangle \) having a depth dependent on the height of the original and \textit{vice versa}.
\box_rotate:Nn \box_rotate:cn \box_grotate:Nn \box_grotate:cn
Rotates the \textit{box} by \textit{angle} (in degrees) anti-clockwise about its reference point. The reference point of the updated box is moved horizontally such that it is at the left side of the smallest rectangle enclosing the rotated material. The updated \textit{box} is an \texttt{hbox}, irrespective of the nature of the \textit{box} before the rotation is applied.

\box_scale:Nnn \box_scale:cn \box_gscale:Nnn \box_gscale:cn
Scales the \textit{box} by factors \textit{x-scale} and \textit{y-scale} in the horizontal and vertical directions, respectively (both scales are integer expressions). The updated \textit{box} is an \texttt{hbox}, irrespective of the nature of the \textit{box} before the scaling is applied. Negative scalings cause the material in the \textit{box} to be reversed in direction, but the reference point of the \textit{box} is unchanged. Thus a negative \textit{y-scale} results in the \textit{box} having a depth dependent on the height of the original and \textit{vice versa}.

### 33.14 Primitive box conditionals

\if_hbox:N \if_vbox:N \if_box_empty:N
\if_hbox:N \texttt{true code}\else:\texttt{false code}\fi:
Tests is \textit{box} is a horizontal box.

\textbf{\TeXhackers note:} This is the \TeX primitive \texttt{ifhbox}.

\if_vbox:N \texttt{true code}\else:\texttt{false code}\fi:
Tests is \textit{box} is a vertical box.

\textbf{\TeXhackers note:} This is the \TeX primitive \texttt{ifvbox}.

\if_box_empty:N \texttt{true code}\else:\texttt{false code}\fi:
Tests is \textit{box} is an empty (void) box.

\textbf{\TeXhackers note:} This is the \TeX primitive \texttt{ifvoid}.
Chapter 34

The l3coffins package
Coffin code layer

The material in this module provides the low-level support system for coffins. For details about the design concept of a coffin, see the xcoffins module (in the l3experimental bundle).

34.1 Creating and initialising coffins

\coffin_new:N \coffin
New: 2011-08-17

Creates a new \texttt{coffin} or raises an error if the name is already taken. The declaration is global. The \texttt{coffin} is initially empty.

\coffin_clear:N \coffin_clear:c
\coffin_gclear:N \coffin_gclear:c
New: 2011-08-17
Updated: 2019-01-21

Clears the content of the \texttt{coffin}.

\coffin_set_eq:NN \coffin
\coffin_set_eq:\langle Nc|cN|cc\rangle
\coffin_gset_eq:NN
\coffin_gset_eq:\langle Nc|cN|cc\rangle
New: 2011-08-17
Updated: 2019-01-21

Sets both the content and poles of \texttt{coffin}_1 equal to those of \texttt{coffin}_2.

\coffin_if_exist_p:N \coffin
\coffin_if_exist_p:c *
\coffin_if_exist:NTF \coffin \{(true code)\} \{(false code)\}
\coffin_if_exist:ctF *
New: 2012-06-20

Tests whether the \texttt{coffin} is currently defined.
34.2 Setting coffin content and poles

\hcoffin_set:Nn \hcoffin_set:cn \hcoffin_gset:Nn \hcoffin_gset:cn

Typesets the \textit{material} in horizontal mode, storing the result in the \textit{coffin}. The standard poles for the \textit{coffin} are then set up based on the size of the typeset material.

\hcoffin_set:Nw \hcoffin_set:cw \hcoffin_gset:Nw \hcoffin_gset:cw

Typesets the \textit{material} in horizontal mode, storing the result in the \textit{coffin}. The standard poles for the \textit{coffin} are then set up based on the size of the typeset material. These functions are useful for setting the entire contents of an environment in a coffin.

\vcoffin_set:Nnn \vcoffin_set:cnn \vcoffin_gset:Nnn \vcoffin_gset:cnn

Typesets the \textit{material} in vertical mode constrained to the given \textit{width} and stores the result in the \textit{coffin}. The standard poles for the \textit{coffin} are then set up based on the size of the typeset material.

\vcoffin_set:Nnw \vcoffin_set:cnw \vcoffin_gset:Nnw \vcoffin_gset:cnw

Typesets the \textit{material} in vertical mode constrained to the given \textit{width} and stores the result in the \textit{coffin}. The standard poles for the \textit{coffin} are then set up based on the size of the typeset material. These functions are useful for setting the entire contents of an environment in a coffin.

\coffin_set_horizontal_pole:Nnn \coffin_set_horizontal_pole:cnn \coffin_gset_horizontal_pole:Nnn \coffin_gset_horizontal_pole:cnn

Sets the \textit{pole} to run horizontally through the \textit{coffin}. The \textit{pole} is placed at the \textit{offset} from the bottom edge of the bounding box of the \textit{coffin}. The \textit{offset} should be given as a dimension expression.
Sets the \textit{(pole)} to run vertically through the \textit{(coffin)}. The \textit{(pole)} is placed at the \textit{(offset)} from the left-hand edge of the bounding box of the \textit{(coffin)}. The \textit{(offset)} should be given as a dimension expression.

### 34.3 Coffin affine transformations

- \textbf{\texttt{\textbackslash coffin_set_vertical_pole:NNn}}
- \textbf{\texttt{\textbackslash coffin_set_vertical_pole:cnn}}

\texttt{\textbackslash coffin_set_vertical_pole:NNn \{coffin\} \{(pole)\} \{(offset)\}}

\texttt{\textbackslash coffin_set_vertical_pole:cnn}

\texttt{\textbackslash coffin_set_vertical_pole:NNn}

\texttt{\textbackslash coffin_set_vertical_pole:cnn}

\textbf{Updated: 2019-01-21}

Sets the \textit{(pole)} to run vertically through the \textit{(coffin)}. The \textit{(pole)} is placed at the \textit{(offset)} from the left-hand edge of the bounding box of the \textit{(coffin)}. The \textit{(offset)} should be given as a dimension expression.

\textbf{34.3 Coffin affine transformations}

\textbf{Updated: 2012-07-20}

\textbf{Updated: 2019-01-21}

- \textbf{\texttt{\textbackslash coffin_resize:NNn}}
- \textbf{\texttt{\textbackslash coffin_resize:cnn}}

\texttt{\textbackslash coffin_resize:NNn \{coffin\} \{\textit{width}\} \{\textit{total-height}\}}

\texttt{\textbackslash coffin_resize:cnn}

\texttt{\textbackslash coffin_resize:NNn}

\texttt{\textbackslash coffin_resize:cnn}

\textbf{Updated: 2019-01-23}

Resized the \textit{(coffin)} to \textit{(width)} and \textit{(total-height)}, both of which should be given as dimension expressions.

- \textbf{\texttt{\textbackslash coffin_rotate:NNn}}
- \textbf{\texttt{\textbackslash coffin_rotate:cn}}

\texttt{\textbackslash coffin_rotate:NNn \{coffin\} \{\textit{angle}\}}

\texttt{\textbackslash coffin_rotate:cn}

\texttt{\textbackslash coffin_rotate:NNn}

\texttt{\textbackslash coffin_rotate:cn}

\textbf{Updated: 2019-01-23}

Rotates the \textit{(coffin)} by the given \textit{(angle)} (given in degrees counter-clockwise). This process rotates both the coffin content and poles. Multiple rotations do not result in the bounding box of the coffin growing unnecessarily.

- \textbf{\texttt{\textbackslash coffin_scale:NNn}}
- \textbf{\texttt{\textbackslash coffin_scale:cnn}}

\texttt{\textbackslash coffin_scale:NNn \{coffin\} \{\textit{x-scale}\} \{\textit{y-scale}\}}

\texttt{\textbackslash coffin_scale:cnn}

\texttt{\textbackslash coffin_scale:NNn}

\texttt{\textbackslash coffin_scale:cnn}

\textbf{Updated: 2019-01-23}

Scales the \textit{(coffin)} by a factors \textit{(x-scale)} and \textit{(y-scale)} in the horizontal and vertical directions, respectively. The two scale factors should be given as real numbers.

### 34.4 Joining and using coffins

\textbf{Updated: 2019-01-22}

This function attaches \textit{(coffin2)} to \textit{(coffin1)} such that the bounding box of \textit{(coffin1)} is not altered, \textit{i.e.} \textit{(coffin2)} can protrude outside of the bounding box of the coffin. The alignment is carried out by first calculating \textit{(handle1)}, the point of intersection of \textit{(coffin1-pole1)} and \textit{(coffin1-pole2)}, and \textit{(handle2)}, the point of intersection of \textit{(coffin2-pole1)} and \textit{(coffin2-pole2)}. \textit{(coffin2)} is then attached to \textit{(coffin1)} such that the relationship between \textit{(handle1)} and \textit{(handle2)} is described by the \textit{(x-offset)} and \textit{(y-offset)}. The two offsets should be given as dimension expressions.
This function joins ⟨coffin1⟩ to ⟨coffin2⟩ such that the bounding box of ⟨coffin1⟩ may expand. The new bounding box covers the area containing the bounding boxes of the two original coffins. The alignment is carried out by first calculating ⟨handle1⟩, the point of intersection of ⟨coffin1-pole1⟩ and ⟨coffin1-pole2⟩, and ⟨handle2⟩, the point of intersection of ⟨coffin2-pole1⟩ and ⟨coffin2-pole2⟩. ⟨coffin2⟩ is then attached to ⟨coffin1⟩ such that the relationship between ⟨handle1⟩ and ⟨handle2⟩ is described by the ⟨x-offset⟩ and ⟨y-offset⟩. The two offsets should be given as dimension expressions.

Typesetting is carried out by first calculating ⟨handle⟩, the point of intersection of ⟨pole1⟩ and ⟨pole2⟩. The coffin is then typeset in horizontal mode such that the relationship between the current reference point in the document and the ⟨handle⟩ is described by the ⟨x-offset⟩ and ⟨y-offset⟩. The two offsets should be given as dimension expressions. Typesetting a coffin is therefore analogous to carrying out an alignment where the “parent” coffin is the current insertion point.

### 34.5 Measuring coffins

This function first calculates the intersections between all of the ⟨poles⟩ of the ⟨coffin⟩ to give a set of ⟨handles⟩. It then prints the ⟨coffin⟩ at the current location in the source, with the position of the ⟨handles⟩ marked on the coffin. The ⟨handles⟩ are labelled as part of this process: the locations of the ⟨handles⟩ and the labels are both printed in the ⟨color⟩ specified.
This function first calculates the \langle handle \rangle for the \langle coffin \rangle as defined by the intersection of \langle pole1 \rangle and \langle pole2 \rangle. It then marks the position of the \langle handle \rangle on the \langle coffin \rangle. The \langle handle \rangle are labelled as part of this process: the location of the \langle handle \rangle and the label are both printed in the \langle color \rangle specified.

This function shows the structural information about the \langle coffin \rangle in the terminal. The width, height and depth of the typeset material are given, along with the location of all of the poles of the coffin.

Notice that the poles of a coffin are defined by four values: the $x$ and $y$ co-ordinates of a point that the pole passes through and the $x$- and $y$-components of a vector denoting the direction of the pole. It is the ratio between the later, rather than the absolute values, which determines the direction of the pole.

This function writes the structural information about the \langle coffin \rangle in the log file. See also \texttt{\coffin\_show\_structure:N} which displays the result in the terminal.

Shows full details of poles and contents of the \langle coffin \rangle in the terminal or log file. See \texttt{\coffin\_show\_structure:N} and \texttt{\box\_show:N} to show separately the pole structure and the contents.

A permanently empty coffin.

Scratch coffins for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Scratch coffins for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX-3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Chapter 35

The l3color package

Color support

35.1 Color in boxes

Controlling the color of text in boxes requires a small number of control functions, so that the boxed material uses the color at the point where it is set, rather than where it is used.

\color_group_begin: ...
\color_group_end:

New: 2011-09-03

Creates a color group: one used to “trap” color settings. This grouping is built in to for example \hbox_set:Nn.

\color_ensure_current:

New: 2011-09-03

Ensures that material inside a box uses the foreground color at the point where the box is set, rather than that in force when the box is used. This function should usually be used within a \color_group_begin: ... \color_group_end: group.

35.2 Color models

A color model is a way to represent sets of colors. Different models are particularly suitable for different output methods, e.g. screen or print. Parameter-based models can describe a very large number of unique colors, and have a varying number of axes which define a color space. In contrast, various proprietary models are available which define spot colors (more formally separations).

Core models are used to pass color information to output; these are “native” to l3color. Core models use real numbers in the range [0, 1] to represent values. The core models supported here are

- gray Grayscale color, with a single axis running from 0 (fully black) to 1 (fully white)
- rgb Red-green-blue color, with three axes, one for each of the components
• cmyk Cyan-magenta-yellow-black color, with four axes, one for each of the components.

There are also interface models: these are convenient for users but have to be manipulated before storing/passing to the backend. Interface models are primarily integer-based: see below for more detail. The supported interface models are

• Gray Grayscale color, with a single axis running from 0 (fully black) to 15 (fully white)
• hsb Hue-saturation-brightness color, with three axes, all real values in the range [0, 1] for hue saturation and brightness
• Hsb Hue-saturation-brightness color, with three axes, integer in the range [0, 360] for hue, real values in the range [0, 1] for saturation and brightness
• HSB Hue-saturation-brightness color, with three axes, integers in the range [0, 240] for hue, saturation and brightness
• HTML HTML format representation of RGB color given as a single six-digit hexadecimal number
• RGB Red-green-blue color, with three axes, one for each of the components, values as integers from 0 to 255
• wave Light wavelength, a real number in the range 380 to 780 (nanometres)

All interface models are internally stored as rgb.

To allow parsing of data from xcolor, any leading model up the first : will be discarded; the approach of selecting an internal form for data is not used in l3color.

Additional models may be created to allow mixing of separation colors with each other or with those from other models. See Section 35.9 for more detail of color support for additional models.

When color is selected by model, the ⟨values⟩ given are specified as a comma-separated list. The length of the list will therefore be determined by the detail of the model involved.

Color models (and interconversion) are complex, and more details are given in the manual to the LATEX2ε xcolor package and in the PostScript Language Reference Manual, published by Addison–Wesley.

### 35.3 Color expressions

In addition to allowing specification of color by model and values, l3color also supports color expressions. These are created by combining one or more color names, with the amount of each specified as a percentage. The latter is given between ! symbols in the expression. Thus for example

red!50!green

is a mixture of 50 % red and 50 % green. A trailing percentage is interpreted as implicitly followed by white, and so

red!25
specifies 25% red mixed with 75% white.

Where the models for the mixed colors are different, the model of the first color is used. Thus

\texttt{red!50!cyan}

will result in a color specification using the \texttt{rgb} model, made up of 50% red and 50% of cyan \textit{expressed in rgb}. This may be important as color model interconversion is not exact.

The one exception to the above is where the first model in an expression is \texttt{gray}. In this case, the order of mixing is “swapped” internally, so that for example

\texttt{black!50!red}

has the same result as

\texttt{red!50!black}

(the predefined colors \texttt{black} and \texttt{white} use the \texttt{gray} model).

Where more than two colors are mixed in an expression, evaluation takes place in a stepwise fashion. Thus in

\texttt{cyan!50!magenta!10!yellow}

the sub-expression

\texttt{cyan!50!magenta}

is first evaluated to give an intermediate color specification, before the second step

\texttt{<intermediate>!10!yellow}

where \texttt{<intermediate>} represents this transitory calculated value.

Within a color expression, . may be used to represent the color active for typesetting (the current color). This allows for example

\texttt{.!50}

to mean a mixture of 50% of current color with white.

(Color expressions supported here are a subset of those provided by the \texttt{LATEX2ε xcolor} package. At present, only such features as are clearly useful have been added here.)

### 35.4 Named colors

Color names are stored in a single namespace, which makes them accessible as part of color expressions. Whilst they are not reserved in a technical sense, the names \texttt{black}, \texttt{white}, \texttt{red}, \texttt{green}, \texttt{blue}, \texttt{cyan}, \texttt{magenta} and \texttt{yellow} have special meaning and should not be redefined. Color names should be made up of letters, numbers and spaces only: other characters are reserved for use in color expressions. In particular, . represents the current color at the start of a color expression.

\begin{verbatim}
\color_set:nn \color_set:nn \{(name)\} \{(color expression)\}
\end{verbatim}

Evaluates the \textit{(color expression)} and stores the resulting color specification as the \textit{(name)}.
\color_set:nn \color_set:nn \{name\} \{model(s)\} \{value(s)\}
Stores the color specification equivalent to the \langle model(s) \rangle and \langle values \rangle as the \langle name \rangle.
\color_set_eq:nn \color_set_eq:nn \{name1\} \{name2\}
Copies the color specification in \langle name2 \rangle to \langle name1 \rangle. The special name . may be used to represent the current color, allowing it to be saved to a name.
\color_show:n \color_log:n \{name\}
Displays the color specification stored in the \langle name \rangle on the terminal or log file.

35.5 Selecting colors

General selection of color is safe when split across pages: a stack is used to ensure that the correct color is re-selected on the new page.
\color_select:n \color_select:n \{color expression\}
Parses the \langle color expression \rangle and then activates the resulting color specification for typeset material.
\color_select:nn \color_select:nn \{model(s)\} \{value(s)\}
Activates the color specification equivalent to the \langle model(s) \rangle and \langle value(s) \rangle for typeset material.
\l_color_fixed_model_tl
When this is set to a non-empty value, colors will be converted to the specified model when they are selected. Note that included images and similar are not influenced by this setting.

35.6 Colors for fills and strokes

Colors for drawing operations and so forth are split into strokes and fills (the latter may also be referred to as non-stroke color). The fill color is used for text under normal circumstances. Depending on the backend, stroke color may use a stack, in which case it exhibits the same page breaking behavior as general color. However, dvips/dvisvgm do not support this, and so color will need to be contained within a scope, such as \draw_begin:/\draw_end:.
Note that the current color is the fill color, as this is used for running text.
\color_fill:n \color_fill:n \{color expression\}
Parses the \langle color expression \rangle and then activates the resulting color specification for filling or stroking.
\color_fill:nn \color_fill:nn \{model(s)\} \{value(s)\}
Activates the color specification equivalent to the \langle model(s) \rangle and \langle value(s) \rangle for filling or stroking.
35.7 Multiple color models

When selecting or setting a color with an explicit model, it is possible to give values for more than one model at one time. This is particularly useful where automated conversion between models does not give the desired outcome. To do this, the list of models and list of values are both subdivided using / characters (as for the similar function in xcolor). For example, to save a color with explicit cmyk and rgb values, one could use:

```
\color_set:nnn { foo } { cmyk / rgb }
{ 0.1 , 0.2 , 0.3 , 0.4 / 0.1, 0.2 , 0.3 }
```

The manually-specified conversion will be used in preference to automated calculation whenever the model(s) listed are used: both in expressions and when a fixed model is active.

Similarly, the same syntax can be applied to directly selecting a color:

```
\color_select:nn { cmyk / rgb }
{ 0.1 , 0.2 , 0.3 , 0.4 / 0.1, 0.2 , 0.3 }
```

Again, this list is used when a fixed model is active: the first entry is used unless there is a fixed model matching one of the other entries.

35.8 Exporting color specifications

The major use of color expressions is in setting typesetting output, but there are other places in which some form of color information is required. These may need data in a different format or using a different model to the internal representation. Thus a set of functions are available to export colors in different formats.

Valid export targets are:

- **backend** Two brace groups: the first containing the model, the second containing space-separated values appropriate for the model; this is the format required by backend functions of expl3
- **comma-sep-cmyk** Comma-separated cyan-magenta-yellow-black values
- **comma-sep-rgb** Comma-separated red-green-blue values suitable for use as a PDF annotation color
- **HTML** Uppercase two-digit hexadecimal values, expressing a red-green-blue color; the digits are not separated
- **space-sep-cmyk** Space-separated cyan-magenta-yellow-black values
- **space-sep-rgb** Space-separated red-green-blue values suitable for use as a PDF annotation color
Parses the \langle color expression \rangle as described earlier, then converts to the \langle format \rangle specified and assigns the data to the \langle tl \rangle.

Expresses the combination of \langle model \rangle and \langle value(s) \rangle in an internal representation, then converts to the \langle format \rangle specified and assigns the data to the \langle tl \rangle.

### 35.9 Creating new color models

Additional color models are required to support specialist workflows, for example those involving separations (see [https://helpx.adobe.com/indesign/using/spot-process-colors.html](https://helpx.adobe.com/indesign/using/spot-process-colors.html) for details of the use of separations in print). Color models may be split into families; for the standard device-based color models (DeviceCMYK, DeviceRGB, DeviceGray), these are synonymous. This is not generally the case: see the PDF reference for more details. (Note that 13color uses the shorter names cmyk, etc.)

Creates a new \langle model \rangle which is derived from the color model \langle family \rangle. The latter should be one of

- DeviceN
- ICCBased
- Separation

(The \langle family \rangle may be given in mixed case as-in the PDF reference: internally, case of these strings is folded.) Depending on the \langle family \rangle, one or more \langle params \rangle are mandatory or optional.

For a Separation space, there are three compulsory keys.

- name The name of the Separation, for example the formal name of a spot color ink. Such a \langle name \rangle may contain spaces, etc., which are not permitted in the \langle model \rangle.

- alternative-model An alternative device colorspace, one of cmyk, rgb, gray or CIELAB. The three parameter-based models work as described above; see below for details of CIELAB colors.

- alternative-values A comma-separated list of values appropriate to the alternative-model. This information is used by the PDF application if the Separation is not available.

CIELAB color separations are created using the alternative-model = CIELAB setting. These colors must also have an illuminant key, one of a, c, e, d50, d55, d65 or d75. The alternative-values in this case are the three parameters \(L^*\), \(a^*\) and \(b^*\) of the CIELAB model. Full details of this device-independent color approach are given in the documentation to the colorspace package.

CIELAB colors cannot be converted into other device-dependent color spaces, and as such, mixing can only occur if colors set up using the CIELAB model are also given
with an alternative parameter-based model. If that is not the case, 13color will fallback to using black as the colorant in any mixing.

For a DeviceN space, there is one compulsory key.

- **names** The names of the components of the DeviceN space. Each should be either the *name* of a Separation model, a process color name (cyan, etc.) or the special name none.

  For a ICCBased space, there is one compulsory key.

- **file** The name of the file containing the profile.

### 35.9.1 Color profiles

Color profiles are used to ensure color accuracy by linking to collaboration. Applying a profile can be used to standardise color which is otherwise device-dependence.

\color_profile_apply:nn

\color_profile_apply:nn \{⟨profile⟩\} \{⟨model⟩\}

This function applies a *profile* to one of the device *models*. The profile will then apply to all color of the selected *model*. The *profile* should specify an ICC profile file. The *model* has to be one the standard device models: cmyk, gray or rgb.
Chapter 36

The l3pdf package
Core PDF support

36.1 Objects

\pdf_object_new:nn
\pdf_object_new:nn {⟨object⟩} {⟨type⟩}
Declares ⟨object⟩ as a PDF object of ⟨type⟩, which should be one of

- array
- dict
- fstream
- stream

The object may be referenced from this point on, and written later using \pdf_object_write:nn.

\pdf_object_if_exist_p:n
\pdf_object_if_exist:nTF
Tests whether an object with name {⟨object⟩} has been defined.

\pdf_object_write:nn
\pdf_object_write:nx
Writes the ⟨content⟩ as content of the ⟨object⟩. Depending on the ⟨type⟩ declared for the object, the format required for the ⟨data⟩ will vary

array A space-separated list of values

dict Key–value pairs in the form /⟨key⟩ ⟨value⟩

fstream Two brace groups: ⟨file name⟩ and ⟨file content⟩

stream Two brace groups: ⟨attributes (dictionary)⟩ and ⟨stream contents⟩
\pdf_object_ref:n * \pdf_object_ref:n \{object\}

Inserts the appropriate information to reference the \{object\} in for example page resource allocation.

\pdf_object_unnamed_write:nn \pdf_object_unnamed_write:nn \{type\} \{content\}
\pdf_object_unnamed_write:nx

Writes the \{content\} as content of an anonymous object. Depending on the \{type\}, the format required for the \{data\} will vary:

- **array** A space-separated list of values
- **dict** Key–value pairs in the form \{key\} \{value\}
- **fstream** Two brace groups: \{attributes (dictionary)\} and \{file name\}
- **stream** Two brace groups: \{attributes (dictionary)\} and \{stream contents\}

\pdf_object_ref_last: *

Inserts the appropriate information to reference the last \{object\} created. This is particularly useful for anonymous objects.

\pdf_pageobject_ref:n *

Inserts the appropriate information to reference the \{pageobject\}.

### 36.2 Version

\pdf_version_compare_p:Nn * \pdf_version_compare:NnTF \{comparator\} \{\{version\}\} \{\{true code\}\} \{\{false code\}\}

Compares the version of the PDF being created with the \{version\} string specified, using the \{comparator\}. Either the \{true code\} or \{false code\} will be left in the output stream.

\pdf_version_gset:n
\pdf_version_min_gset:n

Sets the \{version\} of the PDF being created. The \texttt{min} version will not alter the output version unless it is currently lower than the \{version\} requested.

This function may only be used up to the point where the PDF file is initialised. With \texttt{dvips} it sets \pdf_version_major: and \pdf_version_minor: and allows to compare the values with \pdf_version_compare:Nn, but the PDF version itself still has to be set with the command line option \texttt{-dCompatibilityLevel} of \texttt{ps2pdf}.

\pdf_version:
\pdf_version_major:
\pdf_version_minor:

Expands to the currently-active PDF version.
36.3 Compression

\pdf_uncompress:

Disables any compression of the PDF, where possible.

This function may only be used up to the point where the PDF file is initialised.

36.4 Destinations

Destinations are the places a link jumped too. Unlike the name may suggest they don’t described an exact location in the PDF. Instead a destination contains a reference to a page along with an instruction how to display this page. The normally used “XYZ top left zoom” for example instructs the viewer to show the page with the given zoom and the top left corner at the top left coordinates—which then gives the impression that there is an anchor at this position.

If an instruction takes a coordinate, it is calculated by the following commands relative to the location the command is issued. So to get a specific coordinate one has to move the command to the right place.

\pdf_destination:nn:

This creates a destination. \{(type or integer)\} can be one of fit, fith, fitv, fitb, fitbh, fitbv, fitr, xyz or an integer representing a scale factor in percent. fitr here gives only a lightweight version of /FitR: The backend code defines fitr so that it will with pdfL\TeX{} and LuaL\TeX{} use the coordinates of the surrounding box, with dvips and dvipdfmx it falls back to fit. For full control use \pdf_destination:nnnn.

The keywords match to the PDF names as described in the following tabular.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>PDF</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>fit</td>
<td>/Fit</td>
<td>Fits the page to the window</td>
</tr>
<tr>
<td>fith</td>
<td>/FitH</td>
<td>Fits the width of the page to the window</td>
</tr>
<tr>
<td>fitv</td>
<td>/FitV</td>
<td>Fits the height of the page to the window</td>
</tr>
<tr>
<td>fitb</td>
<td>/FitB</td>
<td>Fits the page bounding box to the window</td>
</tr>
<tr>
<td>fitbh</td>
<td>/FitBH</td>
<td>Fits the width of the page bounding box to the window.</td>
</tr>
<tr>
<td>fitbv</td>
<td>/FitBV</td>
<td>Fits the height of the page bounding box to the window.</td>
</tr>
<tr>
<td>fitr</td>
<td>/FitR</td>
<td>Fits the rectangle specified by the four coordinates to the window (see above for the restrictions)</td>
</tr>
<tr>
<td>xyz</td>
<td>/XYZ</td>
<td>Sets a coordinate but doesn’t change the zoom.</td>
</tr>
<tr>
<td>{(integer)}</td>
<td>/XYZ</td>
<td>Sets a coordinate and a zoom meaning {(integer)}%.</td>
</tr>
</tbody>
</table>
This creates a destination with /FitR type with the given dimensions relative to the current location. The destination is in a box of size zero, but it doesn’t switch to horizontal mode.
Part VII
Additions and removals
Chapter 37

The \texttt{l3candidates} package

Experimental additions to \texttt{l3kernel}

37.1 Important notice

This module provides a space in which functions can be added to \texttt{l3kernel} (\texttt{expl3}) while still being experimental.

As such, the functions here may not remain in their current form, or indeed at all, in \texttt{l3kernel} in the future.

In contrast to the material in \texttt{l3experimental}, the functions here are all \textit{small} additions to the kernel. We encourage programmers to test them out and report back on the \LaTeX-L mailing list.

Thus, if you intend to use any of these functions from the candidate module in a public package offered to others for productive use (e.g., being placed on CTAN) please consider the following points carefully:

- Be prepared that your public packages might require updating when such functions are being finalized.
- Consider informing us that you use a particular function in your public package, e.g., by discussing this on the \LaTeX-L mailing list. This way it becomes easier to coordinate any updates necessary without issues for the users of your package.
- Discussing and understanding use cases for a particular addition or concept also helps to ensure that we provide the right interfaces in the final version so please give us feedback if you consider a certain candidate function useful (or not).

We only add functions in this space if we consider them being serious candidates for a final inclusion into the kernel. However, real use sometimes leads to better ideas, so functions from this module are \textbf{not necessarily stable} and we may have to adjust them!
37.2 Additions to \texttt{l3box}

\begin{verbatim}
\box_clip:N \box_clip:c \box_gclip:N \box_gclip:c
Updated: 2019-01-23
\end{verbatim}

Clips the \langle box \rangle in the output so that only material inside the bounding box is displayed in the output. The updated \langle box \rangle is an hbox, irrespective of the nature of the \langle box \rangle before the clipping is applied.

These functions require the \LaTeX3 native drivers: they do not work with the \LaTeX2e graphics drivers!

\textbf{\LaTeXXhackers note:} Clipping is implemented by the driver, and as such the full content of the box is placed in the output file. Thus clipping does not remove any information from the raw output, and hidden material can therefore be viewed by direct examination of the file.

\begin{verbatim}
\box_set_trim:Nnnnn \box_set_trim:cnnnn \box_gset_trim:Nnnnn \box_gset_trim:cnnnn
New: 2019-01-23
\end{verbatim}

Adjusts the bounding box of the \langle box \rangle \langle left \rangle is removed from the left-hand edge of the bounding box, \langle right \rangle from the right-hand edge and so forth. All adjustments are \langle dimension expressions \rangle. Material outside of the bounding box is still displayed in the output unless \texttt{\box_clip:N} is subsequently applied. The updated \langle box \rangle is an hbox, irrespective of the nature of the \langle box \rangle before the trim operation is applied. The behavior of the operation where the trims requested is greater than the size of the box is undefined.

\begin{verbatim}
\box_set_viewport:Nnnnn \box_set_viewport:cnnnn \box_gset_viewport:Nnnnn \box_gset_viewport:cnnnn
New: 2019-01-23
\end{verbatim}

Adjusts the bounding box of the \langle box \rangle such that it has lower-left co-ordinates (\langle llx \rangle, \langle lly \rangle) and upper-right co-ordinates (\langle urx \rangle, \langle ury \rangle). All four co-ordinate positions are \langle dimension expressions \rangle. Material outside of the bounding box is still displayed in the output unless \texttt{\box_clip:N} is subsequently applied. The updated \langle box \rangle is an hbox, irrespective of the nature of the \langle box \rangle before the viewport operation is applied.

37.3 Additions to \texttt{l3expan}

\begin{verbatim}
\exp_args_generate:n
New: 2018-04-04
Updated: 2019-02-08
\end{verbatim}

Defines \texttt{\exp_args:N(variant)} functions for each \langle variant \rangle given in the comma list \langle variant argument specifiers \rangle. Each \langle variant \rangle should consist of the letters N, c, n, V, v, o, f, e, x, p and the resulting function is protected if the letter x appears in the \langle variant \rangle. This is only useful for cases where \texttt{\cs_generate_variant:Nn} is not applicable.

\begin{verbatim}
\fp_if_nan:n \fp_if_nan:n \fp_nan:n \fp_if_nan:p:n \fp_if_nan:TF
New: 2019-08-25
\end{verbatim}

Evaluates the \langle fexpr \rangle and tests whether the result is exactly \texttt{NaN}. The test returns \texttt{false} for any other result, even a tuple containing \texttt{NaN}.
37.5 Additions to l3file

\iow_allow_break:

In the first argument of \iow_wrap:nn (for instance in messages), inserts a break-point that allows a line break. In other words this is a zero-width breaking space.

\ior_get_term:nN \ior_str_get_term:nN

New: 2019-03-23

Function that reads one or more lines (until an equal number of left and right braces are found) from the terminal and stores the result locally in the \langle token list \rangle variable. Tokenization occurs as described for \ior_get:NN or \ior_str_get:NN, respectively. When the \langle prompt \rangle is empty, \TeX will wait for input without any other indication: typically the programmer will have provided a suitable text using e.g. \ior_term:n.

Where the \langle prompt \rangle is given, it will appear in the terminal followed by an =, e.g.

\ior_shell_open:Nn

New: 2019-05-08

Opens the pseudo-file created by the output of the \langle shell command \rangle for reading using \langle stream \rangle as the control sequence for access. If the \langle stream \rangle was already open it is closed before the new operation begins. The \langle stream \rangle is available for access immediately and will remain allocated to \langle shell command \rangle until a \ior_close:N instruction is given or the \TeX run ends. If piped system calls are disabled an error is raised.

For details of handling of the \langle shell command \rangle, see \sys_get_shell:nnNTF.

37.6 Additions to l3flag

\flag_raise_if_clear:n *

New: 2018-04-02

Ensures the \langle flag \rangle is raised by making its height at least 1, locally.

37.7 Additions to l3intarray

\intarray_gset_rand:Nnn \intarray_gset_rand:cn \intarray_gset_rand:Nn \intarray_gset_rand:cn

New: 2018-05-05

Evaluates the integer expressions \langle minimum \rangle and \langle maximum \rangle then sets each entry (independently) of the \langle integer array variable \rangle to a pseudo-random number between the two (with bounds included). If the absolute value of either bound is bigger than $2^{30} - 1$, an error occurs. Entries are generated in the same way as repeated calls to \int_rand:nn or \int_rand:n respectively, in particular for the second function the \langle minimum \rangle is 1. Assignments are always global. This is not available in older versions of X\TeX.

\intarray_to_clist:N ⊆

New: 2018-05-04

Converts the \langle intarray \rangle to integer denotations separated by commas. All tokens have category code other. If the \langle intarray \rangle has no entry the result is empty; otherwise the result has one fewer comma than the number of items.
### 37.8 Additions to \texttt{l3msg}

\begin{verbatim}
\msg_show_eval:Nn \msg_log_eval:Nn
\begin{verbatim}
\msg_show_eval:Nn \texttt{(function)} \texttt{\{expression\}}
\end{verbatim}
\end{verbatim}

Shows or logs the \texttt{\{expression\}} (turned into a string), an equal sign, and the result of applying the \texttt{\{function\}} to the \texttt{\{expression\}} (with f-expansion). For instance, if the \texttt{\{function\}} is \texttt{\int_eval:n} and the \texttt{\{expression\}} is \texttt{1+2} then this logs \texttt{> 1+2=3}.

\begin{verbatim}
\msg_show_item:n \msg_show_item_unbraced:n \msg_show_item:n \msg_show_item_unbraced:nn
\end{verbatim}

Used in the text of messages for \texttt{\msg_show:nnxxxx} to show or log a list of items or key–value pairs. The one-argument functions are used for sequences, clist or token lists and the others for property lists. These functions turn their arguments to strings.

### 37.9 Additions to \texttt{l3prg}

\begin{verbatim}
\bool_set_inverse:N \bool_set_inverse:c
\bool_set_inverse:Nn \bool_set_inverse:cn
\end{verbatim}

\begin{verbatim}
\bool_set_inverse:N \texttt{(boolean)}
\end{verbatim}

Toggles the \texttt{\{boolean\}} from \texttt{true} to \texttt{false} and conversely: sets it to the inverse of its current value.
\bool_case_true:n * \bool_case_true:nTF
\bool_case_true:nTF * {
  \langle boolean expression cases \rangle
  \langle code case 1 \rangle
  \langle boolean expression cases \rangle
  \langle code case 2 \rangle
  \ldots
  \langle boolean expression cases \rangle
  \langle code case n \rangle
}
\langle true code \rangle
\langle false code \rangle

Evaluates in turn each of the \langle boolean expression cases \rangle until the first one that evaluates to true or to false, for \bool_case_true:n and \bool_case_false:n, respectively. The \langle code \rangle associated to this first case is left in the input stream, followed by the \langle true code \rangle, and other cases are discarded. If none of the cases match then only the \langle false code \rangle is inserted. The functions \bool_case_true:n and \bool_case_false:n, which do nothing if there is no match, are also available. For example

\bool_case_true:nF
{
  \langle\dim_compare_p:n \l__mypkg_wd_dim <= 10pt \rangle
  \langle Fits \rangle
  \langle\int_compare_p:n \l__mypkg_total_int >= 10 \rangle
  \langle Many \rangle
  \langle\l__mypkg_special_bool \rangle
  \langle Special \rangle
}
\langle No idea! \rangle

leaves “Fits” or “Many” or “Special” or “No idea!” in the input stream, in a way similar to some other language’s \texttt{if ... elseif ... elseif ... else ...}.

### 37.10 Additions to \texttt{l3prop}

\prop_rand_key_value:N *
\prop_rand_key_value:c *

\prop_rand_key_value:N \langle prop var \rangle

Selects a pseudo-random key–value pair from the \langle property list \rangle and returns \{\langle key \rangle\} and \{\langle value \rangle\}. If the \langle property list \rangle is empty the result is empty. This is not available in older versions of \texttt{Xe\LaTeX}.

\TeXXhackers note: The result is returned within the \texttt{unexpanded} primitive \texttt{(exp_not:n)}, which means that the \langle value \rangle does not expand further when appearing in an \texttt{x}-type argument expansion.
### 37.11 Additions to \texttt{l3seq}

\begin{verbatim}
\seq_mapthread_function:NNN    \seq_mapthread_function:NNN (seq_1) (seq_2) (function)
\seq_mapthread_function:NcNcN  \seq_mapthread_function:NNN (NcN|cNN|ccN) \tag
\end{verbatim}

Applies \texttt{(function)} to every pair of items \texttt{(seq\_item1)}–\texttt{(seq\_item2)} from the two sequences, returning items from both sequences from left to right. The \texttt{(function)} receives two \texttt{n-type} arguments for each iteration. The mapping terminates when the end of either sequence is reached (\textit{i.e.} whichever sequence has fewer items determines how many iterations occur).

\begin{verbatim}
\seq_set_filter:NNn \seq_set_filter:NnN (sequence_1) (sequence_2) \{\textit{inline boolexpr}\}
\seq_gset_filter:NNn \seq_gset_filter:NnN (sequence_1) (sequence_2) \{\textit{inline boolexpr}\}
\end{verbatim}

Evaluates the \texttt{(inline boolexpr)} for every \texttt{(item)} stored within the \texttt{(sequence_2)}. The \texttt{(inline boolexpr)} receives the \texttt{(item)} as \texttt{#1}. The sequence of all \texttt{(items)} for which the \texttt{(inline boolexpr)} evaluated to \texttt{true} is assigned to \texttt{(sequence_1)}.

\textbf{\texttt{TeXhackers} note:} Contrarily to other mapping functions, \texttt{\seq_map_break:} cannot be used in this function, and would lead to low-level \texttt{TeX} errors.

\begin{verbatim}
\seq_set_from_function:NnN \seq_set_from_function:NNN (seq var \{\textit{loop code}\}) \{\textit{function}\}
\seq_gset_from_function:NnN \seq_gset_from_function:NNN (seq var \{\textit{loop code}\}) \{\textit{function}\}
\end{verbatim}

Sets the \texttt{(seq var)} equal to a sequence whose items are obtained by \texttt{x}-expanding \texttt{(loop code)} \texttt{(function)}. This expansion must result in successive calls to the \texttt{(function)} with no nonexpandable tokens in between. More precisely the \texttt{(function)} is replaced by a wrapper function that inserts the appropriate separators between items in the sequence. The \texttt{(loop code)} must be expandable; it can be for example \texttt{\tl_map_function:NN (tl var)} or \texttt{\clist_map_function:nN \{\textit{clist}\}} or \texttt{\int_step_function:nnnN \{\textit{initial value}\} \{\textit{step}\} \{\textit{final value}\}}.

\begin{verbatim}
\seq_set_from_inline_x:Nnn \seq_set_from_inline_x:NNN (seq var \{\textit{loop code}\}) \{\textit{inline code}\}
\seq_gset_from_inline_x:Nnn \seq_gset_from_inline_x:NNN (seq var \{\textit{loop code}\}) \{\textit{inline code}\}
\end{verbatim}

Sets the \texttt{(seq var)} equal to a sequence whose items are obtained by \texttt{x}-expanding \texttt{(loop code)} applied to a \texttt{(function)} derived from the \texttt{(inline code)}. A \texttt{(function)} is defined, that takes one argument, \texttt{x}-expands the \texttt{(inline code)} with that argument as \texttt{#1}, then adds appropriate separators to turn the result into an item of the sequence. The \texttt{x}-expansion of \texttt{(loop code)} \texttt{(function)} must result in successive calls to the \texttt{(function)} with no nonexpandable tokens in between. The \texttt{(loop code)} must be expandable; it can be for example \texttt{\tl_map_function:NN (tl var)} or \texttt{\clist_map_function:nN \{\textit{clist}\}} or \texttt{\int_step_function:nnnN \{\textit{initial value}\} \{\textit{step}\} \{\textit{final value}\}}, but not the analogous “inline” mappings.
Removes the item of \textit{sequence} at the position given by evaluating the \textit{integer expression} and replaces it by \textit{item}. Items are indexed from 1 on the left/top of the \textit{sequence}, or from −1 on the right/bottom. If the \textit{integer expression} is zero or is larger (in absolute value) than the number of items in the sequence, the \textit{sequence} is not modified. In these cases, \texttt{\seq_set_item:Nnn} raises an error while \texttt{\seq_set_item:NnnTF} runs the \textit{false code}. In cases where the assignment was successful, \textit{true code} is run afterwards.

Removes the \textit{item} at position \textit{integer expression} in the \textit{sequence}, and places it in the \textit{token list variable}. Items are indexed from 1 on the left/top of the \textit{sequence}, or from −1 on the right/bottom. If the position is zero or is larger (in absolute value) than the number of items in the sequence, the \textit{seq var} is not modified, the \textit{token list} is set to the special marker \texttt{\q_no_value}, and the \textit{false code} is left in the input stream; otherwise the \textit{true code} is. The \textit{token list} assignment is local while the \textit{sequence} is assigned locally for \texttt{pop} or globally for \texttt{gpop} functions.

### 37.12 Additions to \texttt{l3sys}

The version string of the current engine, in the same form as given in the banner issued when running a job. For pdf\TeX\ and Lua\TeX\ this is of the form

\begin{verbatim}
\langle major\rangle.\langle minor\rangle.\langle revision\rangle
\end{verbatim}

For X\TeX, the form is

\begin{verbatim}
\langle major\rangle.\langle minor\rangle
\end{verbatim}

For \p\TeX\ and up\TeX, only releases since \TeX\ Live 2018 make the data available, and the form is more complex, as it comprises the \p\TeX\ version, the up\TeX\ version and the e-\p\TeX\ version.

\begin{verbatim}
p\langle major\rangle.\langle minor\rangle.\langle revision\rangle-u\langle major\rangle.\langle minor\rangle-e\langle p\TeX\rangle
\end{verbatim}

where the \texttt{u} part is only present for up\TeX.

Tests if the engine has a pseudo-random number generator. Currently this is the case in pdf\TeX, Lua\TeX, \p\TeX, up\TeX\ and recent releases of X\TeX.
37.13 Additions to l3tl

Leaves in the input stream the items from the \langle start index \rangle to the \langle end index \rangle inclusive, using the same indexing as \tl_range:nnn. Spaces are ignored. Regardless of whether items appear with or without braces in the \langle token list \rangle, the \tl_range_braced:nnn function wraps each item in braces, while \tl_range_unbraced:nnn does not (overall it removes an outer set of braces). For instance,

\begin{verbatim}
\iow_term:x { \tl_range_braced:nnn { abcd~{e{}}f } { 2 } { 5 } }
\iow_term:x { \tl_range_braced:nnn { abcd-{e{}}f } { -4 } { -1 } }
\iow_term:x { \tl_range_braced:nnn { abcd-{e()}f } { -2 } { -1 } }
\iow_term:x { \tl_range_braced:nnn { abcd-{e}f } { 0 } { -1 } }
\end{verbatim}

prints \{b\}{c}{d}{e{}}, \{c\}{d}{e{}}{f}, \{e{}\}{f}, and an empty line to the terminal, while

\begin{verbatim}
\iow_term:x { \tl_range_unbraced:nnn { abcd-{e{}}f } { 2 } { 5 } }
\iow_term:x { \tl_range_unbraced:nnn { abcd-{e()}f } { -4 } { -1 } }
\iow_term:x { \tl_range_unbraced:nnn { abcd-{e}f } { -2 } { -1 } }
\iow_term:x { \tl_range_unbraced:nnn { abcd-{e}f } { 0 } { -1 } }
\end{verbatim}

prints bcde{}, cde{}f, e{}f, and an empty line to the terminal. Because braces are removed, the result of \tl_range_unbraced:nnn may have a different number of items as for \tl_range:nnn or \tl_range_braced:nnn. In cases where preserving spaces is important, consider the slower function \tl_range:nnn.

\textbf{TeXhackers note:} The result is returned within the \unexpanded primitive (\exp_not:n), which means that the \langle item \rangle does not expand further when appearing in an x-type argument expansion.

\begin{verbatim}
\tl_build_begin:N \tl_build_gbegin:N \tl_build_clear:N \tl_build_gclear:N
\end{verbatim}

Clears the \langle tl var \rangle and sets it up to support other \tl_build_... functions, which allow accumulating large numbers of tokens piece by piece much more efficiently than standard l3tl functions. Until \tl_build_end:N (\tl var) is called, applying any function from l3tl other than \tl_build_... will lead to incorrect results. The begin and gbegin functions must be used for local and global (tl var) respectively.

Clears the (tl var) and sets it up to support other \tl_build_... functions. The clear and gclear functions must be used for local and global (tl var) respectively.
\tl_build_put_left:Nn
\tl_build_put_right:Nn
\tl_build_gput_left:Nn
\tl_build_gput_right:Nn

\tl_build_put_left:Nx
\tl_build_put_right:Nx
\tl_build_gput_left:Nx
\tl_build_gput_right:Nx

\tl_build_get:N
\tl_build_gget:N

\tl_build_end:N
\tl_build_gend:N

\c_catcode_active_space_tl
\char_to_utfviii_bytes:n
\char_to_nfd:N

\tl_build_put_left:Nn \{tl var\} \{\tokens\}
\tl_build_put_right:Nn \{tl var\} \{\tokens\}

Adds \{\tokens\} to the left or right side of the current contents of \{tl var\}. The \{tl var\} must have been set up with \tl_build_begin:N or \tl_build_gbegin:N. The put and gput functions must be used for local and global \{tl var\} respectively. The right functions are about twice faster than the left functions.

\tl_build_get:N \{tl var1\} \{tl var2\}

Stores the contents of the \{tl var1\} in the \{tl var2\}. The \{tl var1\} must have been set up with \tl_build_begin:N or \tl_build_gbegin:N. The \{tl var2\} is a “normal” token list variable, assigned locally using \tl_set:Nn.

\tl_build_end:N \{tl var\}
\tl_build_gend:N

Gets the contents of \{tl var\} and stores that into the \{tl var\} using \tl_set:Nn or \tl_gset:Nn. The \{tl var\} must have been set up with \tl_build_begin:N or \tl_build_gbegin:N. The end and gend functions must be used for local and global \{tl var\} respectively. These functions completely remove the setup code that enabled \{tl var\} to be used for other \tl_build_... functions.

37.14 Additions to \texttt{l3token}

\c_catcode_active_space_tl
\char_to_utfviii_bytes:n
\char_to_nfd:N

Token list containing one character with category code 13, (“active”), and character code 32 (space).

\char_to_utfviii_bytes:n \{\codepoint\}

Converts the (Unicode) \{\codepoint\} to UTF-8 bytes. The expansion of this function comprises four brace groups, each of which will contain a hexadecimal value: the appropriate byte. As UTF-8 is a variable-length, one or more of the groups may be empty: the bytes read in the logical order, such that a two-byte codepoint will have groups #1 and #2 filled and #3 and #4 empty.

\char_to_nfd:N \{char\}

Converts the \{char\} to the Unicode Normalization Form Canonical Decomposition. The category code of the generated character is the same as the \{char\}. With 8-bit engines, no change is made to the character.
Collects and removes tokens from the input stream until finding a token that does not match the \textit{test token} (as defined by the test \verb|\token_if_eq_catcode:NNTF| or \verb|\token_if_eq_charcode:NNTF| or \verb|\token_if_eq_meaning:NNTF|). The collected tokens are passed to the \textit{inline code} as \#1. When begin-group or end-group tokens (usually \{ or \}) are collected they are replaced by implicit \texttt{\c_group_begin_token} and \texttt{\c_group_end_token}, and when spaces (including \texttt{\c_space_token}) are collected they are replaced by explicit spaces.

For example the following code prints “Hello” to the terminal and leave “, world!” in the input stream.

\begin{verbatim}
\peek_catcode_collect_inline:Nn A { \iow_term:n {#1} } Hello,~world!
\end{verbatim}

Another example is that the following code tests if the next token is *, ignoring intervening spaces, but putting them back using \#1 if there is no *.

\begin{verbatim}
\peek_meaning_collect_inline:Nn \c_space_token \\{ \peek_charcode:NTF * { star } { no-star #1 } \}
\end{verbatim}

\begin{verbatim}
\peek_remove_spaces:n \\{\code\}
\end{verbatim}

Removes explicit and implicit space tokens (category code 10 and character code 32) from the input stream, then inserts \texttt{\code}.
Part VIII

Implementation
Chapter 38

\texttt{l3bootstrap} implementation

38.1 Lua\TeX{}-specific code

Depending on the versions available, the \LaTeX{} format may not have the raw \texttt{\textbackslash Umath} primitive names available. We fix that globally: it should cause no issues. Older Lua\TeX{} versions do not have a pre-built table of the primitive names here so sort one out ourselves. These end up globally-defined but at that is true with a newer format anyway and as they all start \texttt{\textbackslash U} this should be reasonably safe.

\begin{verbatim}
\begingroup
\expandafter\ifx\csname directlua\endcsname\relax
\else
\directlua{
local i
local t = { }
for _,i in pairs(tex.extraprimitives("luatex")) do
  if string.match(i,"^U") then
    if not string.match(i,"^Uchar$") then %$
      table.insert(t,i)
    end
  end
end
tex.enableprimitives("", t)
}\fi
\endgroup
\end{verbatim}

38.2 The \texttt{\textbackslash pdfstrcmp} primitive in Xg\TeX{}

Only pdf\TeX{} has a primitive called \texttt{\textbackslash pdfstrcmp}. The Xg\TeX{} version is just \texttt{\textbackslash strcmp}, so there is some shuffling to do. As this is still a real primitive, using the pdf\TeX{} name is “safe”.

\begin{verbatim}
\begingroup\expandafter\expandafter\expandafter\endgroup
\expandafter\ifx\csname pdfstrcmp\endcsname\relax
\let\pdfstrcmp\strcmp
\endverbatim
38.3 Loading support Lua code

When LuaTeX is used there are various pieces of Lua code which need to be loaded. The code itself is defined in \texttt{l3luatex} and is extracted into a separate file. Thus here the task is to load the Lua code both now and (if required) at the start of each job.

For LuaTeX we make sure the basic support is loaded: this is only necessary in plain.

As the user might be making a custom format, no assumption is made about matching package mode with only loading the Lua code once. Instead, a query to Lua reveals what mode is in operation.

38.4 Engine requirements

The code currently requires e-\TeX{} and functionality equivalent to \texttt{\texttt{pdfstrcmp}}, and also driver and Unicode character support. This is available in a reasonably-wide range of engines.

For LuaTeX, we require at least Lua 5.3 and the \texttt{token.set_lua} function. This is available at least since LuaTeX 1.10.
These are available in the engines:
- pdfTeX v1.40
- XeTeX v0.99992
- LuaTeX v1.10
- e-(u)pTeX mid-2012 or later.
38.5 Extending allocators

The ability to extend \TeX's allocation routine to allow for \varepsilon-\TeX{} has been around since 1997 in the etex package. Loading this support is delayed until here as we are now sure that the \varepsilon-\TeX{} extensions and \texttt{pdfstrcmp} or equivalent are available. Thus there is no danger of an “uncontrolled” error if the engine requirements are not met.

For \LaTeX{}2e we need to make sure that the extended pool is being used: expl3 uses a lot of registers. For formats from 2015 onward there is nothing to do as this is automatic. For older formats, the etex package needs to be loaded to do the job. In that case, some inserts are reserved also as these have to be from the standard pool. Note that \texttt{reserveinserts} is \texttt{outer} and so is accessed here by csname. In earlier versions, loading etex was done directly and so \texttt{reserveinserts} appeared in the code: this then required a \texttt{\relax} after \texttt{\RequirePackage} to prevent an error with “unsafe” definitions as seen for example with \texttt{capoptions}. The optional loading here is done using a group and \texttt{\ifx} test as we are not quite in the position to have a single name for \texttt{pdfstrcmp} just yet.

\begin{verbatim}
104 \begingroup
105 \def\@tempa{LaTeX2e}\
106 \def\next{}\
107 \ifx\fmtname\@tempa
108 \expandafter\ifx\csname extrafloats\endcsname\relax
109 \def\next
110 {\
111 \RequirePackage{etex}\
112 \csname reserveinserts\endcsname{32}\
113 }\
114 \fi
115 \fi
116 \expandafter\endgroup
117 \next
\end{verbatim}

38.6 The \LaTeX{}3 code environment

The code environment is now set up.

\texttt{\ExplSyntaxOff} Before changing any category codes, in package mode we need to save the situation before loading. Note the set up here means that once applied \texttt{\ExplSyntaxOff} becomes a “do nothing” command until \texttt{\ExplSyntaxOn} is used.

\begin{verbatim}
118 \protectededef\ExplSyntaxOff
119 {\
120 \protectededef\noexpand\ExplSyntaxOff{}\
121 \catcode 9 = \the\catcode 9\relax
122 \catcode 32 = \the\catcode 32\relax
123 \catcode 34 = \the\catcode 34\relax
124 \catcode 38 = \the\catcode 38\relax
125 \catcode 58 = \the\catcode 58\relax
126 \catcode 94 = \the\catcode 94\relax
127 \catcode 95 = \the\catcode 95\relax
128 \catcode 124 = \the\catcode 124\relax
129 \catcode 126 = \the\catcode 126\relax
130 \endlinechar = \the\endlinechar\relax
\end{verbatim}
The code environment is now set up.

\l__kernel_expl_bool The status for code syntax: this is on at present.

\ExplSyntaxOn The idea here is that multiple \ExplSyntaxOn calls are not going to mess up category codes, and that multiple calls to \ExplSyntaxOff are also not wasting time. Applying \ExplSyntaxOn alters the definition of \ExplSyntaxOff and so in package mode this function should not be used until after the end of the loading process!

\protected \def \ExplSyntaxOn
\boolexpr:nF \l__kernel_expl_bool
\cs_set_protected:Npx \ExplSyntaxOff
{
\char_set_catcode:nn { 9 } { \char_value_catcode:n { 9 } }
\char_set_catcode:nn { 32 } { \char_value_catcode:n { 32 } }
\char_set_catcode:nn { 34 } { \char_value_catcode:n { 34 } }
\char_set_catcode:nn { 38 } { \char_value_catcode:n { 38 } }
\char_set_catcode:nn { 58 } { \char_value_catcode:n { 58 } }
\char_set_catcode:nn { 94 } { \char_value_catcode:n { 94 } }
\char_set_catcode:nn { 95 } { \char_value_catcode:n { 95 } }
\char_set_catcode:nn { 124 } { \char_value_catcode:n { 124 } }
\char_set_catcode:nn { 126 } { \char_value_catcode:n { 126 } }
\tex_endlinechar:D = \tex_the:D \tex_endlinechar:D \scan_stop:
\bool_set_false:N \l__kernel_expl_bool
\cs_set_protected:NPn \ExplSyntaxOff { }
}
\char_set_catcode_ignore:n { 9 } \% tab
\char_set_catcode_ignore:n { 32 } \% space
\char_set_catcode_other:n { 34 } \% double quote
\char_set_catcode_alignment:n { 38 } \% ampersand
\char_set_catcode_letter:n { 58 } \% colon
\char_set_catcode_math_superscript:n { 94 } \% circumflex
\char_set_catcode_letter:n { 95 } \% underscore
\char_set_catcode_other:n { 124 } \% pipe
\char_set_catcode_space:n \{ 126 \} \% tilde
\tex_endlinechar:D = 32 \scan_stop:
\bool_set_true:N \_\_kernel_expr\_bool
}

(End definition for \ExplSyntaxOn. This function is documented on page 9.)

⟨/package⟩
Chapter 39

l3names implementation

The prefix here is \texttt{kernel}. A few places need \texttt{@@} to be left as is; this is obtained as \texttt{@@@}. The code here simply renames all of the primitives to new, internal, names. The \texttt{\let} primitive is renamed by hand first as it is essential for the entire process to follow. This also uses \texttt{\global}, as that way we avoid leaving an unneeded \texttt{csname} in the hash table.

\begin{verbatim}
\let \tex_global:D \global
\let \tex_let:D \let
\end{verbatim}

Everything is inside a (rather long) group, which keeps \texttt{\_kernel\_primitive:NN} trapped.

\verb!\_kernel\_primitive:NN! A temporary function to actually do the renaming.

\begin{verbatim}
\long \def \_kernel\_primitive:NN #1#2
{ \tex_global:D \tex_let:D #2 #1 }
\end{verbatim}

(End definition for \texttt{\_kernel\_primitive:NN}.)

To allow extracting “just the names”, a bit of DocStrip fiddling.

\begin{verbatim}
{/package & tex}
{/names | tex}
{/names | package}
\end{verbatim}

In the current incarnation of this package, all \TeX\ primitives are given a new name of the form \texttt{\tex\_oldname:D}. But first three special cases which have symbolic original names. These are given modified new names, so that they may be entered without catcode tricks.

\begin{verbatim}
\_kernel\_primitive:NN \tex\_space:D
\_kernel\_primitive:NN \slash \tex\_italiccorrection:D
\_kernel\_primitive:NN \- \tex\_hyphen:D
\end{verbatim}

Now all the other primitives.

\begin{verbatim}
\_kernel\_primitive:NN \above \tex\_above:D
\_kernel\_primitive:NN \abovedisplayshortskip \tex\_abovedisplayshortskip:D
\_kernel\_primitive:NN \abovedisplayskip \tex\_abovedisplayskip:D
\_kernel\_primitive:NN \abovewithdelims \tex\_abovewithdelims:D
\end{verbatim}
\_kernel\_primitive:NN \_accent \tex\_accent:D
\_kernel\_primitive:NN \_adjdemerits \tex\_adjdemerits:D
\_kernel\_primitive:NN \_advance \tex\_advance:D
\_kernel\_primitive:NN \_afterassignment \tex\_afterassignment:D
\_kernel\_primitive:NN \_aftergroup \tex\_aftergroup:D
\_kernel\_primitive:NN \_atop \tex\_atop:D
\_kernel\_primitive:NN \_atopwithdelims \tex\_atopwithdelims:D
\_kernel\_primitive:NN \_badness \tex\_badness:D
\_kernel\_primitive:NN \_baselineskip \tex\_baselineskip:D
\_kernel\_primitive:NN \_batchmode \tex\_batchmode:D
\_kernel\_primitive:NN \_begingroup \tex\_begingroup:D
\_kernel\_primitive:NN \_belowdisplayshortskip \tex\_belowdisplayshortskip:D
\_kernel\_primitive:NN \_belowdisplayskip \tex\_belowdisplayskip:D
\_kernel\_primitive:NN \_binoppenalty \tex\_binoppenalty:D
\_kernel\_primitive:NN \_botmark \tex\_botmark:D
\_kernel\_primitive:NN \_box \tex\_box:D
\_kernel\_primitive:NN \_boxmaxdepth \tex\_boxmaxdepth:D
\_kernel\_primitive:NN \_brokenpenalty \tex\_brokenpenalty:D
\_kernel\_primitive:NN \_catcode \tex\_catcode:D
\_kernel\_primitive:NN \_char \tex\_char:D
\_kernel\_primitive:NN \_chardef \tex\_chardef:D
\_kernel\_primitive:NN \_cleaders \tex\_cleaders:D
\_kernel\_primitive:NN \_closein \tex\_closein:D
\_kernel\_primitive:NN \_closeout \tex\_closeout:D
\_kernel\_primitive:NN \_clubpenalty \tex\_clubpenalty:D
\_kernel\_primitive:NN \_copy \tex\_copy:D
\_kernel\_primitive:NN \_count \tex\_count:D
\_kernel\_primitive:NN \_countdef \tex\_countdef:D
\_kernel\_primitive:NN \_cr \tex\_cr:D
\_kernel\_primitive:NN \_crrc \tex\_crrc:D
\_kernel\_primitive:NN \_csname \tex\_csname:D
\_kernel\_primitive:NN \_day \tex\_day:D
\_kernel\_primitive:NN \_deadcycles \tex\_deadcycles:D
\_kernel\_primitive:NN \_def \tex\_def:D
\_kernel\_primitive:NN \_defaulthyphenchar \tex\_defaulthyphenchar:D
\_kernel\_primitive:NN \_defaultskewchar \tex\_defaultskewchar:D
\_kernel\_primitive:NN \_delcode \tex\_delcode:D
\_kernel\_primitive:NN \_delimiter \tex\_delimiter:D
\_kernel\_primitive:NN \_delimiterfactor \tex\_delimiterfactor:D
\_kernel\_primitive:NN \_delimitershortfall \tex\_delimitershortfall:D
\_kernel\_primitive:NN \_dimen \tex\_dimen:D
\_kernel\_primitive:NN \_dimendef \tex\_dimendef:D
\_kernel\_primitive:NN \_discretionary \tex\_discretionary:D
\_kernel\_primitive:NN \_displayindent \tex\_displayindent:D
\_kernel\_primitive:NN \_displaylimits \tex\_displaylimits:D
\_kernel\_primitive:NN \_displaystyle \tex\_displaystyle:D
\_kernel\_primitive:NN \_displaywidowpenalty \tex\_displaywidowpenalty:D
\_kernel\_primitive:NN \_divide \tex\_divide:D
\_kernel\_primitive:NN \_doublehyphenpenemerges \tex\_doublehyphenpenemerges:D
\_kernel\_primitive:NN \_dp \tex\_dp:D
\_kernel\_primitive:NN \_dump \tex\_dump:D
\_kernel\_primitive:NN \_edef \tex\_edef:D
\_kernel\_primitive:NN \_else \tex\_else:D
Primitives introduced by \TeX. 

\begin{verbatim}
\_kernel_primitive:NN \toks \tex_toks:D
\_kernel_primitive:NN \toksdef \tex_toksdef:D
\_kernel_primitive:NN \tolerance \tex_tolerance:D
\_kernel_primitive:NN \topmark \tex_topmark:D
\_kernel_primitive:NN \topskip \tex_topskip:D
\_kernel_primitive:NN \tracingcommands \tex_tracingcommands:D
\_kernel_primitive:NN \tracinglostchars \tex_tracinglostchars:D
\_kernel_primitive:NN \tracingmacros \tex_tracingmacros:D
\_kernel_primitive:NN \tracingonline \tex_tracingonline:D
\_kernel_primitive:NN \tracingoutput \tex_tracingoutput:D
\_kernel_primitive:NN \tracingpages \tex_tracingpages:D
\_kernel_primitive:NN \tracingparagraphs \tex_tracingparagraphs:D
\_kernel_primitive:NN \tracingrestores \tex_tracingrestores:D
\_kernel_primitive:NN \tracingstats \tex_tracingstats:D
\_kernel_primitive:NN \uccode \tex_uccode:D
\_kernel_primitive:NN \uchyph \tex_uchyph:D
\_kernel_primitive:NN \underline \tex_underline:D
\_kernel_primitive:NN \unhbox \tex_unhbox:D
\_kernel_primitive:NN \unhcopy \tex_unhcopy:D
\_kernel_primitive:NN \unkern \tex_unkern:D
\_kernel_primitive:NN \unpenalty \tex_unpenalty:D
\_kernel_primitive:NN \unskip \tex_unskip:D
\_kernel_primitive:NN \unvbox \tex_unvbox:D
\_kernel_primitive:NN \unvcopy \tex_unvcopy:D
\_kernel_primitive:NN \uppercase \tex_uppercase:D
\_kernel_primitive:NN \vadjust \tex_vadjust:D
\_kernel_primitive:NN \valign \tex_valign:D
\_kernel_primitive:NN \vbadness \tex_vbadness:D
\_kernel_primitive:NN \vbox \tex_vbox:D
\_kernel_primitive:NN \vcenter \tex_vcenter:D
\_kernel_primitive:NN \vfil \tex_vfil:D
\_kernel_primitive:NN \vfill \tex_vfill:D
\_kernel_primitive:NN \vfilneg \tex_vfilneg:D
\_kernel_primitive:NN \vfuzz \tex_vfuzz:D
\_kernel_primitive:NN \voffset \tex_voffset:D
\_kernel_primitive:NN \vrule \tex_vrule:D
\_kernel_primitive:NN \vaize \tex_vaize:D
\_kernel_primitive:NN \vakip \tex_vakip:D
\_kernel_primitive:NN \vsplit \tex_vsplit:D
\_kernel_primitive:NN \vss \tex_vss:D
\_kernel_primitive:NN \vtop \tex_vtop:D
\_kernel_primitive:NN \wd \tex_wd:D
\_kernel_primitive:NN \widowpenalty \tex_widowpenalty:D
\_kernel_primitive:NN \write \tex_write:D
\_kernel_primitive:NN \xdef \tex_xdef:D
\_kernel_primitive:NN \xleaders \tex_xleaders:D
\_kernel_primitive:NN \xspaceskip \tex_xspaceskip:D
\_kernel_primitive:NN \year \tex_year:D
\end{verbatim}

Primitives introduced by \$\epsilon\$-\TeX.
Post-ε-TeX primitives do not always end up with the same name in all engines, if indeed they are available cross-engine anyway. We therefore take the approach of preferring the shortest name that makes sense. First, we deal with the primitives introduced by pdfTeX which directly relate to PDF output: these are copied with the names unchanged.
These are not related to PDF output and either already appear in other engines without the `pdf` prefix, or might reasonably do so at some future stage. We therefore drop the leading `pdf` here.
The version primitives are not related to PDF mode but are pdfTeX-specific, so again are carried forward unchanged.

These ones appear in pdfTeX but don’t have pdf in the name at all: no decisions to make.

Post pdfTeX primitive availability gets more complex. Both XeTeX and LuaTeX have varying names for some primitives from pdfTeX. Particularly for LuaTeX tracking all of that would be hard. Instead, we now check that we only save primitives if they actually exist.

Some pdfTeX primitives are handled here because they got dropped in LuaTeX but the corresponding internal names are emulated later. The Lua code is already loaded at this point, so we shouldn’t overwrite them.
Xe\TeX-specific primitives. Note that Xe\TeX’s `strcmp` is handled earlier and is “rolled up” into \pdfstrcmp. A few cross-compatibility names which lack the `pdf` of the original are handled later.

\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
\pdfstrcmp
Primitives from pdfTeX that XeTeX renames: also helps with LuaTeX.

Primitives from LuaTeX, some of which have been ported back to XeTeX.
Primitives from pdfTeX that LuaTeX renames.

\__kernel_primitive:NN \adjustspacing \tex_adjustspacing:D
\__kernel_primitive:NN \copyfont \tex_copyfont:D
\__kernel_primitive:NN \draftmode \tex_draftmode:D
\__kernel_primitive:NN \expandglyphsinfont \tex_fontexpand:D
\__kernel_primitive:NN \ifabedefim \tex_ifabedefim:D
\__kernel_primitive:NN \ifabenum \tex_ifabenum:D
\__kernel_primitive:NN \ignoreligaturesinfont \tex_ignoreligaturesinfont:D
\__kernel_primitive:NN \inserttext \tex_inserttext:D
\__kernel_primitive:NN \lastsavedboxresourceindex
\tex_pdflastxform:D
The set of Unicode math primitives were introduced by \texttt{\LaTeX} and \texttt{\LaTeXe} in a somewhat complex fashion: a few first as \texttt{\LaTeXe}... which were then renamed with \texttt{\LaTeXe} having a lot more. These names now all start \texttt{\Um...} and mainly \texttt{\Umath...}.
Primitives from pTeX.

__kernel_primitive:NN \Umathsubshiftdrop \tex_Umathsubshiftdrop:D
__kernel_primitive:NN \Umathsubsupshiftdown \tex_Umathsubsupshiftdown:D
__kernel_primitive:NN \Umathsubsupvgap \tex_Umathsubsupvgap:D
__kernel_primitive:NN \Umathsubtopmax \tex_Umathsubtopmax:D
__kernel_primitive:NN \Umathsupbottommin \tex_Umathsupbottommin:D
__kernel_primitive:NN \Umathsupshiftdrop \tex_Umathsupshiftdrop:D
__kernel_primitive:NN \Umathsupshiftup \tex_Umathsupshiftup:D
__kernel_primitive:NN \Umathsupsubbottommax \tex_Umathsupsubbottommax:D
__kernel_primitive:NN \Umathunderbarkern \tex_Umathunderbarkern:D
__kernel_primitive:NN \Umathunderbarrule \tex_Umathunderbarrule:D
__kernel_primitive:NN \Umathunderbarvgap \tex_Umathunderbarvgap:D
__kernel_primitive:NN \Umathunderdelimiterbgap \tex_Umathunderdelimiterbgap:D
__kernel_primitive:NN \Umathunderdelimitervgap \tex_Umathunderdelimitervgap:D
__kernel_primitive:NN \Unosubscript \tex_Unosubscript:D
__kernel_primitive:NN \Unosuperscript \tex_Unosuperscript:D
__kernel_primitive:NN \Uoverdelimiter \tex_Uoverdelimiter:D
__kernel_primitive:NN \Uradical \tex_Uradical:D
__kernel_primitive:NN \Uskewed \tex_Uskewed:D
__kernel_primitive:NN \Uskewedwithdelims \tex_Uskewedwithdelims:D
__kernel_primitive:NN \Ustack \tex_Ustack:D
__kernel_primitive:NN \Ustartdisplaymath \tex_Ustartdisplaymath:D
__kernel_primitive:NN \Ustartmath \tex_Ustartmath:D
__kernel_primitive:NN \Ustopdisplaymath \tex_Ustopdisplaymath:D
__kernel_primitive:NN \Ustopmath \tex_Ustopmath:D
__kernel_primitive:NN \Unosubscript \tex_Unosubscript:D
__kernel_primitive:NN \Unosuperscript \tex_Unosuperscript:D
__kernel_primitive:NN \Uvextensible \tex_Uvextensible:D

Primitives from pTeX.

__kernel_primitive:NN \autospacing \tex_autospacing:D
__kernel_primitive:NN \autoxspacing \tex_autoxspacing:D
__kernel_primitive:NN \currentcjktoken \tex_currentcjktoken:D
__kernel_primitive:NN \currentspacingmode \tex_currentspacingmode:D
__kernel_primitive:NN \currentxspacingmode \tex_currentxspacingmode:D
__kernel_primitive:NN \disinhibitglue \tex_disinhibitglue:D
__kernel_primitive:NN \dtou \tex_dtou:D
__kernel_primitive:NN \epTeXinputencoding \tex_epTeXinputencoding:D
__kernel_primitive:NN \epTeXVersion \tex_epTeXVersion:D
__kernel_primitive:NN \eucl \tex_euc:D
__kernel_primitive:NN \hfi \tex_hfi:D
__kernel_primitive:NN \ifdbox \tex_ifdbox:D
__kernel_primitive:NN \ifddir \tex_ifddir:D
__kernel_primitive:NN \iffont \tex_iffont:D
__kernel_primitive:NN \ifmbox \tex_ifmbox:D
__kernel_primitive:NN \ifmdir \tex_ifmdir:D
__kernel_primitive:NN \iftbox \tex_iftbox:D
__kernel_primitive:NN \iftfont \tex_iftfont:D
__kernel_primitive:NN \iftdir \tex_iftdir:D
__kernel_primitive:NN \ifybox \tex_ifybox:D
__kernel_primitive:NN \ifydir \tex_ifydir:D
__kernel_primitive:NN \inhibitglue \tex_inhibitglue:D

334
Primitives from upTEX.

Omega primitives provided by μTEX (listed separately mainly to allow understanding of their source).
Newer cross-engine primitives.

End of the “just the names” part of the source.

The job is done: close the group (using the primitive renamed!).

\LATEX moves a few primitives, so these are sorted out. In newer versions of \LATEX, expl3 is loaded rather early, so only some primitives are already renamed, so we need two tests here. At the beginning of the \LATEX format, the primitives \end and \input are renamed, and only later on the other ones.

If \@@hyph is defined, we are loading expl3 in a pre-2020/10/01 release of \LATEX, so a few other primitives have to be tested as well.

The \shipout primitive is particularly tricky as a number of packages want to hook in here. First, we see if a sufficiently-new kernel has saved a copy: if it has, just use that. Otherwise, we need to check each of the possible packages/classes that might move it: here, we are looking for those which do not delay action to the \AtBeginDocument hook. (We cannot use \primitive as that doesn’t allow us to make a direct copy of the primitive itself.) As we know that \LATEX is in use, we use it’s \@for loop here.
Some tidying up is needed for \textit{pdf}tracingfonts. Newer LuaTeX has this simply as \texttt{tracingfonts}, but that is overwritten by the \LaTeXe\ kernel. So any spurious definition has to be removed, then the real version saved either from the pdfTeX name or from LuaTeX. In the latter case, we leave \texttt{@@tracingfonts} available: this might be useful and almost all \LaTeXe\ users will have expl3 loaded by fontspec. (We follow the usual kernel convention that \texttt{@@} is used for saved primitives.)

That is also true for the LuaTeX primitives under \LaTeXe\ (depending on the format-building date). There are a few primitives that get the right names anyway so are missing here!
\texttt{\textunderscore let:D \texttt{\textunderscore late lua:D}} \texttt{\luatexlate lua}
\texttt{\textunderscore let:D \texttt{\textunderscore lua escape string:D}} \texttt{\luatexluaescapestring}
\texttt{\textunderscore let:D \texttt{\textunderscore lua function:D}} \texttt{\luatexluafunction}
\texttt{\textunderscore let:D \texttt{\textunderscore math style:D}} \texttt{\luatexmathstyle}
\texttt{\textunderscore let:D \texttt{\textunderscore nokerns:D}} \texttt{\luatexnokerns}
\texttt{\textunderscore let:D \texttt{\textunderscore noligs:D}} \texttt{\luatexnoligs}
\texttt{\textunderscore let:D \texttt{\textunderscore output box:D}} \texttt{\luatexoutputbox}
\texttt{\textunderscore let:D \texttt{\textunderscore page left offset:D}} \texttt{\luatexpagel eft offset}
\texttt{\textunderscore let:D \texttt{\textunderscore page top offset:D}} \texttt{\luatexpagetop offset}
\texttt{\textunderscore let:D \texttt{\textunderscore post ex hyphen char:D}} \texttt{\luatexpostexhyphenchar}
\texttt{\textunderscore let:D \texttt{\textunderscore post hyphen char:D}} \texttt{\luatexposthyphenchar}
\texttt{\textunderscore let:D \texttt{\textunderscore pre ex hyphen char:D}} \texttt{\luatexpreexhyphenchar}
\texttt{\textunderscore let:D \texttt{\textunderscore pre hyphen char:D}} \texttt{\luatexprehyphen char}
\texttt{\textunderscore let:D \texttt{\textunderscore save cat code table:D}} \texttt{\luatexsavecatcodetable}
\texttt{\textunderscore let:D \texttt{\textunderscore scantext tokens:D}} \texttt{\luatexscantexttokens}
\texttt{\textunderscore let:D \texttt{\textunderscore suppress ifcs name error:D}} \texttt{\luatexsuppressifcsnameerror}
\texttt{\textunderscore let:D \texttt{\textunderscore suppress long error:D}} \texttt{\luatexsuppresslongerror}
\texttt{\textunderscore let:D \texttt{\textunderscore suppress mat par error:D}} \texttt{\luatexsuppressmatparerror}
\texttt{\textunderscore let:D \texttt{\textunderscore suppress outer error:D}} \texttt{\luatexsuppressoutererror}
\texttt{\textunderscore let:D \texttt{\textunderscore U char:D}} \texttt{\luatexUchar}
\texttt{\textunderscore let:D \texttt{\textunderscore suppress font not found error:D}} \texttt{\luatexsuppressfontnotfounderror}
\texttt{\textunderscore let:D \texttt{\textunderscore body dir:D}} \texttt{\luatextextdir}
\texttt{\textunderscore let:D \texttt{\textunderscore box dir:D}} \texttt{\luatexboxdir}
\texttt{\textunderscore let:D \texttt{\textunderscore left ghost:D}} \texttt{\luatexleftghost}
\texttt{\textunderscore let:D \texttt{\textunderscore local broken penalty:D}} \texttt{\luatexlocalbrokenpenalty}
\texttt{\textunderscore let:D \texttt{\textunderscore local inter line penalty:D}} \texttt{\luatexlocalinterlinepenalty}
\texttt{\textunderscore let:D \texttt{\textunderscore local left box:D}} \texttt{\luatexlocalleftbox}
\texttt{\textunderscore let:D \texttt{\textunderscore local right box:D}} \texttt{\luatexlocalrightbox}
\texttt{\textunderscore let:D \texttt{\textunderscore math dir:D}} \texttt{\luatexmathdir}
\texttt{\textunderscore let:D \texttt{\textunderscore page bottom offset:D}} \texttt{\luatexpagebottomoffset}
\texttt{\textunderscore let:D \texttt{\textunderscore page dir:D}} \texttt{\luatexpagedir}
\texttt{\textunderscore let:D \texttt{\textunderscore page height:D}} \texttt{\luatexpageheight}
\texttt{\textunderscore let:D \texttt{\textunderscore page right offset:D}} \texttt{\luatexpagerightoffset}
\texttt{\textunderscore let:D \texttt{\textunderscore page width:D}} \texttt{\luatexpage width}
\texttt{\textunderscore let:D \texttt{\textunderscore pardir:D}} \texttt{\luatexpardir}
\texttt{\textunderscore let:D \texttt{\textunderscore right ghost:D}} \texttt{\luatexrightghost}
\texttt{\textunderscore let:D \texttt{\textunderscore text dir:D}} \texttt{\luatextextdir}
\texttt{\textunderscore fi:D}

Which also covers those slightly odd ones.

\texttt{\textunderscore let:D \texttt{\textunderscore body dir:D}} \texttt{\luatexbodydir}
\texttt{\textunderscore let:D \texttt{\textunderscore box dir:D}} \texttt{\luatexboxdir}
\texttt{\textunderscore let:D \texttt{\textunderscore left ghost:D}} \texttt{\luatexleftghost}
\texttt{\textunderscore let:D \texttt{\textunderscore local broken penalty:D}} \texttt{\luatexlocalbrokenpenalty}
\texttt{\textunderscore let:D \texttt{\textunderscore local inter line penalty:D}} \texttt{\luatexlocalinterlinepenalty}
\texttt{\textunderscore let:D \texttt{\textunderscore local left box:D}} \texttt{\luatexlocalleftbox}
\texttt{\textunderscore let:D \texttt{\textunderscore local right box:D}} \texttt{\luatexlocalrightbox}
\texttt{\textunderscore let:D \texttt{\textunderscore math dir:D}} \texttt{\luatexmathdir}
\texttt{\textunderscore let:D \texttt{\textunderscore page bottom offset:D}} \texttt{\luatexpagebottomoffset}
\texttt{\textunderscore let:D \texttt{\textunderscore page dir:D}} \texttt{\luatexpagedir}
\texttt{\textunderscore let:D \texttt{\textunderscore page height:D}} \texttt{\luatexpageheight}
\texttt{\textunderscore let:D \texttt{\textunderscore page right offset:D}} \texttt{\luatexpagerightoffset}
\texttt{\textunderscore let:D \texttt{\textunderscore page width:D}} \texttt{\luatexpage width}
\texttt{\textunderscore let:D \texttt{\textunderscore pardir:D}} \texttt{\luatexpardir}
\texttt{\textunderscore let:D \texttt{\textunderscore right ghost:D}} \texttt{\luatexrightghost}
\texttt{\textunderscore let:D \texttt{\textunderscore text dir:D}} \texttt{\luatextextdir}
\texttt{\textunderscore fi:D}

Only \texttt{pdf\TeX} and \texttt{Lua\TeX} define \texttt{\pdfmapfile} and \texttt{\pdfmapline}: Tidy up the fact that some format-building processes leave a couple of questionable decisions about that!

\texttt{\textunderscore ifnum:D 0}
\texttt{\textunderscore ifdefined:D \texttt{\textunderscore pdf\textversion:D}} \texttt{1 \textunderscore fi:D}
\texttt{\textunderscore ifdefined:D \texttt{\textunderscore lua\version:D}} \texttt{1 \textunderscore fi:D}
\texttt{= 0 \%}
\texttt{\textunderscore let:D \texttt{\textunderscore map file:D \texttt{\textunderscore undefined:D}}}
\texttt{\textunderscore let:D \texttt{\textunderscore map line:D \texttt{\textunderscore undefined:D}}}
\texttt{\textunderscore fi:D}

338
A few packages do unfortunate things to date-related primitives.

\begin{verbatim}
\textbegingroup:D
\textedef:D \l_tmpa_tl { \textmeaning:D \texttime:D }
\textedef:D \l_tmpb_tl { \textstring:D \time }
\textif:D \l_tmpa_tl \l_tmpb_tl
\textelse:D
\textglobal:D \textlet:D \texttime:D \textundefined:D
\textfi:D
\textedef:D \l_tmpa_tl { \textmeaning:D \textday:D }
\textedef:D \l_tmpb_tl { \textstring:D \day }
\textif:D \l_tmpa_tl \l_tmpb_tl
\textelse:D
\textglobal:D \textlet:D \textday:D \textundefined:D
\textfi:D
\textedef:D \l_tmpa_tl { \textmeaning:D \textmonth:D }
\textedef:D \l_tmpb_tl { \textstring:D \month }
\textif:D \l_tmpa_tl \l_tmpb_tl
\textelse:D
\textglobal:D \textlet:D \textmonth:D \textundefined:D
\textfi:D
\textedef:D \l_tmpa_tl { \textmeaning:D \textyear:D }
\textedef:D \l_tmpb_tl { \textstring:D \year }
\textif:D \l_tmpa_tl \l_tmpb_tl
\textelse:D
\textglobal:D \textlet:D \textyear:D \textundefined:D
\textfi:D
\textendgroup:D
\end{verbatim}

Up to v0.80, LuaTeX defines the pdfTeX version data: rather confusing. Removing them means that \textpdftexversion:D is a marker for pdfTeX alone: useful in engine-dependent code later.

\begin{verbatim}
\textife:D \textluatexversion:D
\textlet:D \textpdftexbanner:D \textundefined:D
\textlet:D \textpdftexrevision:D \textundefined:D
\textlet:D \textpdftexversion:D \textundefined:D
\textfi:D
\end{verbatim}

cslatex moves a couple of primitives which we recover here; as there is no other marker, we can only work by looking for the names.

\begin{verbatim}
\textife:D \orietveryjob
\textlet:D \texteveryjob:D \orietveryjob
\textfi:D
\end{verbatim}

For ConTeXt, two tests are needed. Both Mark II and Mark IV move several primitives: these are all covered by the first test, again using \textend as a marker. For Mark IV, a few more primitives are moved: they are implemented using some Lua code in the current ConTeXt.

\begin{verbatim}
\textife:D \normalend
\textlet:D \textend:D \normalend
\textlet:D \texteveryjob:D \normaleveryjob
\textlet:D \textinput:D \normalinput
\end{verbatim}

339
In \ref{luatex}, we additionally emulate some primitives using Lua code.

\begin{verbatim}
\let\language:D \normallanguage
\let\mathop:D \normalmathop\let\outer:D \normalouter
\let\over:D \normalover\let\vcenter:D \normalvcenter
\let\unexpanded:D \normalunexpanded
\let\expanded:D \normalexpanded
\fi
\ifdefined:D \normalitaliccorrection\let\hoffset:D \normalhoffset
\let\italiccorrection:D \normalitaliccorrection\let\voffset:D \normalvoffset
\let\showtokens:D \normalshowtokens\let\bodydir:D \spac_directions_normal_body_dir
\let\pagedir:D \spac_directions_normal_page_dir
\fi
\let\left:D \normalleft\let\middle:D \normalmiddle\let\right:D \normalright
\fi
\langle/\text\rangle

\text\_strcmp:D Compare two strings, expanding to 0 if they are equal, -1 if the first one is smaller and 1 if the second one is smaller. Here “smaller” refers to codepoint order which does not correspond to the user expected order for most non-ASCII strings.
\begin{verbatim}
local minus_tok = token.new(string.byte'\-', 12)
local zero_tok = token.new(string.byte'0', 12)
local one_tok = token.new(string.byte'1', 12)
luacmd('\text\_strcmp:D', function()
  local first = scan_string()
  local second = scan_string()
  if first < second then
    put_next(minus_tok, one_tok)
  else
    put_next(first == second and zero_tok or one_tok)
  end
end, 'global')
\end{verbatim}

(End definition for \text\_strcmp:D. This function is documented on page ??.)

\text\_Ucharcat:D Creating arbitrary chars using \text\_cprint. The alternative approach using token.put_next(token.create(...)) would be about 10% slower.
\begin{verbatim}
local cprint = \text\_cprint
luacmd('\text\_Ucharcat:D', function()
  local charcode = scan_int()
  local catcode = scan_int()
  \text\_cprint(catcode, utf8_char(charcode))
end, 'global')
\end{verbatim}

340
\texttt{\textbackslash tex\_filesize:D}  Wrap the function from \texttt{ltxtutils}.

```
1390 \texttt{luacmd(\texttt{\textbackslash tex\_filesize:D}, \texttt{function()})}
1391 \texttt{local size = filesize(scan\_string())}
1392 \texttt{if size then write(size) end}
1393 \texttt{end, \textquote{global}}
```

(End definition for \texttt{\textbackslash tex\_filesize:D}. This function is documented on page ??.)

\texttt{\textbackslash tex\_mdfivesum:D}  There are two cases: Either hash a file or a string. Both are already implemented in \LaTeX{} or built-in.

```
1394 \texttt{luacmd(\texttt{\textbackslash tex\_mdfivesum:D}, \texttt{function()})}
1395 \texttt{local hash}
1396 \texttt{if scan\_keyword\texttt{\textbackslash textquote{file}}} then
1397 \texttt{hash = filemd5sum(scan\_string())}
1398 \texttt{else}
1399 \texttt{hash = md5\_HEX(scan\_string())}
1400 \texttt{end}
1401 \texttt{if hash then write(hash) end}
1402 \texttt{end, \textquote{global}}
```

(End definition for \texttt{\textbackslash tex\_mdfivesum:D}. This function is documented on page ??.)

\texttt{\textbackslash tex\_filemoddate:D}  A primitive for getting the modification date of a file.

```
1403 \texttt{luacmd(\texttt{\textbackslash tex\_filemoddate:D}, \texttt{function()})}
1404 \texttt{local date = filemoddate(scan\_string())}
1405 \texttt{if date then write(date) end}
1406 \texttt{end, \textquote{global}}
```

(End definition for \texttt{\textbackslash tex\_filemoddate:D}. This function is documented on page ??.)

\texttt{\textbackslash tex\_filedump:D}  An emulated primitive for getting a hexdump from a (partial) file. The length has a default of 0. This is consistent with pdfTEX, but it effectively makes the primitive useless without an explicit \texttt{length}. Therefore we allow the keyword \texttt{whole} to be used instead of a length, indicating that the whole remaining file should be read.

```
1414 \texttt{luacmd(\texttt{\textbackslash tex\_filedump:D}, \texttt{function()})}
1415 \texttt{local offset = scan\_keyword\texttt{\textbackslash textquote{offset}}} and \texttt{scan\_int()} or \texttt{nil}
1416 \texttt{local length = scan\_keyword\texttt{\textbackslash textquote{length}}} and \texttt{scan\_int()}\texttt{or not scan\_keyword\texttt{\textbackslash textquote{whole}}} and \texttt{0} or \texttt{nil}
1417 \texttt{local data = filedump(scan\_string(), offset, length)}
1418 \texttt{if data then write(data) end}
1419 \texttt{end, \textquote{global}}
```

(End definition for \texttt{\textbackslash tex\_filedump:D}. This function is documented on page ??.)
Chapter 40

l3kernel-functions: kernel-reserved functions

40.1 Internal kernel functions

\_\_kernel_chk_cs_exist:N \_\_kernel_chk_cs_exist:N \{cs\}
This function is only created if debugging is enabled. It checks that \{cs\} exists according to the criteria for \cs_if_exist_p:N\, and if not raises a kernel-level error.

\_\_kernel_chk_cs_exist:c
\_\_kernel_chk_defined:NT \{variable\} \{true code\}
If \{variable\} is not defined (according to \cs_if_exist:NT\), this triggers an error, otherwise the \{true code\} is run.

\_\_kernel_chk_expr:nNnN \{expr\} \{eval\} \{convert\} \{caller\}
This function is only created if debugging is enabled. By default it is equivalent to \use_i:nnnn. When expression checking is enabled, it leaves in the input stream the result of \tex_the:D \{eval\} \{expr\} \tex_relax:D after checking that no token was left over. If any token was not taken as part of the expression, there is an error message displaying the result of the evaluation as well as the \{caller\}. For instance \{eval\} can be \_\_int_eval:w and \{caller\} can be \int_eval:n or \int_set:Nn. The argument \{convert\} is empty except for \mu expressions where it is \tex_mutoglue:D, used for internal purposes.

\_\_kernel_chk_tl_type:NnnT \_\_kernel_chk_tl_type:NNnT \{control sequence\} \{specific type\}
\{reconstruction\} \{true code\}
Helper to test that the \{control sequence\} is a variable of the given \{specific type\} of token list. Produces suitable error messages if the \{control sequence\} does not exist, or if it is not a token list variable at all, or if the \{control sequence\} differs from the result of \string-x-expanding \{reconstruction\}. If all of these tests succeed then the \{true code\} is run.
\__kernel_cs_parm_from_arg_count:nnF \__kernel_cs_parm_from_arg_count:nnF {\{follow-on\}} {\{args\}}
{\{false code\}}

Evaluates the number of \{args\} and leaves the \{follow-on\} code followed by a brace group containing the required number of primitive parameter markers (#1, etc.). If the number of \{args\} is outside the range [0, 9], the \{false code\} is inserted instead of the \{follow-on\}.

\__kernel_dependency_version_check:Nn \__kernel_dependency_version_check:Nn {\{date\}} {\{file\}}
\__kernel_dependency_version_check:nn \__kernel_dependency_version_check:nn {\{date\}} {\{file\}}

Checks if the loaded version of the expl3 kernel is at least \{date\}, required by \{file\}. If the kernel date is older than \{date\}, the loading of \{file\} is aborted and an error is raised.

\__kernel_deprecation_code:nn \__kernel_deprecation_code:nn {\{error code\}} {\{working code\}}

Stores both an \{error\} and \{working\} definition for given material such that they can be exchanged by \debug_on: and \debug_off:.

\__kernel_exp_not:w * \__kernel_exp_not:w \{expandable tokens\} \{\{content\}\}

Carries out expansion on the \{expandable tokens\} before preventing further expansion of the \{content\} as for \exp_not:n. Typically, the \{expandable tokens\} will alter the nature of the \{content\}, i.e. allow it to be generated in some way.

\l__kernel_expl_bool
A boolean which records the current code syntax status: true if currently inside a code environment. This variable should only be set by \ExplSyntaxOn/\ExplSyntaxOff.

(End definition for \l__kernel_expl_bool.)

\c__kernel_expl_date_tl
A token list containing the release date of the l3kernel preloaded in \LaTeX\ used to check if dependencies match.

(End definition for \c__kernel_expl_date_tl.)

\__kernel_file_missing:n \__kernel_file_missing:n {\{name\}}

Expands the \{name\} as per \__kernel_file_name_sanitize:n then produces an error message indicating that this file was not found.

\__kernel_file_name_sanitize:n * \__kernel_file_name_sanitize:n {\{name\}}

Expands the file name using a \csname-based approach, and relies on active characters (for example from UTF-8 characters) being properly set up to expand to a expansion-safe version using \ifcsname. This is less conservative than the token-by-token approach used before, but it is much faster.

\__kernel_file_input_push:n \__kernel_file_input_push:n {\{name\}}
\__kernel_file_input_pop:

Used to push and pop data from the internal file stack: needed only in package mode, where interfacing with the \LaTeX\ kernel is necessary.

Updated: 2021-04-17
\__kernel_int_add:nnn \__kernel_int_add:nnn \{\text{integer}_1\} \{\text{integer}_2\} \{\text{integer}_3\}

Expands to the result of adding the three \textit{integers} (which must be suitable input for \texttt{\textbackslash int_eval:w}), avoiding intermediate overflow. Overflow occurs only if the overall result is outside \([-2^{31}+1, 2^{31}-1]\). The \textit{integers} may be of the form \texttt{\textbackslash int_eval:w... \textbackslash scan_stop}: but may be evaluated more than once.

\__kernel_intarray_gset:Nnn \__kernel_intarray_gset:Nnn \{\text{intarray var}\} \{\text{index}\} \{\text{value}\}

Faster version of \texttt{\textbackslash intarray_gset:Nnn}. Stores the \textit{value} into the \textit{integer array variable} at the \textit{position}. The \textit{index} and \textit{value} must be suitable for a direct assignment to a \TeX{} count register, for instance expanding to an integer denotation or obtained through the primitive \texttt{\textbackslash numexpr} (which may be un-terminated). No bound checking is performed: the caller is responsible for ensuring that the \textit{position} is between 1 and the \texttt{\textbackslash intarray_count:N}, and the \textit{value}'s absolute value is at most \(2^{30}-1\). Assignments are always global.

\__kernel_intarray_item:Nn \__kernel_intarray_item:Nn \{\text{intarray var}\} \{\text{index}\}

Faster version of \texttt{\textbackslash intarray_item:Nn}. Expands to the integer entry stored at the \textit{index} in the \textit{integer array variable}. The \textit{index} must be suitable for a direct assignment to a \TeX{} count register and must be between 1 and the \texttt{\textbackslash intarray_count:N}, lest a low-level \TeX{} error occur.

\__kernel_intarray_range_to_clist:Nnn \__kernel_intarray_range_to_clist:Nnn \{\text{intarray var}\} \{\text{start index}\} \{\text{end index}\}

Converts to integer denotations separated by commas the entries of the \textit{intarray} from positions \textit{start index} to \textit{end index} included. The \textit{start index} and \textit{end index} must be suitable for a direct assignment to a \TeX{} count register, must be between 1 and the \texttt{\textbackslash intarray_count:N}, and be suitably ordered. All tokens have category code other.

\__kernel_intarray_gset_range_from_clist:Nnn \__kernel_intarray_gset_range_from_clist:Nnn \{\text{intarray var}\} \{\text{start index}\} \{\text{integer comma list}\}

Stores the entries of the \textit{clist} as entries of the \textit{intarray var} starting from the \textit{start index}, upwards. This is done without any bound checking. The \textit{start index} and all entries of the \textit{integer comma list} (which do not undergo space trimming and brace stripping as in normal clist mappings) must be suitable for a direct assignment to a \TeX{} count register. An empty entry may stop the loop.

\__kernel_ior_open:Nn \__kernel_ior_open:No \{\text{stream}\} \{\text{file name}\}

This function has identical syntax to the public version. However, it does not take precautions against active characters in the \textit{file name}, and it does not attempt to add a \textit{path} to the \textit{file name}: it is therefore intended to be used by higher-level functions which have already fully expanded the \textit{file name} and which need to perform multiple open or close operations. See for example the implementation of \texttt{\textbackslash file_get_full\_name:nN},
\_\_kernel_iow_with:Nnn  \_\_kernel_iow_with:Nnn \( integer \) \{value\} \{code\}

If the \( integer \) is equal to the \( value \) then this function simply runs the \( code \). Otherwise it saves the current value of the \( integer \), sets it to the \( value \), runs the \( code \), and restores the \( integer \) to its former value. This is used to ensure that the newline_char is 10 when writing to a stream, which lets \texttt{iow_newline} work, and that errorcontextlines is $-1$ when displaying a message.

\_\_kernel_kern:n  \_\_kernel_kern:n \{length\}

Inserts a kern of the specified \( length \), a dimension expression.

(\textit{End definition for \_\_kernel_kern:n}.)

\g__kernel_prg_map_int  This integer is used by non-expandable mapping functions to track the level of nesting in force. The functions \( \texttt{\textbackslash (type)\_map\_1:w, \texttt{(type)\_map\_2:w, etc.} \), labelled by \texttt{\g__kernel_prg_map_int} hold functions to be mapped over various list datatypes in inline and variable mappings.

(\textit{End definition for \g__kernel_prg_map_int}.)
\__kernel_quark_new_test:N \__kernel_quark_new_test:N \langle name \rangle:(arg spec)

Defines a quark-test function \langle name \rangle:(arg spec) which tests if its argument is \q__-\langle namespace \rangle_recursion_tail, then acts accordingly, as described below for each possible (arg spec).

The \langle namespace \rangle is determined as the first (nonempty) \_delimited word in \langle name \rangle and is used internally in the definition of auxiliaries. The function \__kernel_quark_new_test:N does not define the \q__-\langle namespace \rangle_recursion_tail and \q__-\langle namespace \rangle_recursion_stop quarks. They should be manually defined with \quark_new:N.

There are 6 different types of quark-test functions. Which one is defined depends on the (arg spec), which must be one of the options listed now. Four of them are modeled after \quark_if_recursion_tail:(N|n) and \quark_if_recursion_tail_do:(N|n)n.

n defines \langle name \rangle:n such that it checks if #1 contains only \q__-\langle namespace \rangle_recursion_tail, and if so consumes all tokens up to \q__-\langle namespace \rangle_recursion_stop (c.f. \quark_if_recursion_tail_stop:n).

nn defines \langle name \rangle:nn such that it checks if #1 contains only \q__-\langle namespace \rangle_recursion_tail, and if so consumes all tokens up to \q__-\langle namespace \rangle_recursion_stop, then executes the code #2 after that (c.f. \quark_if_recursion_tail_stop_do:nn).

N defines \langle name \rangle:N such that it checks if #1 is \q__-\langle namespace \rangle_recursion_tail, and if so consumes all tokens up to \q__-\langle namespace \rangle_recursion_stop (c.f. \quark_if_recursion_tail_stop:N).

Nn defines \langle name \rangle:Nn such that it checks if #1 is \q__-\langle namespace \rangle_recursion_tail, and if so consumes all tokens up to \q__-\langle namespace \rangle_recursion_stop, then executes the code #2 after that (c.f. \quark_if_recursion_tail_stop_do:Nn).

The last two are modeled after \quark_if_recursion_tail_break:(n|N)N, and in those cases the quark \q__-\langle namespace \rangle_recursion_stop is not used (and thus needs not be defined).

nN defines \langle name \rangle:nN such that it checks if #1 contains only \q__-\langle namespace \rangle_recursion_tail, and if so uses the \langle type \rangle_map_break: function #2.

NN defines \langle name \rangle:NN such that it checks if #1 is \q__-\langle namespace \rangle_recursion_tail, and if so uses the \langle type \rangle_map_break: function #2.

Any other signature, as well as a function without signature are errors, and in such case the definition is aborted.
\__kernel_quark_new_conditional:Nn  \__kernel_quark_new_conditional:Nn  \__\langle namespace\rangle\_quark_if\_\langle name\rangle:\langle arg spec\rangle \{\langle conditions\rangle\}

Defines a collection of quark conditionals that test if their argument is the quark \q\_\__\langle namespace\rangle\_\langle name\rangle and perform suitable actions. The \langle conditions\rangle are a comma-separated list of one or more of p, T, F, and TF, and one conditional is defined for each \langle condition\rangle in the list, as described for \prg_new_conditional:NNnn. The conditionals are defined using \prg_new_conditional:NNnn, so that their name is obtained by adding p, T, F, or TF to the base name \__\langle namespace\rangle\_quark_if\_\langle name\rangle:\langle arg spec\rangle.

The first argument of \__kernel_quark_new_conditional:Nn must contain _quark_if_ and :, as these markers are used to determine the \langle name\rangle of the quark \q\_\__\langle namespace\rangle\_\langle name\rangle to be tested. This quark should be manually defined with \quark_new:N, as \__kernel_quark_new_conditional:Nn does not define it.

The function \__kernel_quark_new_conditional:Nn can define 2 different types of quark conditionals. Which one is defined depends on the \langle arg spec\rangle, which must be one of the following options, modeled after \quark_if_nil:(N|n)(TF):

n defines \__\langle namespace\rangle\_quark_if\_\langle name\rangle:n(TF) such that it checks if \#1 contains only \q\_\__\langle namespace\rangle\_\langle name\rangle, and executes the proper conditional branch.

N defines \__\langle namespace\rangle\_quark_if\_\langle name\rangle:N(TF) such that it checks if \#1 is \q\_\__\langle namespace\rangle\_\langle name\rangle, and executes the proper conditional branch.

Any other signature, as well as a function without signature are errors, and in such case the definition is aborted.

\c\__kernel_randint_max_int  Maximal allowed argument to \__kernel_randint:n. Equal to \(2^{17} - 1\).

(End definition for \c\__kernel_randint_max_int.)

\__kernel_randint:n  \__kernel_randint:n \{\langle max\rangle\}

Used in an integer expression this gives a pseudo-random number between 1 and \langle max\rangle included. One must have \langle max\rangle \leq 2^{17} - 1. The \langle max\rangle must be suitable for \int_value:w (and any \int_eval:w must be terminated by \scan_stop: or equivalent).

\__kernel_randint:nn  \__kernel_randint:nn \{\langle min\rangle\} \{\langle max\rangle\}

Used in an integer expression this gives a pseudo-random number between \langle min\rangle and \langle max\rangle included. The \langle min\rangle and \langle max\rangle must be suitable for \int_value:w (and any \int_eval:w must be terminated by \scan_stop: or equivalent). For small ranges \(R = \langle max\rangle - \langle min\rangle + 1 \leq 2^{17} - 1, \langle min\rangle - 1 + \__kernel_randint:n\{R\} is faster.

\__kernel_register_show:N  \__kernel_register_show:N \langle register\rangle

Used to show the contents of a \TeX register at the terminal, formatted such that internal parts of the mechanism are not visible.

\__kernel_register_log:N  \__kernel_register_log:N \langle register\rangle

Used to write the contents of a \TeX register to the log file in a form similar to \__kernel_register_show:N.
\texttt{\_\_kernel\_str\_to\_other:n} \hspace{1em} \texttt{\_\_kernel\_str\_to\_other:n \{(token list)\}}

Converts the (\textit{token list}) to a (\textit{other string}), where spaces have category code “other”. This function can be \texttt{f}-expanded without fear of losing a leading space, since spaces do not have category code 10 in its result. It takes a time quadratic in the character count of the string.

\texttt{\_\_kernel\_str\_to\_other\_fast:n} \hspace{1em} \texttt{\_\_kernel\_str\_to\_other\_fast:n \{(token list)\}}

Same behaviour \texttt{\_\_kernel\_str\_to\_other:n} but only restricted-expandable. It takes a time linear in the character count of the string.

\texttt{\_\_kernel\_tl\_to\_str:w} \hspace{1em} \texttt{\_\_kernel\_tl\_to\_str:w \{expandable tokens\} \{tokens\}}

Carries out expansion on the (\textit{expandable tokens}) before conversion of the (\textit{tokens}) to a string as describe for \texttt{\_\_kernel\_tl\_to\_str:n}. Typically, the (\textit{expandable tokens}) will alter the nature of the (\textit{tokens}), \textit{i.e.} allow it to be generated in some way. This function requires only a single expansion.

\texttt{\_\_kernel\_tl\_set:Nx} \hspace{1em} \texttt{\_\_kernel\_tl\_set:Nx \{tl var\} \{tokens\}}
\texttt{\_\_kernel\_tl\_gset:Nx}

Fully expands (\textit{tokens}) and assigns the result to (\textit{tl var}). (\textit{tokens}) must be given in braces and there must be no token between (\textit{tl var}) and (\textit{tokens}).

\section{Kernel backend functions}

These functions are required to pass information to the backend. The nature of these means that they are defined only when the relevant backend is in use.

\texttt{\_\_kernel\_backend\_literal:n} \hspace{1em} \texttt{\_\_kernel\_backend\_literal:n \{content\}}
\texttt{\_\_kernel\_backend\_literal:x}

Adds the (\textit{content}) literally to the current vertical list as a whatsit. The nature of the (\textit{content}) will depend on the backend in use.

\texttt{\_\_kernel\_backend\_literal\_postscript:n} \hspace{1em} \texttt{\_\_kernel\_backend\_literal\_postscript:n \{PostScript\}}
\texttt{\_\_kernel\_backend\_literal\_postscript:x}

Adds the (\textit{PostScript}) literally to the current vertical list as a whatsit. No positioning is applied.

\texttt{\_\_kernel\_backend\_literal\_pdf:n} \hspace{1em} \texttt{\_\_kernel\_backend\_literal\_pdf:n \{PDF instructions\}}
\texttt{\_\_kernel\_backend\_literal\_pdf:x}

Adds the (\textit{PDF instructions}) literally to the current vertical list as a whatsit. No positioning is applied.

\texttt{\_\_kernel\_backend\_literal\_svg:n} \hspace{1em} \texttt{\_\_kernel\_backend\_literal\_svg:n \{SVG instructions\}}
\texttt{\_\_kernel\_backend\_literal\_svg:x}

Adds the (\textit{SVG instructions}) literally to the current vertical list as a whatsit. No positioning is applied.
\__kernel_backend_postscript:n {\langle\textit{PostScript}\rangle}
\__kernel_backend_postscript:x

Adds the \langle\textit{PostScript}\rangle to the current vertical list as a whatsis. The PostScript reference point is adjusted to match the current position. The PostScript is inserted inside a SDict begin/end pair.

\__kernel_backend_align_begin:
\__kernel_backend_align_end:
\langle\textit{PostScript literals}\rangle
\__kernel_backend_align_end:

Arranges to align the PostScript and DVI current positions and scales.

\__kernel_backend_scope_begin:
\__kernel_backend_scope_end:
\langle\textit{content}\rangle
\__kernel_backend_scope_end:

Creates a scope for instructions at the backend level.

\__kernel_backend_matrix:n {\langle\textit{matrix}\rangle}
\__kernel_backend_matrix:x

Applies the \langle\textit{matrix}\rangle to the current transformation matrix.

\g__kernel_backend_header_bool

Specifies whether to write headers for the backend.

\l__kernel_color_stack_int

The color stack used in \texttt{pdftex} and \texttt{luatex} for the main color.
Chapter 41

\texttt{lst3\texttt{basics implementation}}

\section*{41.1 Renaming some \TeX\ primitives (again)}

Having given all the \TeX\ primitives a consistent name, we need to give sensible names to the ones we actually want to use. These will be defined as needed in the appropriate modules, but we do a few now, just to get started.\footnote{This renaming gets expensive in terms of csname usage, an alternative scheme would be to just use the \texttt{\textbackslash tex...} name in the cases where no good alternative exists.}

\begin{verbatim}
\def\iftrue{\def\iftrue{\expandafter\ifnum\csname if\endcsname\relax}
\def\iffalse{\def\iffalse{\expandafter\ifnum\csname if\endcsname\relax}
\def\or{\def\or{\expandafter\ifnum\csname if\endcsname\relax}
\def\else{\def\else{\expandafter\ifnum\csname if\endcsname\relax}
\def\fi{\def\fi{\expandafter\ifnum\csname if\endcsname\relax}
\def\reverse_if{\def\reverse_if{\expandafter\ifnum\csname if\endcsname\relax}
\def\ifw{\def\ifw{\expandafter\ifnum\csname if\endcsname\relax}
\def\if_charcode{\def\if_charcode{\expandafter\ifnum\csname if\endcsname\relax}
\def\if_catcode{\def\if_catcode{\expandafter\ifnum\csname if\endcsname\relax}
\def\if_meaning{\def\if_meaning{\expandafter\ifnum\csname if\endcsname\relax}
(End definition for \texttt{\iftrue} and others. These functions are documented on page \pageref{page}.)
\def\if_mode_math{\def\if_mode_math{\expandafter\ifnum\csname if\endcsname\relax}
\def\if_mode_horizontal{\def\if_mode_horizontal{\expandafter\ifnum\csname if\endcsname\relax}
\def\if_mode_vertical{\def\if_mode_vertical{\expandafter\ifnum\csname if\endcsname\relax}
\def\if_mode_inner{\def\if_mode_inner{\expandafter\ifnum\csname if\endcsname\relax}
(End definition for \texttt{\if_mode_math} and others. These functions are documented on page \pageref{page}.)
\def\if_cs_exist{\def\if_cs_exist{\expandafter\ifnum\csname if\endcsname\relax}
\def\ifcsname{\def\ifcsname{\expandafter\ifnum\csname if\endcsname\relax}
\def\csname{\def\csname{\expandafter\ifnum\csname if\endcsname\relax}
\def\csend{\def\csend{\expandafter\ifnum\csname if\endcsname\relax}
Building csnames and testing if control sequences exist.
\def\ifcsname{\def\ifcsname{\expandafter\ifnum\csname if\endcsname\relax}
\end{verbatim}

Then some conditionals.

\begin{verbatim}
(End definition for \texttt{\iftrue} and others. These functions are documented on page \pageref{page}.)
\end{verbatim}
The five `\exp_` functions are used in the `\l3\expa` module where they are described.

```
\begin{verbatim}
\exp_after:wN \exp_not:N \exp_not:n
\end{verbatim}
```

Exchanging a control sequence or token.

```
\begin{verbatim}
\token_to_meaning:N \cs_meaning:N
\end{verbatim}
```

Making strings.

```
\begin{verbatim}
\tl_to_str:n \token_to_str:N \__kernel_tl_to_str:w
\end{verbatim}
```

Adding material after the end of a group.

```
\begin{verbatim}
\text\text\text\if_int_compare:w \__int_to_roman:w
\text\text\text\exp_args:Nc \exp_args:cc
\text\text\text\group_insert_after:N
\text\text\text\exp_args:Nc
\text\text\text\exp_args:cc
\end{verbatim}
```

For integers.

```
\begin{verbatim}
\if_int_compare:w \__int_to_roman:w
\end{verbatim}
```

Adding material after the end of a group.

```
\begin{verbatim}
\group_insert_after:N
\end{verbatim}
```

Discussed in `\l3\expa`, but needed much earlier.

```
\begin{verbatim}
\exp_args:Nc \exp_args:cc
\end{verbatim}
```

(End definition for `\exp_after:wN`, `\exp_not:N`, and `\exp_not:n`. These functions are documented on page 38.)

(End definition for `\token_to_meaning:N` and `\cs_meaning:N`. These functions are documented on page 186.)

(End definition for `\token_to_str:n`, `\token_to_str:N`, and `\__kernel_tl_to_str:w`. These functions are documented on page 106.)

(End definition for `\scan_stop:`, `\group_begin:`, and `\group_end:`. These functions are documented on page 13.)

(End definition for `\if_int_compare:w` and `\__int_to_roman:w`. This function is documented on page 166.)

(End definition for `\group_insert_after:N`. This function is documented on page 14.)

(End definition for `\exp_args:Nc` and `\exp_args:cc`. These functions are documented on page 34.)
A small number of variants defined by hand. Some of the necessary functions (\use_i:nn, \use_ii:nn, and \exp_args:NNc) are not defined at that point yet, but will be defined before those variants are used. The \cs_meaning:c command must check for an undefined control sequence to avoid defining it mistakenly.

\begin{verbatim}
\text_def:D \token_to_str:c { \exp_args:Nc \token_to_str:N }
\text_long:D \text_def:D \cs_meaning:c #1
{
  \ifcs_exist:w #1 \cs_end:
  \exp_after:wN \use_i:nn
  \else:
  \exp_after:wN \use_ii:nn
  \fi:
  { \exp_args:Nc \cs_meaning:N {#1} }
  { \tl_to_str:n {undefined} }
}
\text_let:D \token_to_meaning:c = \cs_meaning:c
\end{verbatim}

(End definition for \token_to_meaning:N. This function is documented on page 186.)

1457 \tex_def:D \token_to_str:c { \exp_args:Nc \token_to_str:N }
1458 \tex_def:D \cs_meaning:c #1
1459 {
  \ifcs_exist:w #1 \cs_end:
  \exp_after:wN \use_i:nn
  \else:
  \exp_after:wN \use_ii:nn
  \fi:
  { \exp_args:Nc \cs_meaning:N {#1} }
  { \tl_to_str:n {undefined} }
}
1460 \tex_let:D \token_to_meaning:c = \cs_meaning:c

(End definition for \token_to_meaning:N. This function is documented on page 186.)

\subsection{Defining some constants}

\c_zero_int We need the constant \c_zero_int which is used by some functions in the l3alloc module. The rest are defined in the l3int module — at least for the ones that can be defined with \tex_chardef:D or \tex_mathchardef:D. For other constants the l3int module is required but it can’t be used until the allocation has been set up properly!

\begin{verbatim}
\tex_chardef:D \c_zero_int = 0 ~
\tex_else:D
\tex_ifdefined:D \tex_luatexversion:D
\tex_chardef:D \c_max_register_int = 65535 ~
\tex_else:D
\tex_mathchardef:D \c_max_register_int = 32767 ~
\tex_fi:D
\end{verbatim}

(End definition for \c_zero_int. This variable is documented on page 165.)

\c_max_register_int This is here as this particular integer is needed both in package mode and to bootstrap \l3alloc, and is documented in l3int. \LaTeX{} and those which contain parts of the Omega extensions have more registers available than \TeX{}.

\begin{verbatim}
\tex_ifdefined:D \tex_luatexversion:D
\tex_chardef:D \c_max_register_int = 65535 ~
\tex_else:D
\tex_ifdefined:D \tex_omathchardef:D
\tex_omathchardef:D \c_max_register_int = 65535 ~
\tex_else:D
\tex_mathchardef:D \c_max_register_int = 32767 ~
\tex_fi:D
\end{verbatim}

(End definition for \c_max_register_int. This variable is documented on page 165.)

\subsection{Defining functions}

We start by providing functions for the typical definition functions. First the local ones.

All assignment functions in \TeX{} should be naturally protected; after all, the \TeX{} primitives for assignments are and it can be a cause of problems if others aren’t.

\begin{verbatim}
\cs_set:Npn \cs_set_protected:Npx
\cs_set_protected:Nx
\cs_set:Nx
\cs_set_protected:N
\cs_set:N
\cs_set_nopar:Npx
\cs_set_nopar:N
\cs_set_protected_nopar:Npx
\cs_set_protected_nopar:N
\cs_set_protected:N
\cs_set_protected:Npx
\end{verbatim}

352
41.4 Selecting tokens

\l__exp_internal_tl \Scratch token list variable for \texttt{l3expansion}, used by \texttt{\use:x}, used in defining conditionals. We don’t use \texttt{tl} methods because \texttt{l3basics} is loaded earlier.

(End definition for \texttt{\l__exp_internal_tl}.)

\use:c This macro grabs its argument and returns a csname from it.

(End definition for \texttt{\use:c}.)

\use:x Fully expands its argument and passes it to the input stream. Uses the reserved \texttt{\l__-_exp_internal_tl} which we’ve set up above.

(End definition for \texttt{\use:x}. This function is documented on page 20.)
\use:e In non-LuaTeXengines older than 2019, \expanded{} is emulated.
\cs_set:Npn \use:e #1 { \tex_expanded:D {#1} }
\tex_ifdefined:D \tex_expanded:D \tex_else:D
\cs_set:Npn \use:e #1 { \exp_args:Ne \use:n {#1} }
\tex_fi:D

(End definition for \use:e. This function is documented on page 25.)

\use:n
\use:nn
\use:nnn
\use:nnnn
These macros grab their arguments and return them back to the input (with outer braces removed).
\cs_set:Npn \use:n #1 {#1}
\cs_set:Npn \use:nn #1#2 {#1#2}
\cs_set:Npn \use:nnn #1#2#3 {#1#2#3}
\cs_set:Npn \use:nnnn #1#2#3#4 {#1#2#3#4}

(End definition for \use:n and others. These functions are documented on page 23.)

\use_i:nn
\use_ii:nn
\use_i:nnn
\use_ii:nnn
\use_iii:nnn
\use_i_ii:nnn
\use_i:nnnn
\use_ii:nnnn
\use_iii:nnnn
\use_iv:nnnn
We also need something for picking up arguments from a longer list.
\cs_set:Npn \use_i:nnn #1#2#3 {#1}
\cs_set:Npn \use_ii:nnn #1#2#3 {#2}
\cs_set:Npn \use_iii:nnn #1#2#3 {#3}
\cs_set:Npn \use_i_ii:nnn #1#2#3 {#1#2}
\cs_set:Npn \use_i:nnnn #1#2#3#4 {#1}
\cs_set:Npn \use_ii:nnnn #1#2#3#4 {#2}
\cs_set:Npn \use_iii:nnnn #1#2#3#4 {#3}
\cs_set:Npn \use_iv:nnnn #1#2#3#4 {#4}

(End definition for \use_i:nnn and others. These functions are documented on page 24.)

\use_ii_i:nn
\cs_set:Npn \use_ii_i:nn #1#2 { #2 #1 }

(End definition for \use_ii_i:nn. This function is documented on page 24.)

\use_none_delimit_by_q_nil:w
\use_none_delimit_by_q_stop:w
\use_none_delimit_by_q_recursion_stop:w
Functions that gobble everything until they see either \q_nil, \q_stop, or \q_recursion_stop, respectively.
\cs_set:Npn \use_none_delimit_by_q_nil:w #1 \q_nil { }
\cs_set:Npn \use_none_delimit_by_q_stop:w #1 \q_stop { }
\cs_set:Npn \use_none_delimit_by_q_recursion_stop:w #1 \q_recursion_stop { }

(End definition for \use_none_delimit_by_q_nil:w, \use_none_delimit_by_q_stop:w, and \use_none_delimit_by_q_recursion_stop:w. These functions are documented on page 25.)
Same as above but execute first argument after gobbling. Very useful when you need to skip the rest of a mapping sequence but want an easy way to control what should be expanded next.

```
\cs_set:Npn \use_i_delimit_by_q_nil:nw #1#2 \q_nil {#1}
\cs_set:Npn \use_i_delimit_by_q_stop:nw #1#2 \q_stop {#1}
\cs_set:Npn \use_i_delimit_by_q_recursion_stop:nw 
\#1\#2 \q_recursion_stop \{#1\}
```

(End definition for \use_i_delimit_by_q_nil:nw, \use_i_delimit_by_q_stop:nw, and \use_i_delimit_by_q_recursion_stop:nw. These functions are documented on page 25.)

### 41.5 Gobbling tokens from input

To gobble tokens from the input we use a standard naming convention: the number of tokens gobbled is given by the number of \textit{n}'s following the : in the name. Although we could define functions to remove ten arguments or more using separate calls of \textbackslash use_{\text{-none}}:nnnnn, this is very non-intuitive to the programmer who will assume that expanding such a function once takes care of gobbling all the tokens in one go.

```
\cs_set:Npn \use_none:n #1 { }
\cs_set:Npn \use_none:nn #1#2 { }
\cs_set:Npn \use_none:nnn #1#2#3 { }
\cs_set:Npn \use_none:nnnn #1#2#3#4 { }
\cs_set:Npn \use_none:nnnnn #1#2#3#4#5 { }
\cs_set:Npn \use_none:nnnnnn #1#2#3#4#5#6 { }
\cs_set:Npn \use_none:nnnnnnn #1#2#3#4#5#6#7 { }
\cs_set:Npn \use_none:nnnnnnnn #1#2#3#4#5#6#7#8 { }
\cs_set:Npn \use_none:nnnnnnnn #1#2#3#4#5#6#7#8#9 { }
```

(End definition for \textbackslash use_{\text{-none}}:n and others. These functions are documented on page 25.)

### 41.6 Debugging and patching later definitions

A more meaningful test of whether debugging is enabled than messing up with guards. We can also more easily change the logic in one place then. This is needed primarily for deprecations.

```
\cs_set_protected:Npn \__kernel_if_debug:TF #1#2 {#2}
```

(End definition for \textbackslash \_\_kernel_if_debug:TF.)

```
\__kernel_if_debug:TF
\debug_on:n
\debug_off:n
```

(Stubs.)
Some commands were more recently deprecated and not yet removed; only make these into errors if the user requests it. This relies on two token lists, filled up in \l3deprecation.

\cs_set_nopar:Npn \g__debug_deprecation_on_tl { }
\cs_set_nopar:Npn \g__debug_deprecation_off_tl { }
\cs_set_protected:Npn \__kernel_deprecation_code:nn #1#2
\tl_gput_right:Nn \g__debug_deprecation_on_tl {#1}
\tl_gput_right:Nn \g__debug_deprecation_off_tl {#2}

(End definition for \__kernel_deprecation_code:nn, \g__debug_deprecation_on_tl, and \g__debug_deprecation_off_tl.)

41.7 Conditional processing and definitions

\langle @@=prg\rangle

Underneath any predicate function (_p) or other conditional forms (TF, etc.) is a built-in logic saying that it after all of the testing and processing must return the \langle state\rangle this leaves \TeX{} in. Therefore, a simple user interface could be something like
\begin{verbatim}
\if_meaning:w #1#2 \prg_return_true: \else: \if_meaning:w #1#3 \prg_return_true: \else: \prg_return_false: \fi: \fi:
\end{verbatim}

Usually, a \TeX{} programmer would have to insert a number of \exp_after:wN\_s to ensure the state value is returned at exactly the point where the last conditional is finished. However, that obscures the code and forces the \TeX{} programmer to prove that he/she knows the $2^n - 1$ table. We therefore provide the simpler interface.

\begin{verbatim}
\cs_set:Npn \prg_return_true: { \exp_after:wN \use_i:nn \exp:w }
\cs_set:Npn \prg_return_false: { \exp_after:wN \use_ii:nn \exp:w }
\end{verbatim}

The idea here is that \exp:w expands fully any \else: and \fi: that are waiting to be discarded, before reaching the \exp_end: which leaves an empty expansion. The code can then leave either the first or second argument in the input stream. This means that all of the branching code has to contain at least two tokens: see how the logical tests are actually implemented to see this.

\begin{verbatim}
\cs_set:Npn \prg_return_true: { \{ \exp_after:wN \use_i:nn \exp:w \} }
\cs_set:Npn \prg_return_false: { \{ \exp_after:wN \use_ii:nn \exp:w \} }
\end{verbatim}
An extended state space could be implemented by including a more elaborate function in place of `\use_i:nn/\use_ii:nn`. Provided two arguments are absorbed then the code would work.

(\textit{End definition for `\prg_return_true': and `\prg_return_false':. These functions are documented on page 62.})

Private version of `\use_none_delimit_by_q_recursion_stop:w`.

\begin{verbatim}
\cs_set:Npn \__prg_use_none_delimit_by_q_recursion_stop:w
  #1 \q__prg_recursion_stop { }
\end{verbatim}

(\textit{End definition for `\__prg_use_none_delimit_by_q_recursion_stop:w`.)

The user functions for the types using parameter text from the programmer. The various functions only differ by which function is used for the assignment. For those \texttt{Npnn} type functions, we must grab the parameter text, reading everything up to a left brace before continuing. Then split the base function into name and signature, and feed `{⟨name⟩}` `{⟨signature⟩}` `{⟨boolean⟩}` `{⟨set or new⟩}` `{⟨maybe protected⟩}` `{⟨parameters⟩}` `{TF,...}` `{⟨code⟩}` to the auxiliary function responsible for defining all conditionals. Note that \texttt{e} stands for expandable and \texttt{p} for protected.

\begin{verbatim}
\cs_set_protected:Npn \prg_set_conditional:Npnn
  { \__prg_generate_conditional_parm:NNNpnn \cs_set:Npn e }
\cs_set_protected:Npn \prg_new_conditional:Npnn
  { \__prg_generate_conditional_parm:NNNpnn \cs_new:Npn e }
\cs_set_protected:Npn \prg_set_protected_conditional:Npnn
  { \__prg_generate_conditional_parm:NNNpnn \cs_set_protected:Npn p }
\cs_set_protected:Npn \prg_new_protected_conditional:Npnn
  { \__prg_generate_conditional_parm:NNNpnn \cs_new_protected:Npn p }
\cs_set_protected:Npn \__prg_generate_conditional_parm:NNNpnn #1#2#3#4#
  \begin{verbatim}
  \use:x
  \begin{verbatim}
  \__prg_generate_conditional:nnNNNnnn
     \cs_split_function:N #3
     \cs_set:N #1 \q #2 \#3 \#4
  \end{verbatim}
  \end{verbatim}
\end{verbatim}

(\textit{End definition for `\prg_set_conditional:Npnn' and others. These functions are documented on page 60.})

The user functions for the types automatically inserting the correct parameter text based on the signature. The various functions only differ by which function is used for the assignment. Split the base function into name and signature. The second auxiliary generates the parameter text from the number of letters in the signature. Then feed `{⟨name⟩}` `{⟨signature⟩}` `{⟨boolean⟩}` `{⟨set or new⟩}` `{⟨maybe protected⟩}` `{⟨parameters⟩}` `{TF,...}` `{⟨code⟩}` to the auxiliary function responsible for defining all conditionals. If the \texttt{⟨signature⟩} has more than 9 letters, the definition is aborted since \TeX{} macros have at most 9 arguments. The erroneous case where the function name contains no colon is captured later.

\begin{verbatim}
\cs_set_protected:Npn \prg_set_conditional:Nnn
  { \__prg_generate_conditional_count:NNNnn }
\cs_set_protected:Npn \prg_new_conditional:Nnn
  { \__prg_generate_conditional_count:NNNnn }
\cs_set_protected:Npn \prg_set_protected_conditional:Nnn
  { \__prg_generate_conditional_count:NNNnn }
\cs_set_protected:Npn \prg_new_protected_conditional:Nnn
  { \__prg_generate_conditional_count:NNNnn }
\cs_set_protected:Npn \__prg_generate_conditional_count:nnNNNnn
  \begin{verbatim}
  \begin{verbatim}
  \cs_set:N #1 \q #2 \#3 \#4
  \end{verbatim}
\end{verbatim}
\end{verbatim}

357
The workhorse here is going through a list of desired forms, i.e., P, TF, T and F. The first three arguments come from splitting up the base form of the conditional, which gives the name, signature and a boolean to signal whether or not there was a colon in the name. In the absence of a colon, we throw an error and don’t define any conditional. The fourth and fifth arguments build up the defining function. The sixth is the parameters to use (possibly empty), the seventh is the list of forms to define, the eighth is the replacement text which we will augment when defining the forms. The use of \__prg_generate_conditional:nnn makes the later loop more robust.

A large number of our low-level conditionals look like \code \prg_return_true: \else: \prg_return_false: \fi: so we optimize this special case by calling \__prg_generate_conditional_fast:nw \{\code\}. This passes \use_i:nn instead of \use_i_i:nn to functions such as \__prg_generate_p_form:wnnnnnnN.

\cs_set_protected:Nnn \prg_set_protected_conditional:nnn
\__prg_generate_conditional_count:NNnn \cs_set_protected:Nnn p
\__prg_new_protected_conditional:nnn \cs_set_protected:Npn p
\__prg_generate_conditional_count:NNNnn \cs_split_function:N #3
\__kernel_cs_parm_from_arg_count:nnF { \__prg_generate_conditional:nnNnnn #3 #4 #5 }
\msg_error:nxx \token_to_str:c \{ #1 : #2 \}
\use_none:nn

\__prg_generate_conditional:nnNNNnnn #1#2#3#4#5#6#7#8
\if_meaning:w \c_false_bool #3 \msg_error:nxx \token_to_str:c \{ missing-colon \}
\exp_after:wN \use_none:nn \exp_after:wN \exp_not:N \__prg_generate_conditional:nnNnnnNw \exp_not:n \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:N \exp_not:nn

(End definition for \prg_set_conditional:nnn and others. These functions are documented on page 60.)
Loopy through the list of desired forms. First are six arguments and seventh is the form. Use the form to call the correct type. If the form does not exist, the \use:c construction results in \relax, and the error message is displayed (unless the form is empty, to allow for \{T, , F\}), then \use_none:nnnnnnnnn cleans up. Otherwise, the error message is removed by the variant form.

\cs_set_protected:Npn \__prg_generate_conditional:NNnnnnNw #1#2#3#4#5#6#7#8 ,
{\if_meaning:w \q__prg_recursion_tail #8
\exp_after:wN \__prg_use_none_delimit_by_q_recursion_stop:w
\fi:
\use:c { \__prg_generate_ #8 _form:wNNnnnnN }
\tl_if_empty:nF {#8}
{ \msg_error:nnxx
 { kernel } { conditional-form-unknown }
 {#8} { \token_to_str:c { #3 : #4 } } }
\use_none:nnnnnnnn
\s__prg_stop
#1 #2 {#3} {#4} {#5} {#6} #7
\__prgGenerateConditional:NNnnnnnnn #1 #2 {#3} {#4} {#5} {#6} #7
}

(End definition for \__prgGenerateConditional:nnNNnnnn and others.)

\__prg Generate_p_form:wNNnnnnnn
\__prg Generate_TF_form:wNNnnnnnn
\__prg Generate_f_form:wNNnnnnnn
\__prg Generate_P_form:wNNnnnnnn
\__prg P_true:w
\__prg T_true:w
\__prg F_true:w
\__prg TF_true:w

How to generate the various forms. Those functions take the following arguments: 1: junk, 2: \cs_set:Npn or similar, 3: p (for protected conditionals) or e, 4: function name, 5: signature, 6: parameter text, 7: replacement (possibly trimmed by \__prgGenerateConditional:fast:nw), 8: \use_i:ii:nnn or \use_i:nn (for "fast" conditionals). Remember that the logic-returning functions expect two arguments to be present after \exp_end:. notice the construction of the different variants relies on this, and that the TF and F variants will be slightly faster than the T version. The p form is only valid for expandable tests, we check for that by making sure that the second argument is empty. For "fast" conditionals, #7 has an extra \if___. To optimize a bit further we don't use \exp_after:wN \use_i:ii:nnn and similar but instead use \__prg TF true:w and similar to swap out the macro after \fi:. It would be a tiny bit faster if we directly grabbed the T and F arguments there, but if those are actually missing, the recovery from the runaway argument would not insert \fi: back, messing up nesting of conditionals.
The setting-equal functions. Split both functions and feed \( \langle \text{name1} \rangle \) \( \langle \text{signature1} \rangle \) \( \langle \text{boolean1} \rangle \) \( \langle \text{name2} \rangle \) \( \langle \text{signature2} \rangle \) \( \langle \text{boolean2} \rangle \) \( \langle \text{copying function} \rangle \) \( \langle \text{conditions} \rangle \), \q__prg_recursion_tail , \q__prg_recursion_stop to a first auxiliary.
\cs_set_protected:Npn \prg_set_eq_conditional:NNn { \__prg_set_eq_conditional:NNNn \cs_new_eq:cc }
\cs_set_protected:Npn \prg_new_eq_conditional:NNn { \__prg_set_eq_conditional:NNNn \cs_new_eq:cc }
\cs_set_protected:Npn \__prg_set_eq_conditional:NNNn #1#2#3#4
{ \use:x
{ \exp_not:N \__prg_set_eq_conditional:nnNnnNNw \cs_split_function:N #2 \cs_split_function:N #3 \exp_not:N #1 \tl_to_str:n {#4} \exp_not:n { , \q__prg_recursion_tail , \q__prg_recursion_stop } }
}

(End definition for \prg_set_eq_conditional:NNn, \prg_new_eq_conditional:NNn, and \__prg_set_eq_conditional:NNNn. These functions are documented on page 62.)


Split the function to be defined, and setup a manual clist loop over argument #6 of the first auxiliary. The second auxiliary receives twice three arguments coming from splitting the function to be defined and the function to copy. Make sure that both functions contained a colon, otherwise we don’t know how to build conditionals, hence abort. Call the looping macro, with arguments \{⟨name⟩\} \{⟨signature⟩\} \{⟨name⟩\} \{⟨signature⟩\} ⟨copying function⟩ and followed by the comma list. At each step in the loop, make sure that the conditional form we copy is defined, and copy it, otherwise abort.

\cs_set_protected:Npn \__prg_set_eq_conditional_loop:nnnnNw #1#2#3#4#5#6 ,
{ \if_meaning:w \c_false_bool #3 \msg_error:nnx { kernel } { missing-colon } \{ \token_to_str:c {#1} \} \exp_after:wN \__prg_use_none_delimit_by_q_recursion_stop:w \fi:
\if_meaning:w \c_false_bool #6 \msg_error:nnx { kernel } { missing-colon } \{ \token_to_str:c {#4} \} \exp_after:wN \__prg_use_none_delimit_by_q_recursion_stop:w \fi:
\__prg_set_eq_conditional_loop:nnnnNw \{#1\} \{#2\} \{#4\} \{#5\}
}
\cs_set_protected:Npn \__prg_set_eq_conditional_loop:nnnnNw #1#2#3#4#5#6 ,
{ \if_meaning:w \q__prg_recursion_tail #6 \exp_after:wN \__prg_use_none_delimit_by_q_recursion_stop:w \fi:
\use:c { \__prg_set_eq_conditional_ #6 _form:wNnnnNw \tl_if_empty:nF \{#6\} \{ \token_to_str:c \{ #1 : #2 \} \} }
\use_none:nnnnnn \s__prg_stop
All that is left is to define the canonical boolean true and false. I think Michael originated the idea of expandable boolean tests. At first these were supposed to expand into either \texttt{TT} or \texttt{TF} to be tested using \texttt{\if:w} but this was later changed to \texttt{00} and \texttt{01}, so they could be used in logical operations. Later again they were changed to being numerical constants with values of 1 for true and 0 for false. We need this from the get-go.\

\texttt{\c_true_bool} \texttt{\c_false_bool} Here are the canonical boolean values.\

\begin{verbatim}
1771 \tex_chardef:D \c_true_bool = 1 -
1772 \tex_chardef:D \c_false_bool = 0 -
\end{verbatim}

(End definition for \texttt{\c_true_bool} and \texttt{\c_false_bool}. These variables are documented on page 26.)

41.8 Dissecting a control sequence

\begin{verbatim}
\__cs_count_signature:N \__cs_count_signature:N \langle function \rangle
Splits the \langle function \rangle into the \langle name \rangle (i.e. the part before the colon) and the \langle signature \rangle (i.e. after the colon). The \langle number \rangle of tokens in the \langle signature \rangle is then left in the input stream. If there was no \langle signature \rangle then the result is the marker value $-1$.
\end{verbatim}

\begin{verbatim}
\__cs_get_function_name:N \__cs_get_function_name:N \langle function \rangle
Splits the \langle function \rangle into the \langle name \rangle (i.e. the part before the colon) and the \langle signature \rangle (i.e. after the colon). The \langle name \rangle is then left in the input stream without the escape character present made up of tokens with category code 12 (other).
\end{verbatim}

\begin{verbatim}
\__cs_get_function_signature:N \__cs_get_function_signature:N \langle function \rangle
Splits the \langle function \rangle into the \langle name \rangle (i.e. the part before the colon) and the \langle signature \rangle (i.e. after the colon). The \langle signature \rangle is then left in the input stream made up of tokens with category code 12 (other).
\end{verbatim}

\begin{verbatim}
\__cs_tmp:w Function used for various short-term usages, for instance defining functions whose definition involves tokens which are hard to insert normally (spaces, characters with category other).
\end{verbatim}
This converts a control sequence into the character string of its name, removing the leading escape character. This turns out to be a non-trivial matter as there are different cases:

- The usual case of a printable escape character;
- the case of a non-printable escape characters, e.g., when the value of the \escapechar is negative;
- when the escape character is a space.

One approach to solve this is to test how many tokens result from \token_to_str:N \a. If there are two tokens, then the escape character is printable, while if it is non-printable then only one is present.

However, there is an additional complication: the control sequence itself may start with a space. Clearly that should not be lost in the process of converting to a string. So the approach adopted is a little more intricate still. When the escape character is printable, \token_to_str:N \a yields the escape character itself and a space. The character codes are different, thus the \if:w test is false, and \TeX reads \cs_to_str:w \a after turning the following control sequence into a string: this auxiliary removes the escape character, and stops the expansion of the initial \tex_romannumeral:D. The second case is that the escape character is not printable. Then the \if:w test is unfinished after reading a the space from \token_to_str:N \a, and the auxiliary \cs_to_str:w is expanded, feeding - as a second character for the test; the test is false, and \TeX skips to \fi:, then performs \token_to_str:N, and stops the \tex_romannumeral:D with \c_zero_int. The last case is that the escape character is itself a space. In this case, the \if:w test is true, and the auxiliary \cs_to_str:w comes into play, inserting \c_zero_int to the character 0. The initial \tex_romannumeral:D then sees 0, which is not a terminated number, followed by the escape character, a space, which is removed, terminating the expansion of \tex_romannumeral:D. In all three cases, \cs_to_str:N takes two expansion steps to be fully expanded.

We implement the expansion scheme using \tex_romannumeral:D terminating it with \c_zero_int rather than using \exp:w and \exp_end: as we normally do. The reason is that the code heavily depends on terminating the expansion with \c_zero_int so we make this dependency explicit.

If speed is a concern we could use \csstring in \TeX. For the empty csname that primitive gives an empty result while the current \cs_to_str:N gives incorrect results in all engines (this is impossible to fix without huge performance hit).

(End definition for \cs_to_str:N, \cs_to_str:w, and \cs_to_str:w. This function is documented on page 21.)
This function takes a function name and splits it into name with the escape char removed and argument specification. In addition to this, a third argument, a boolean \( \langle \text{true} \rangle \) or \( \langle \text{false} \rangle \) is returned with \( \langle \text{true} \rangle \) for when there is a colon in the function and \( \langle \text{false} \rangle \) if there is not.

We cannot use : directly as it has the wrong category code so an x-type expansion is used to force the conversion.

First ensure that we actually get a properly evaluated string by expanding \cs_to_str:N twice. If the function contained a colon, the auxiliary takes as \#1 the function name, delimited by the first colon, then the signature \#2, delimited by \s__cs_mark, then \c_true_bool as \#3, and \#4 cleans up until \s__cs_stop. Otherwise, the \#1 contains the function name and \s__cs_mark \c_true_bool \#2 is empty, \#3 is \c_false_bool, and \#4 cleans up. The second auxiliary trims the trailing \s__cs_mark from the function name if present (that is, if the original function had no colon).

\begin{verbatim}
\exp_after:wN \exp_after:wN \exp_after:wN \__cs_split_function_auxi:w
\cs_to_str:N ##1 \s__cs_mark \c_true_bool
#1 \s__cs_mark \c_false_bool \s__cs_stop
\}
\cs_set:Npn \__cs_split_function_auxi:w
##1 #1 ##2 \s__cs_mark ##3##4 \s__cs_stop
{ \__cs_split_function_auxii:w ##1 \s__cs_mark \s__cs_stop {##2} ##3 }
\cs_set:Npn \__cs_split_function_auxii:w
##1 \s__cs_mark ##2 \s__cs_stop
{ {##1} }
\}
\exp_after:wN \__cs_tmp:w \token_to_str:N :
\end{verbatim}

(End definition for \cs_split_function:N, \__cs_split_function_auxi:w, and \__cs_split_function_auxii:w. This function is documented on page 22.)

### 41.9 Exist or free

A control sequence is said to exist (to be used) if has an entry in the hash table and its meaning is different from the primitive \relax token. A control sequence is said to be free (to be defined) if it does not already exist.

Two versions for checking existence. For the \N form we firstly check for \scan_stop: and then if it is in the hash table. There is no problem when inputting something like \else: or \fi: as \TeX will only ever skip input in case the token tested against is \scan_stop:.

\begin{verbatim}
\exp_after:wN \__cs_tmp:w \token_to_str:N :
\end{verbatim}
For the c form we firstly check if it is in the hash table and then for \texttt{\textbackslash scan\_stop}: so that we do not add it to the hash table unless it was already there. Here we have to be careful as the text to be skipped if the first test is false may contain tokens that disturb the scanner. Therefore, we ensure that the second test is performed after the first one has concluded completely.

\begin{verbatim}
\prg_set_conditional:Npnn \cs_if_exist:c #1 { p , T , F , TF }
{ \if_cs_exist:w #1 \cs_end:
  \exp_after:wN \use_i:nn
  \else:
  \exp_after:wN \use_i:nn
  \fi:
  { \exp_after:wN \if_meaning:w \cs:w #1 \cs_end: \scan_stop:
    \prg_return_false:
    \else:
    \prg_return_true:
    \fi:
  }
  \prg_return_false:
\}
\end{verbatim}

(End definition for \texttt{\textbackslash cs\_if\_exist:NTF}. This function is documented on page 27.)

\texttt{\textbackslash cs\_if\_free:p:N} The logical reversal of the above.
\begin{verbatim}
\prg_set_conditional:Npnn \cs_if_free_p:N #1 { p , T , F , TF }
{ \if_meaning:w #1 \scan_stop:
  \prg_return_true:
  \else:
  \if_cs_exist:N #1
  \prg_return_false:
  \else:
  \prg_return_true:
  \fi:
  \fi:
\}
\end{verbatim}

\texttt{\textbackslash cs\_if\_free:p:c} The logical reversal of the above.
\begin{verbatim}
\prg_set_conditional:Npnn \cs_if_free_p:c #1 { p , T , F , TF }
{ \if_meaning:w #1 \scan_stop:
  \prg_return_true:
  \else:
  \if_cs_exist:N #1
  \prg_return_false:
  \else:
  \prg_return_true:
  \fi:
  \fi:
\}
\end{verbatim}

\texttt{\textbackslash cs\_if\_free:p:NTF} The logical reversal of the above.
\begin{verbatim}
\prg_set_conditional:Npnn \cs_if_free_p:NTF #1 { p , T , F , TF }
{ \if_meaning:w #1 \scan_stop:
  \prg_return_true:
  \else:
  \if_cs_exist:N #1
  \prg_return_false:
  \else:
  \prg_return_true:
  \fi:
  \fi:
\}
\end{verbatim}

\texttt{\textbackslash cs\_if\_free:c:TF} The logical reversal of the above.
\begin{verbatim}
\prg_set_conditional:Npnn \cs_if_free:c:TF #1 { p , T , F , TF }
{ \if_meaning:w #1 \scan_stop:
  \prg_return_true:
  \else:
  \if_cs_exist:N #1
  \prg_return_false:
  \else:
  \prg_return_true:
  \fi:
  \fi:
\}
\end{verbatim}

(End definition for \texttt{\textbackslash cs\_if\_free:c:NTF}. This function is documented on page 27.)
\cs_if_exist_use:N
\cs_if_exist_use:c
\cs_if_exist_use:NTF
\cs_if_exist_use:cTF

The \texttt{\cs_if_exist_use:...} functions cannot be implemented as conditionals because the true branch must leave both the control sequence itself and the true code in the input stream. For the \texttt{c} variants, we are careful not to put the control sequence in the hash table if it does not exist. In Lua\TeX{} we could use the \texttt{\lastnamedcs} primitive.

\begin{verbatim}
\cs_set:Npn \cs_if_exist_use:NTF #1#2 { \cs_if_exist:NTF #1 { #1 #2 } }
\cs_set:Npn \cs_if_exist_use:NF #1 { \cs_if_exist:NTF #1 { #1 } }
\cs_set:Npn \cs_if_exist_use:NT #1 #2 { \cs_if_exist:NTF #1 { #1 #2 } { } }
\cs_set:Npn \cs_if_exist_use:N #1 { \cs_if_exist:NTF #1 { #1 } { } }
\cs_set:Npn \cs_if_exist_use:cTF #1#2 { \cs_if_exist:cTF {#1} { \use:c {#1} #2 } }
\cs_set:Npn \cs_if_exist_use:cF #1 { \cs_if_exist:cTF {#1} { \use:c {#1} } }
\cs_set:Npn \cs_if_exist_use:cT #1#2 { \cs_if_exist:cTF {#1} { \use:c {#1} #2 } { } }
\cs_set:Npn \cs_if_exist_use:c #1 { \cs_if_exist:cTF {#1} { \use:c {#1} } { } }
\end{verbatim}

(End definition for \texttt{\cs_if_exist_use:NTF}. This function is documented on page 21.)

41.10 Preliminaries for new functions

We provide two kinds of functions that can be used to define control sequences. On the one hand we have functions that check if their argument doesn’t already exist, they are called \texttt{\ldots_new}. The second type of defining functions doesn’t check if the argument is already defined.

Before we can define them, we need some auxiliary macros that allow us to generate error messages. The next few definitions here are only temporary, they will be redefined later on.

\begin{verbatim}
\msg_error:nnxx
\msg_error:nnx
\msg_error:nn
\end{verbatim}

If an internal error occurs before \LaTeX{} has loaded \texttt{l3msg} then the code should issue a usable if terse error message and halt. This can only happen if a coding error is made by the team, so this is a reasonable response. Setting the \texttt{\newlinechar} is needed, to turn \texttt{\^{}\^{}J} into a proper line break in plain \TeX{}.

\begin{verbatim}
\cs_set_protected:Npn \msg_error:nnxx #1#2#3#4
{ \tex_newlinechar:D = '\\J \scan_stop:
\texErrMsg:D
{ !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! \^{}\^{}J
Argh,-internal-\LaTeX{}-error! \^{}\^{}J \^{}\^{}J
Module - #1 , - message-name="#2": \^{}\^{}J
Arguments-"#3'-and-'#4' \^{}\^{}J \^{}\^{}J
This-is-one-for-The-\LaTeX{}-Project:-bailing-out
}
\end{verbatim}

366
Another one from l3msg which will be altered later.
\msg_line_context: \cs_set:Npn \msg_line_context:
\{ on-line- \tex_the:D \tex_inputlineno:D \}
\iow_log:x \iow_term:x
We define a routine to write only to the log file. And a similar one for writing to both
the log file and the terminal. These will be redefined later by l3io.
\__kernel_chk_if_free_cs:N \__kernel_chk_if_free_cs:c
This command is called by \cs_new_nopar:Npn and \cs_new_eq:NN etc. to make sure
that the argument sequence is not already in use. If it is, an error is signalled. It checks
if (csname) is undefined or \scan_stop:. Otherwise an error message is issued. We have
to make sure we don’t put the argument into the conditional processing since it may be
an \if... type function!
\__kernel_chk_if_free_cs:N \__kernel_chk_if_free_cs:c
\cs_set:Npn \__cs_tmp:w #1#2
\{ \cs_set_protected:Npn \msg_error:nnx \#1 \#2 \{ \}
\cs_set_protected:Npn \msg_error:nnx \#1 \#2 \{ \}
\}
(End definition for \msg_error:nn, \msg_error:nnx, and \msg_error:nn. These functions are docu-
mented on page ??.)
\msg_line_context: Another one from l3msg which will be altered later.
\msg_line_context: \cs_set:Npn \msg_line_context:
\{ on-line- \tex_the:D \tex_inputlineno:D \}
\iow_log:x \iow_term:x
We define a routine to write only to the log file. And a similar one for writing to both
the log file and the terminal. These will be redefined later by l3io.
\__kernel_chk_if_free_cs:N \__kernel_chk_if_free_cs:c
This command is called by \cs_new_nopar:Npn and \cs_new_eq:NN etc. to make sure
that the argument sequence is not already in use. If it is, an error is signalled. It checks
if (csname) is undefined or \scan_stop:. Otherwise an error message is issued. We have
to make sure we don’t put the argument into the conditional processing since it may be
an \if... type function!
\__kernel_chk_if_free_cs:N \__kernel_chk_if_free_cs:c
\cs_set:Npn \__cs_tmp:w #1#2
\{ \cs_set_protected:Npn #1 ##1
\{ \__kernel_chk_if_free_cs:N ##1
\}
\}
\}
\}
\cs_set_protected:Npn \__kernel_chk_if_free_cs:c
\{ \exp_args:Nc \__kernel_chk_if_free_cs:N \}
(End definition for \__kernel_chk_if_free_cs:N)

41.11 Defining new functions
\cs_new_nopar:Npn \cs_new_nopar:Npx \cs_new:Npn \cs_new:Npx \cs_new_protected_nopar:Npn \cs_new_protected_nopar:Npx \cs_new_protected:Npn \cs_new_protected:Npx
\__cs_tmp:w
\cs_set:Npn \__cs_tmp:w #1#2
\{ \cs_set_protected:Npn \#1 \#2
\{ \}
\}
\}
\}
\}
\cs_set_protected:Npn \__kernel_chk_if_free_cs:N \#1
\}
\}
\}
\}
\cs_set_protected:Npn \__kernel_chk_if_free_cs:c
\{ \exp_args:Nc \__kernel_chk_if_free_cs:N \}
(End definition for \__kernel_chk_if_free_cs:N)
Like \cs_set_nopar:Npn and \cs_new_nopar:Npn, except that the first argument consists of the sequence of characters that should be used to form the name of the desired control sequence (the c stands for csname argument, see the expansion module). Global versions are also provided.

\cs_set:cpn \langle string \rangle \langle rep-text \rangle turns \langle string \rangle into a csname and then assigns \langle rep-text \rangle to it by using \cs_set_nopar:Npn. This means that there might be a parameter string between the two arguments.

Variants of the \cs_set:Npn versions which make a csname out of the first arguments. We may also do this globally.

Variants of the \cs_set_protected_nopar:Npn versions which make a csname out of the first arguments. We may also do this globally.
Variants of the \cs_set_protected:Npn versions which make a csname out of the first arguments. We may also do this globally.

\__cs_tmp:w \cs_set_protected:cpn \cs_set_protected:Npn
\__cs_tmp:w \cs_set_protected:cpx \cs_set_protected:Npx
\__cs_tmp:w \cs_gset_protected:cpn \cs_gset_protected:Npn
\__cs_tmp:w \cs_gset_protected:cpx \cs_gset_protected:Npx
\__cs_tmp:w \cs_new_protected:cpn \cs_new_protected:Npn
\__cs_tmp:w \cs_new_protected:cpx \cs_new_protected:Npx
\__cs_tmp:w \cs_gset_protected:cpn \cs_gset_protected:Npn
\__cs_tmp:w \cs_gset_protected:cpx \cs_gset_protected:Npx
\__cs_tmp:w \cs_new_protected:cpn \cs_new_protected:Npn
\__cs_tmp:w \cs_new_protected:cpx \cs_new_protected:Npx

(End definition for \cs_set_protected:Npn. This function is documented on page 16.)

41.12 Copying definitions

\cs_set_eq:NN \cs_set_eq:cN \cs_set_eq:Nc \cs_set_eq:cc
\cs_gset_eq:NN \cs_gset_eq:cN \cs_gset_eq:Nc \cs_gset_eq:cc
\cs_new_eq:NN \cs_new_eq:cN \cs_new_eq:Nc \cs_new_eq:cc

These macros allow us to copy the definition of a control sequence to another control sequence.

The = sign allows us to define funny char tokens like = itself or \par with this function.

For the definition of \c_space_char{~} to work we need the ~ after the =.

\cs_set_eq:NN is long to avoid problems with a literal argument of \par. While
\cs_new_eq:NN will probably never be correct with a first argument of \par, define it long in order to throw an “already defined” error rather than “runaway argument”.

\cs_new_protected:Npn \cs_set_eq:NN #1 { \tex_let:D #1 =~ }
\cs_new_protected:Npn \cs_set_eq:cN { \exp_args:Nc \cs_set_eq:NN }
\cs_new_protected:Npn \cs_set_eq:Nc { \exp_args:NNc \cs_set_eq:NN }
\cs_new_protected:Npn \cs_set_eq:cc { \exp_args:Ncc \cs_set_eq:NN }
\cs_new_protected:Npn \cs_gset_eq:NN { \tex_global:D \cs_set_eq:NN }
\cs_new_protected:Npn \cs_gset_eq:Nc { \exp_args:NNc \cs_gset_eq:NN }
\cs_new_protected:Npn \cs_gset_eq:cN { \exp_args:Nc \cs_gset_eq:NN }
\cs_new_protected:Npn \cs_gset_eq:cc { \exp_args:Ncc \cs_gset_eq:NN }
\cs_new_protected:Npn \cs_new_eq:NN #1 { \__kernel_chk_if_free_cs:N #1 \tex_global:D \cs_set_eq:NN #1 }
\cs_new_protected:Npn \cs_new_eq:cN { \exp_args:Nc \cs_new_eq:NN }
\cs_new_protected:Npn \cs_new_eq:Nc { \exp_args:NNc \cs_new_eq:NN }
\cs_new_protected:Npn \cs_new_eq:cc { \exp_args:Ncc \cs_new_eq:NN }

(End definition for \cs_set_eq:NN, \cs_gset_eq:NN, and \cs_new_eq:NN. These functions are documented on page 19.)

41.13 Undefining functions

\cs_undefine:N \cs_undefine:c

The following function is used to free the main memory from the definition of some function that isn’t in use any longer. The c variant is careful not to add the control sequence to the hash table if it isn’t there yet, and it also avoids nesting \TeX conditionals in case #1 is unbalanced in this matter.

\cs_new_protected:Npn \cs_undefine:N #1 { \tex_undefined:D } \cs_new_protected:Npn \cs_undefine:c #1 { \if_cs_exist:w #1 \cs_end: }

369
41.14 Generating parameter text from argument count

\TeX3 provides shorthands to define control sequences and conditionals with a simple parameter text, derived directly from the signature, or more generally from knowing the number of arguments, between 0 and 9. This function expands to its first argument, untouched, followed by a brace group containing the parameter text \{#1\ldots#n\}, where \(n\) is the result of evaluating the second argument (as described in \texttt{\int_{eval}:n}). If the second argument gives a result outside the range \([0,9]\), the third argument is returned instead, normally an error message. Some of the functions use here are not defined yet, but will be defined before this function is called.

\begin{verbatim}
\cs_set_protected:Npn \__kernel_cs_parm_from_arg_count:nnF #1#2 #3
\exp_args:Nx \__cs_parm_from_arg_count_test:nnF #1#2 #3
\exp_after:wN \exp_not:n \if_case:w \int_eval:n {#2} { } \or: { ##1 } \or: { ##1##2 } \or: { ##1##2##3 } \or: { ##1##2##3##4 } \or: { ##1##2##3##4##5 } \or: { ##1##2##3##4##5##6 } \or: { ##1##2##3##4##5##6##7 } \or: { ##1##2##3##4##5##6##7##8 } \or: { ##1##2##3##4##5##6##7##8##9 } \else: { \c_false_bool } \fi:
{#1}
\end{verbatim}

(End definition for \texttt{\cs_{undefine}:N}. This function is documented on page 20.)
41.15 Defining functions from a given number of arguments

Counting the number of tokens in the signature, i.e., the number of arguments the function
should take. Since this is not used in any time-critical function, we simply use \tl_count:n if there is a signature, otherwise −1 arguments to signal an error. We need
a variant form right away.

\cs_new:Npn \__cs_count_signature:N #1
{ \exp_args:Nf \__cs_count_signature:n { \cs_split_function:N #1 } }
\cs_new:Npn \__cs_count_signature:n #1
{ \int_eval:n { \__cs_count_signature:nnN #1 } }
\cs_new:Npn \__cs_count_signature:nnN #1#2#3
{ \if_meaning:w \c_true_bool #3 \tl_count:n {#2} \else: -1 \fi: }
\cs_new:Npn \__cs_count_signature:c
{ \exp_args:Nc \__cs_count_signature:N }

We provide a constructor function for defining functions with a given number of arguments. For this we need to choose the correct parameter text and then use that when defining. Since \TeX supports from zero to nine arguments, we use a simple switch to choose the correct parameter text, ensuring the result is returned after finishing the conditional. If it is not between zero and nine, we throw an error.

1: function to define, 2: with what to define it, 3: the number of args it requires and
4: the replacement text

\cs_new_protected:Npn \cs_generate_from_arg_count:NNnn #1#2#3#4
{ \__kernel_cs_parm_from_arg_count:nnF { \use:nnn #2 #1 } {#3} {#4} }
\cs_new_protected:Npn \cs_generate_from_arg_count:cNnn
{ \exp_args:Nc \cs_generate_from_arg_count:NNnn }
\cs_new_protected:Npn \cs_generate_from_arg_count:Ncnn
{ \exp_args:NNc \cs_generate_from_arg_count:NNnn }

A variant form we need right away, plus one which is used elsewhere but which is most
logically created here.

\cs_new_protected:Npn \cs_generate_from_arg_count:NNnn #1#2#3#4
{ \__kernel_cs_parm_from_arg_count:nnF { \use:nnn #2 #1 } {#3} {#4} }
\cs_new_protected:Npn \cs_generate_from_arg_count:cNnn
{ \exp_args:Nc \cs_generate_from_arg_count:NNnn }
\cs_new_protected:Npn \cs_generate_from_arg_count:Ncnn
{ \exp_args:NNc \cs_generate_from_arg_count:NNnn }

(End definition for \__cs_count_signature:N, \__cs_count_signature:n, and \__cs_count_signature:nnN.)

(End definition for \cs_generate_from_arg_count:NNnn. This function is documented on page 19.)
41.16
2034

Using the signature to define functions

⟨@@=cs⟩

We can now combine some of the tools we have to provide a simple interface for
defining functions, where the number of arguments is read from the signature. For
instance, \cs_set:Nn \foo_bar:nn {#1,#2}.
\cs_set:Nn We want to define \cs_set:Nn as
\cs_set:Nx
\cs_set_nopar:Nn
\cs_set_nopar:Nx
\cs_set_protected:Nn
\cs_set_protected:Nx
\cs_set_protected_nopar:Nn
\cs_set_protected_nopar:Nx
\cs_gset:Nn
\cs_gset:Nx
\cs_gset_nopar:Nn
\cs_gset_nopar:Nx
\cs_gset_protected:Nn
\cs_gset_protected:Nx
\cs_gset_protected_nopar:Nn
\cs_gset_protected_nopar:Nx
\cs_new:Nn
\cs_new:Nx
\cs_new_nopar:Nn
\cs_new_nopar:Nx
\cs_new_protected:Nn
\cs_new_protected:Nx
\cs_new_protected_nopar:Nn
\cs_new_protected_nopar:Nx

\cs_set_protected:Npn \cs_set:Nn #1#2
{
\cs_generate_from_arg_count:NNnn #1 \cs_set:Npn
{ \@@_count_signature:N #1 } {#2}
}
In short, to define \cs_set:Nn we need just use \cs_set:Npn, everything else is the same
for each variant. Therefore, we can make it simpler by temporarily defining a function
to do this for us.
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070

\cs_set:Npn \__cs_tmp:w #1#2#3
{
\cs_new_protected:cpx { cs_ #1 : #2 }
{
\exp_not:N \__cs_generate_from_signature:NNn
\exp_after:wN \exp_not:N \cs:w cs_ #1 : #3 \cs_end:
}
}
\cs_new_protected:Npn \__cs_generate_from_signature:NNn #1#2
{
\use:x
{
\__cs_generate_from_signature:nnNNNn
\cs_split_function:N #2
}
#1 #2
}
\cs_new_protected:Npn \__cs_generate_from_signature:nnNNNn #1#2#3#4#5#6
{
\bool_if:NTF #3
{
\cs_set_nopar:Npx \__cs_tmp:w
{ \tl_map_function:nN {#2} \__cs_generate_from_signature:n }
\tl_if_empty:oF \__cs_tmp:w
{
\msg_error:nnxxx { kernel } { non-base-function }
{ \token_to_str:N #5 } {#2} { \__cs_tmp:w }
}
\cs_generate_from_arg_count:NNnn
#5 #4 { \tl_count:n {#2} } {#6}
}
{
\msg_error:nnx { kernel } { missing-colon }
{ \token_to_str:N #5 }
}
}

372


Then we define the 24 variants beginning with \texttt{N}.

\begin{verbatim}
\_\_cs_tmp:w { set } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { set } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { set_nopar } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { set_nopar } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { set_protected } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { set_protected } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { set_protected_nopar } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { set_protected_nopar } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { gset } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { gset } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { gset_nopar } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { gset_nopar } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { gset_protected } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { gset_protected } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { gset_protected_nopar } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { gset_protected_nopar } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { new } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { new } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { new_nopar } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { new_nopar } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { new_protected } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { new_protected } \{ Nx \} \{ Npx \}
\_\_cs_tmp:w { new_protected_nopar } \{ Nn \} \{ Npn \}
\_\_cs_tmp:w { new_protected_nopar } \{ Nx \} \{ Npx \}
\end{verbatim}

(End definition for \texttt{\cs_set:Nn} and others. These functions are documented on page 17.)

The 24 \texttt{c} variants simply use \texttt{\exp_args:Nc}.

\begin{verbatim}
\_\_cs_tmp:w { set } \{ n \}
\_\_cs_tmp:w { set } \{ x \}
\_\_cs_tmp:w { set_nopar } \{ n \}
\_\_cs_tmp:w { set_nopar } \{ x \}
\_\_cs_tmp:w { set_protected } \{ n \}
\_\_cs_tmp:w { set_protected } \{ x \}
\_\_cs_tmp:w { set_protected_nopar } \{ n \}
\_\_cs_tmp:w { set_protected_nopar } \{ x \}
\_\_cs_tmp:w { gset } \{ n \}
\_\_cs_tmp:w { gset } \{ x \}
\_\_cs_tmp:w { gset_nopar } \{ n \}
\_\_cs_tmp:w { gset_nopar } \{ x \}
\_\_cs_tmp:w { gset_nopar } \{ n \}
\_\_cs_tmp:w { gset_nopar } \{ x \}
\_\_cs_tmp:w { gset_nopar } \{ n \}
\_\_cs_tmp:w { gset_nopar } \{ x \}
\end{verbatim}

41.17 Checking control sequence equality

Check if two control sequences are identical.

\cs_if_eq_p:NN
\cs_if_eq_p:cN
\cs_if_eq_p:cc
\cs_if_eq:NNTF
\cs_if_eq:NNTF
\cs_if_eq:NcTF
\cs_if_eq:NcF
\cs_if_eq_p:cc

(End definition for \cs_if_eq:NNTF. This function is documented on page 27.)

41.18 Diagnostic functions

\__kernel_chk_defined:NT Error if the variable #1 is not defined.

\cs_new_protected:Npm \__kernel_chk_defined:NT #1#2

(End definition for \cs_set:Nn. This function is documented on page 17.)
Simply using the `\show` primitive does not allow for line-wrapping, so instead use `\tl_show:n` and `\tl_log:n` (defined in `l3tl` and that performs line-wrapping). This displays \textless (variable)\textgreater = (value). We expand the value before-hand as otherwise some integers (such as `\currentgrouplevel` or `\currentgrouptype`) altered by the line-wrapping code would show wrong values.

\begin{verbatim}
cs_new_protected:Npn \_kernel_register_show:N { \_kernel_register_show_aux:NN \tl_show:n }
cs_new_protected:Npn \_kernel_register_show:c { \exp_args:Nc \_kernel_register_show:N }
cs_new_protected:Npn \_kernel_register_log:N { \_kernel_register_show_aux:NN \tl_log:n }
cs_new_protected:Npn \_kernel_register_log:c { \exp_args:Nc \_kernel_register_log:N }
cs_new_protected:Npn \_kernel_register_show_aux:NN #1#2 { \__kernel_chk_defined:NT #2 { \exp_args:No \_kernel_register_show:nNN { \tex_the:D #2 } #2 #1 } }
cs_new_protected:Npn \_kernel_register_show_aux:nNN #1#2#3 { \exp_args:No #3 { \token_to_str:N #2 = \cs_meaning:N #2 } }
\end{verbatim}

Some control sequences have a very long name or meaning. Thus, simply using \TeX’s primitive `\show` could lead to overlong lines. The output of this primitive is mimicked to some extent, then the re-built string is given to `\tl_show:n` or `\tl_log:n` for line-wrapping. We must expand the meaning before passing it to the wrapping code as otherwise we would wrongly see the definitions that are in place there. To get correct escape characters, set the `\escapechar` in a group; this also localizes the assignment performed by \texttt{x}\-expansion. The `\cs_show:c` and `\cs_log:c` commands convert their argument to a control sequence within a group to avoid showing \texttt{\relax} for undefined control sequences.

\begin{verbatim}
cs_new_protected:Npn \cs_show:N { \_kernel_register_show:N \tl_show:n }
cs_new_protected:Npn \cs_show:c { \group_begin: \exp_args:NNc \group_end: \cs_show:N }
cs_new_protected:Npn \cs_log:N { \_kernel_register_show:N \tl_log:n }
cs_new_protected:Npn \cs_log:c { \group_begin: \exp_args:NNc \group_end: \cs_log:N }
cs_new_protected:Npn \_kernel_register_show:NN #1#2 { \group_begin: \int_set:Nn \tex_escapechar:D { '\\ } \exp_args:NNx \group_end: #1 \{ \token_to_str:N #2 = \cs_meaning:N #2 \} }
\end{verbatim}

(End definition for \_kernel\_chk\_defined:NT.)
 Wrapper around \showgroups. Getting \TeX{} to write to the log without interruption the
run is done by altering the interaction mode.

\begin{Verbatim}
\cs_new_protected:Npn \group_show_list:
\{ \__kernel_group_show:NN \use_none:n 1 \}
\cs_new_protected:Npn \group_log_list:
\{ \__kernel_group_show:NN \int_zero:N 0 \}
\cs_new_protected:Npn \__kernel_group_show:NN #1#2
\{ \use:x \{ \__kernel_prefix_arg_replacement:wN #1 \}
\tl_to_str:n \{ macro : \} \exp_not:n \{ #2 -> #3 \}
\exp_after:wN \scan_stop: \\}
\cs_new:Npn \cs_prefix_spec:N #1
\{ \token_if_macro:NTF #1 \exp_after:wN \__kernel_prefix_arg_replacement:wN \token_to_meaning:N \use_i:nnn \}
\cs_new:Npn \cs_argument_spec:N #1
\{ \token_if_macro:NTF #1 \exp_after:wN \__kernel_prefix_arg_replacement:wN \}
\cs_new:Npn \cs_replacement_spec:N #1
\{ \token_if_macro:NTF #1 \exp_after:wN \__kernel_prefix_arg_replacement:wN \}
\cs_new:Npn \__kernel_prefix_arg_replacement:wN #1
\{ \use:x \{ #1 \} \}
\end{Verbatim}

(End definition for \group_show_list:, \group_log_list:, and \__kernel_group_show:NN. These func-
tions are documented on page \pageref{page:group_defs}.)

41.19 Decomposing a macro definition

We sometimes want to test if a control sequence can be expanded to reveal a hidden value.
However, we cannot just expand the macro blindly as it may have arguments and none
might be present. Therefore we define these functions to pick either the prefix(es), the
argument specification, or the replacement text from a macro. All of this information is
returned as characters with catcode 12. If the token in question isn’t a macro, the token
\scan_stop: is returned instead.
41.20 Doing nothing functions

\prg_do_nothing: This does not fit anywhere else!
\cs_new:Npn \prg_do_nothing: { }

(End definition for \prg_do_nothing:. This function is documented on page 13.)

41.21 Breaking out of mapping functions

\prg_break_point:Nn \prg_map_break:Nn

In inline mappings, the nesting level must be reset at the end of the mapping, even when
the user decides to break out. This is done by putting the code that must be performed
as an argument of \__prg_break_point:Nn. The breaking functions are then defined to
jump to that point and perform the argument of \__prg_break_point:Nn, before the
user’s code (if any). There is a check that we close the correct loop, otherwise we continue
breaking.

\cs_new_eq:NN \prg_break_point:Nn \use_ii:nn
\cs_new:Npn \prg_map_break:Nn #1#2#3 \prg_break_point:Nn #4#5
\cs_new:Npn \prg_break: #1 \prg_break_point: { }
\cs_new:Npn \prg_break:n #1#2 \prg_break_point: {#1}

(End definition for \prg_break_point:Nn and \prg_map_break:Nn. These functions are documented on page 68.)

\prg_break_point: \prg_break: \prg_break:n

Very simple analogues of \prg_break_point:Nn and \prg_map_break:Nn, for use in fast
short-term recursions which are not mappings, do not need to support nesting, and in
which nothing has to be done at the end of the loop.

\cs_new_eq:NN \prg_break_point: \prg_do_nothing:
\cs_new:Npn \prg_break: #1 \prg_break_point: { }
\cs_new:Npn \prg_break:n #1#2 \prg_break_point: {#1}
41.22 Starting a paragraph

\mode_leave_vertical: The approach here is different to that used by \LaTeXe or plain \TeX, which unbox a
void box to force horizontal mode. That inserts the \everypar tokens before the re-inserted unboxing tokens. The approach here uses either the \quitvmode primitive or the equivalent protected macro. In vertical mode, the \indent primitive is inserted: this
will switch to horizontal mode and insert \everypar tokens and nothing else. Unlike the \LaTeXe version, the availability of \TeX means using a mode test can be done at for
example the start of an \halign.

\begin{verbatim}
\cs_new_protected:Npn \mode_leave_vertical:
  {\if_mode_vertical: \exp_after:wN \tex-indent:D \fi:}
\end{verbatim}

(End definition for \mode_leave_vertical:. This function is documented on page 28.)
Chapter 42

l3expan implementation

The \exp_ module has its private variable to temporarily store the result of x-type argument expansion. This is done to avoid interference with other functions using temporary variables.

(End definition for \l__exp_internal_tl.)

These are defined in l3basics, as they are needed “early”. This is just a reminder of that fact!

(End definition for \exp_after:wN, \exp_not:N, and \exp_not:n. These functions are documented on page 38.)

42.1 General expansion

In this section a general mechanism for defining functions that handle arguments is defined. These general expansion functions are expandable unless x is used. (Any version of x is going to have to use one of the l3TFX3 names for \cs_set:Npx at some point, and so is never going to be expandable.)

The definition of expansion functions with this technique happens in section 42.8. In section 42.2 some common cases are coded by a more direct method for efficiency, typically using calls to \exp_after:wN.

This scratch token list variable is defined in l3basics.

(End definition for \l__exp_internal_tl.)

This code uses internal functions with names that start with \\:: to perform the expansions. All macros are long since the tokens undergoing expansion may be arbitrary user input.

An argument manipulator \\::(Z) always has signature #1:::#2#3 where #1 holds the remaining argument manipulations to be performed, \\:: serves as an end marker for the list of manipulations, #2 is the carried over result of the previous expansion steps and #3 is the argument about to be processed. One exception to this rule is \\::p, which has to grab an argument delimited by a left brace.
#1 is the result of an expansion step, #2 is the remaining argument manipulations and #3 is the current result of the expansion chain. This auxiliary function moves #1 back after #3 in the input stream and checks if any expansion is left to be done by calling #2. In by far the most cases we need to add a set of braces to the result of an argument manipulation so it is more effective to do it directly here. Actually, so far only the \texttt{c} of the final argument manipulation variants does not require a set of braces.

\begin{verbatim}
\cs_new:Npn \_\_exp_arg_next:nnn \#1\#2\#3 { \#2 \::: { \#3 \{\#1\} } }
\cs_new:Npn \_\_exp_arg_next:Nnn \#1\#2\#3 { \#2 \::: { \#3 \#1 } }
\end{verbatim}

(End definition for \_\_exp_arg_next:nnn and \_\_exp_arg_next:Nnn.)

\texttt{\:::} The end marker is just another name for the identity function.

\begin{verbatim}
\cs_new:Npn \::: \#1 {\#1}
\end{verbatim}

(End definition for \:::. This function is documented on page 42.)

\texttt{\::n} This function is used to skip an argument that doesn’t need to be expanded.

\begin{verbatim}
\cs_new:Npn \::n \#1 \::: \#2\#3 { \#1 \::: { \#2 \{\#3\} } }
\end{verbatim}

(End definition for \::n. This function is documented on page 42.)

\texttt{\::N} This function is used to skip an argument that consists of a single token and doesn’t need to be expanded.

\begin{verbatim}
\cs_new:Npn \::N \#1 \::: \#2\#3 { \#1 \::: \{\#2\#3\} }
\end{verbatim}

(End definition for \::N. This function is documented on page 42.)

\texttt{\::p} This function is used to skip an argument that is delimited by a left brace and doesn’t need to be expanded. It is not wrapped in braces in the result.

\begin{verbatim}
\cs_new:Npn \::p \#1 \::: \#2\#3\# { \#1 \::: \{\#2\#3\} }
\end{verbatim}

(End definition for \::p. This function is documented on page 42.)

\texttt{\::c} This function is used to skip an argument that is turned into a control sequence without expansion.

\begin{verbatim}
\cs_new:Npn \::c \#1 \::: \#2\#3
\{ \exp_after:wN \_\_\_exp_arg_next:Nnn \cs:w \#3 \cs_end: \{\#1\} \{\#2\} \}
\end{verbatim}

(End definition for \::c. This function is documented on page 42.)

\texttt{\::o} This function is used to expand an argument once.

\begin{verbatim}
\cs_new:Npn \::o \#1 \::: \#2\#3
\{ \exp_after:wN \_\_\_exp_arg_next:nnn \exp_after:wN \{\#3\} \{\#1\} \{\#2\} \}
\end{verbatim}

(End definition for \::o. This function is documented on page 42.)

\texttt{\::e} With the \texttt{\ expanded} primitive available, just expand. Otherwise defer to \texttt{\exp_args:Ne} implemented later.

\begin{verbatim}
\cs_if_exist:NTF \text_expanded:D
\{
\cs_new:Npn \::e \#1 \::: \#2\#3
\{ \text_expanded:D \{ \exp_not:n \{ \#1 \::: \} \{ \exp_not:n \{\#2\} \{\#3\} \} \}
\}
\}
\cs_new:Npn \::e \#1 \::: \#2\#3
\{ \exp_args:Ne \_\_\_exp_arg_next:nnn \{\#3\} \{\#1\} \{\#2\} \}
\}
\end{verbatim}
\exp_stop_f: This function is used to expand a token list until the first unexpandable token is found. This is achieved through \exp:w \exp_end Continue_f:w that expands everything in its way following it. This scanning procedure is terminated once the expansion hits something non-expandable (if that is a space it is removed). We introduce \exp_stop_f: to mark such an end-of-expansion marker. For example, f-expanding \cs_set_eq:Nc \aaa { b \l_tmpa_tl b } where \l_tmpa_tl contains the characters lur gives \tex_let:D \aaa = \blurb which then turns out to start with the non-expandable token see. Since the expansion of \exp:w \exp_end Continue_f:w is empty, we wind up with a fully expanded list, only \TeX{} has not tried to execute any of the non-expandable tokens. This is what differentiates this function from the x argument type.

\cs_new:Npn \::f \#1 \::: \#2\#3
\exp_after:wN \__exp_arg_next:nnn
\exp_after:wN { \exp:w \exp_end_continue_f:w \#3 }\#1\#2
\}
\use:nn { \cs_new_eq:NN \exp_stop_f: } { - }

(End definition for \::f and \exp_stop_f:. These functions are documented on page 42.)

\::x This function is used to expand an argument fully. We build in the expansion of \_\_exp_arg_next:nnn.
\cs_new_protected:Npn \::x \#1 \::: \#2\#3
\cs_set_nopar:Npx \l__exp_internal_tl
{ \exp_not:n { \::: } { \exp_not:n {\#2} {\#3} } }\#1\#2
\}

(End definition for \::x. This function is documented on page 42.)

\::v muskip, or built-in \TeX{} register. The V version expects a single token whereas v like c creates a csname from its argument given in braces and then evaluates it as if it was a V. The \exp:w sets off an expansion similar to an f-type expansion, which we terminate using \exp_end:. The argument is returned in braces.
\cs_new:Npn \::v \#1 \::: \#2\#3
\exp_after:wN \_\_exp_arg_next:nnn
\exp_after:wN { \exp:w \_\_exp_eval_register:N \#3 }\#1\#2
\}
\cs_new:Npn \::V \#1 \::: \#2\#3
\exp_after:wN \_\_exp_arg_next:nnn
\exp_after:wN { \exp:w \_\_exp_eval_register:c \#3 }\#1\#2
\}

(End definition for \::v and \::V. These functions are documented on page 42.)
This function evaluates a register. Now a register might exist as one of two things: A parameter-less macro or a built-in \TeX register such as \texttt{\textbackslash count}. For the \TeX registers we have to utilize a \texttt{\textbackslash the} whereas for the macros we merely have to expand them once. The trick is to compare the meaning of the token in question when it has been prefixed with \texttt{\exp_not:N} and the token itself. If it is a macro, the prefixed \texttt{\exp_not:N} temporarily turns it into the primitive \texttt{\scan_stop:}.

If the token was not a macro it may be a malformed variable from a c expansion in which case it is equal to the primitive \texttt{\scan_stop:}. In that case we throw an error. We could let \TeX do it for us but that would result in the rather obscure

\begin{verbatim}
! You can't use \texttt{\textbackslash relax} after \texttt{\textbackslash the}.
\end{verbatim}

which while quite true doesn't give many hints as to what actually went wrong. We provide something more sensible.

The next bit requires some explanation. The function must be initiated by \texttt{\exp:w} and we want to terminate this expansion chain by inserting the \texttt{\exp_end:} token. However, we have to expand the register \texttt{\#1} before we do that. If it is a \TeX register, we need to execute the sequence \texttt{\exp_after:wN \exp_end: \tex_the:D \#1} and if it is a macro we need to execute \texttt{\exp_after:wN \exp_end: \#1}. We therefore issue the longer of the two sequences and if the register is a macro, we remove the \texttt{\tex_the:D}.

Clean up nicely, then call the undefined control sequence. The result is an error message looking like this:

\begin{verbatim}
! Undefined control sequence.
<argument> \LaTeX3 error: Erroneous variable used!
1.55 \tl_set:Nv \l_tmpa_tl {undefined_tl}
\end{verbatim}

(End definition for \texttt{\__exp_eval_register:N} and \texttt{\__exp_eval_error_msg:w}.)

382
42.2 Hand-tuned definitions

One of the most important features of these functions is that they are fully expandable.

\begin{verbatim}
\exp_args:Nc \exp_args:cc
(End definition for \exp_args:Nc and \exp_args:cc. These functions are documented on page 34.)

\exp_args:NNc \exp_args:Ncc \exp_args:Nccc
Here are the functions that turn their argument into csnames but are expandable.

\begin{verbatim}
\cs_new:Npn \exp_args:NNc #1#2#3
\{ \exp_after:wN #1 \exp_after:wN #2 \cs:w #3\cs_end: \}
\cs_new:Npn \exp_args:Ncc #1#2#3
\{ \exp_after:wN #1 \cs:w #2 \exp_after:wN \cs_end: \cs:w #3 \cs_end: \}
\cs_new:Npn \exp_args:Nccc #1#2#3#4
\{ \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN \cs_end: \cs:w #3 \exp_after:wN \cs_end: \cs:w #4 \cs_end: \}
\end{verbatim}
(End definition for \exp_args:NNc, \exp_args:Ncc, and \exp_args:Nccc. These functions are documented on page 36.)

\exp_args:No \exp_args:NNo \exp_args:NNNo
Those lovely runs of expansion!

\begin{verbatim}
\cs_new:Npn \exp_args:No #1#2
\{ \exp_after:wN #1 \exp_after:wN {#2} \}
\cs_new:Npn \exp_args:NNo #1#2#3
\{ \exp_after:wN #1 \exp_after:wN \cs:w #2 \exp_after:wN \cs_end: \exp_after:wN \cs:w #3 \exp_after:wN \cs_end: \}
\cs_new:Npn \exp_args:NNNo #1#2#3#4
\{ \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN \cs_end: \exp_after:wN \cs:w #3 \exp_after:wN \cs_end: \exp_after:wN \cs:w #4 \exp_after:wN \cs_end: \}
\end{verbatim}
(End definition for \exp_args:No, \exp_args:NNo, and \exp_args:NNNo. These functions are documented on page 35.)

\exp_args:Ne
When the \texttt{expanded} primitive is available, use it. Otherwise use \texttt{_\_exp_e:nn}, defined later, to fully expand tokens.
\begin{verbatim}
\cs_if_exist:NTF \tex_expanded:D
\{ \cs_new:Npn \exp_args:Ne #1#2 \{ \exp_after:wN \tex_expanded:D \{ {#2} \} \}
\}
\end{verbatim}
(End definition for \exp_args:Ne. This function is documented on page 35.)
\end{verbatim}
Some more hand-tuned function with three arguments. If we forced that an \texttt{a} argument always has braces, we could implement \texttt{exp_args:Nco} with less tokens and only two arguments.
\exp_args:NNV
\exp_args:NNV
\exp_args:NcNc
\exp_args:Ncco

A few more that we can hand-tune.

\exp_args:NNV #1#2#3
\exp_args:NNV #1#2#3
\exp_args:NcNc #1#2#3
\exp_args:Ncco #1#2#3

(End definition for \exp_args:NNV and others. These functions are documented on page 36.)
(End definition for \exp_args:NNNV and others. These functions are documented on page 37.)

\exp_args:Nx

\cs_new_protected:Npn \exp_args:Nx \#1\#2
\{ \use:x { \exp_not:N #1 {#2} } \}

(End definition for \exp_args:Nx. This function is documented on page 36.)

### 42.3 Last-unbraced versions

There are a few places where the last argument needs to be available unbraced. First some helper macros.

\cs_new:Npn \__exp_arg_last_unbraced:nn #1#2 { #2#1 }
\cs_new:Npn \::o_unbraced \::: #1#2
\{ \exp_after:wN \__exp_arg_last_unbraced:nn \exp_after:wN {#2} {#1} \}
\cs_new:Npn \::V_unbraced \::: #1#2
\{ \exp_after:wN \__exp_arg_last_unbraced:nn
\exp_after:wN { \exp:w \__exp_eval_register:N #2 } {#1} \}
\cs_new:Npn \::v_unbraced \::: #1#2
\{ \exp_after:wN \__exp_arg_last_unbraced:nn
\exp_after:wN { \exp:w \__exp_eval_register:c {#2} } {#1} \}
\cs_if_exist:NTF \tex_expanded:D
\{ \cs_new:Npn \::e_unbraced \::: #1#2
\{ \tex_expanded:D { \exp_not:n {#1} #2 } \}
\}
\cs_new_protected:Npn \::e_unbraced \::: #1#2
\{ \exp:w \__exp_e:nn {#2} {#1} \}
\cs_new:Npn \::f_unbraced \::: #1#2
\{ \exp_after:wN \__exp_arg_last_unbraced:nn
\exp_after:wN { \exp:w \exp_end_continue_f:w #2 } {#1} \}
\cs_new_protected:Npn \::x_unbraced \::: #1#2
\{ \cs_set_nopar:Npx \l__exp_internal_tl { \exp_not:n {#1} #2 } \}

386
Now the business end: most of these are hand-tuned for speed, but the general system is in place.

\begin{verbatim}
\exp_last_unbraced:No \exp_last_unbraced:NV \exp_last_unbraced:Nv \exp_last_unbraced:Ne \exp_last_unbraced:Nf \exp_last_unbraced:NNo \exp_last_unbraced:NNV \exp_last_unbraced:NNf \exp_last_unbraced:Nco \exp_last_unbraced:NcV \exp_last_unbraced:NNNo \exp_last_unbraced:NNNV \exp_last_unbraced:NNf \exp_last_unbraced:Nx
\end{verbatim}
42.4 Preventing expansion

At the kernel level, we need the primitive behaviour to allow expansion before the brace group.

(End definition for \texttt{\_kernel\_exp\not:w}.)

\texttt{\_kernel\_exp\not:w}  

All these except \texttt{\_kernel\_exp\not:c} call the kernel-internal \texttt{\_kernel\_exp\not:w} namely

(End definition for \texttt{\_kernel\_exp\not:w}.)

\texttt{\_kernel\_exp\not:c}  

All these except \texttt{\_kernel\_exp\not:c} call the kernel-internal \texttt{\_kernel\_exp\not:w} namely
42.5 Controlled expansion

To trigger a sequence of “arbitrarily” many expansions we need a method to invoke \TeX{}'s expansion mechanism in such a way that (a) we are able to stop it in a controlled manner and (b) the result of what triggered the expansion in the first place is null, i.e., that we do not get any unwanted side effects. There aren’t that many possibilities in \TeX{}; in fact the one explained below might well be the only one (as normally the result of expansion is not null).

The trick here is to make use of the fact that \texttt{\textbackslash tex\_roman\textnumber\textbackslash D} expands the tokens following it when looking for a number and that its expansion is null if that number turns out to be zero or negative. So we use that to start the expansion sequence: \texttt{\exp:w} is set equal to \texttt{\textbackslash tex\_roman\textnumber\textbackslash D} in \texttt{l3basics}. To stop the expansion sequence in a controlled way all we need to provide is a constant integer zero as part of expanded tokens. As this is an integer constant it immediately stops \texttt{\textbackslash tex\_roman\textnumber\textbackslash D}'s search for a number. Again, the definition of \texttt{\exp: end} as the integer constant zero is in \texttt{l3basics}. (Note that according to our specification all tokens we expand initiated by \texttt{\exp:w} are supposed to be expandable (as well as their replacement text in the expansion) so we will not encounter a “number” that actually result in a roman numeral being generated. Or if we do then the programmer made a mistake.)

If on the other hand we want to stop the initial expansion sequence but continue with an f-type expansion we provide the alphabetic constant \texttt{''@} that also represents 0 but this time \TeX{}’s syntax for a \texttt{⟨number⟩} continues searching for an optional space (and it continues expansion doing that) — see \TeX{}book page 269 for details.

\texttt{\textbackslash group\_begin:}
\texttt{\textbackslash tex\_cat\textcode:D '``@ = 13}
\texttt{\textbackslash cs\_new\_protected:Npn \exp: end\_continue\_f:w { '``@ } }

If the above definition ever appears outside its proper context the active character ``@ will be executed so we turn this into an error. The test for existence covers the (unlikely) case that some other code has already defined ``@: this is true for example for \texttt{xmltex.tex}.

\texttt{\textbackslash if\_cs\_exist:N ``@}
\texttt{\textbackslash else:}

389
The same but grabbing an argument to remove spaces and braces.

\cs_new:Npn \exp_end_continue_f:nw #1 \{ \msg_expandable_error:nn { kernel } { bad-exp-end-f } \}
\fi:

42.6 Emulating e-type expansion

When the \expanded primitive is available it is used to implement e-type expansion; otherwise we emulate it.

\cs_new:Npn \__exp_e:nn #1
{ \tl_if_head_is_N_type:nTF {#1} { \__exp_e:N } { \tl_if_head_is_group:nTF {#1} { \__exp_e_group:n } { \tl_if_empty:nTF {#1} { \exp_after:wN \__exp_e_end:nn } { \exp_after:wN \__exp_e_space:nn } \exp_after:wN \if_false: \fi: } } \if_false: \fi: #1 \}
\cs_new:Npn \__exp_e_end:nn #1#2 { \exp_end: #2 }
\cs_new:Npn \__exp_e_space:nn #1#2 { \exp_args:Nf \__exp_e:nn {#1}{#2 ~} }

\__exp_e:nn \__exp_e_end:nn

Repeatedly expand tokens, keeping track of fully-expanded tokens in the second argument to \__exp_e:nn; this function eventually calls \__exp_e_end:nn to leave \exp_end: in the input stream, followed by the result of the expansion. There are many special cases: spaces, brace groups, \noexpand, \unexpanded, \the, \primitive. While we use brace tricks \if_false: \fi:, the expansion of this function is always triggered by \exp:w so brace balance is eventually restored after that is hit with a single step of expansion. Otherwise we could not nest e-type expansions within each other.

\cs_new:Npn \__exp_e:nn #1
{ \tl_if_head_is_N_type:nTF {#1} { \__exp_e:N } { \tl_if_head_is_group:nTF {#1} { \__exp_e_group:n } { \tl_if_empty:nTF {#1} { \exp_after:wN \__exp_e_end:nn } { \exp_after:wN \__exp_e_space:nn } \exp_after:wN \if_false: \fi: } } \if_false: \fi: #1 \}
\cs_new:Npn \__exp_e_end:nn #1#2 { \exp_end: #2 }
\cs_new:Npn \__exp_e_space:nn #1#2 { \exp_args:Nf \__exp_e:nn {#1}{#2 ~} }

\__exp_e:nn \__exp_e_end:nn
\__exp_e_space:nn

For an explicit space character, remove it by f-expansion and put it in the (future) output.
For a group, expand its contents, wrap it in two pairs of braces, and call \_\_exp_e_group:n. This function places the first item (the double-brace wrapped result) into the output. Importantly, \tl_head:n works even if the input contains quarks.

\begin{Verbatim}
\cs_new:Npn \_\_exp_e_group:n #1
  \{ 
  \exp_after:wN \_\_exp_e_put:nn 
  \exp_after:wN { \exp_after:wN { \exp_after:wN { 
    \exp:w \if_false: } \fi: \_\_exp_e:nn #1 \} \} \}
\end{Verbatim}

(End definition for \_\_exp_e_group:n, \_\_exp_e_put:nn, and \_\_exp_e_put:nnn.)

For an N-type token, call \_\_exp_e:N with arguments the (first token), the remaining tokens to expand and what’s already been expanded. If the (first token) is non-expandable, including \protected (\long or not) macros, it is put in the result by \_\_exp_e_protected:Nn. The four special primitives \unexpanded, \noexpand, \the, \primitive are detected; otherwise the token is expanded by \_\_exp_e_expandable:Nn.

\begin{Verbatim}
\cs_new:Npn \_\_exp_e:N #1
  \{ 
  \exp_after:wN \_\_exp_e:Nnn 
  \exp_after:wN #1
  \exp_after:wN { \if_false: } \fi: \}
\end{Verbatim}

\begin{Verbatim}
\cs_new:Npn \_\_exp_e:Nnn #1#2#3
  \{ \exp_args:No \_\_exp_e:nn \{ \use_none:n #2 \} \{ #3 #1 \} \}
\end{Verbatim}

(End definition for \_\_exp_e:N, \_\_exp_e:Nnn, \_\_exp_e_protected:Nn, and \_\_exp_e_expandable:Nn.)
We don't try hard to make sensible error recovery since the error recovery of \text/tex_-
primitive:D when followed by something else than a primitive depends on the engine. The
only valid case is when what follows is N-type. Then distinguish special primitives
unexpanded, noexpand, \the, \primitive from other primitives. In the “other” case,
the only reasonable way to check if the primitive that follows \text/tex_primitive:D is
expandable is to expand and compare the before-expansion and after-expansion results.
If they coincide then probably the primitive is non-expandable and should be put in the
output together with \text/tex_primitive:D (one can cook up contrived counter-examples
where the true \expanded would have an infinite loop), and otherwise one should continue
expanding.

\cs_new:Npn \__exp_e_primitive:Nnn \#1\#2
\{\if_false: { \fi:
\tl_if_head_is_N_type:nTF {\#2}
{ \__exp_e_primitive_aux:NNw \#1 }
{ \msg_expandable_error:nnn { kernel } { e-type }
{ Missing-primitive-name }
\__exp_e_primitive_aux:NNw \#1 \c_empty_tl
#2
}
\}
\exp_after:wN \__exp_e_primitive_aux:NNn
\exp_after:wN \#1
\exp_after:wN \#2
\exp_after:wN { \if_false: } \fi:
\}
\cs_new:Npn \__exp_e_primitive_aux:NNw \#1\#2
\{\exp_args:Nf \str_case_e:nnTF { \cs_to_str:N \#2 }
{ \expanded \#2 }
{ \unexpanded \#2 } \{ \__exp_e_unexpanded:Nnn \exp_not:n \}
{ \noexpand } \{ \__exp_e_noexpand:Nnn \exp_not:N \}
{ \the } \{ \__exp_e_the:Nnn \tex_the:D \}
{ \sys_if_engine_xetex:T { pdf }
\sys_if_engine_luatex:T { pdf }
\primitive
\} \{ \__exp_e_primitive:Nnn \#1 }
\}
{ \__exp_e_primitive_other:Nn \#1 \#2 }
\}
\cs_new:Npn \__exp_e_primitive_other:Nn
\{\exp_args:No \str_case_e:nnTF { \cs_to_str:N \#3 }
{ \expanded \#3 }
#1 \#2 \{\#3\}
\}
\cs_new:Npn \__exp_e_primitive_other_aux:nNNnn \#1#2#3#4#5
\{
  \str_if_eq:nnTF {#1} { #2 #3 #4 }
  { \__exp_e:nn \#4 \{ #5 #2 #3 \} }
  { \__exp_e:nn \#1 \{ \#5 \} }
\}

(End definition for \__exp_e_primitive:Nnn and others.)

\__exp_e_noexpand:Nnn

The \noexpand primitive has no effect when followed by a token that is not N-type; otherwise \__exp_e_put:nn can grab the next token and put it in the result unchanged.
\cs_new:Npn \__exp_e_noexpand:Nnn \#1\#2
\{
  \tl_if_head_is_N_type:nTF {#2}
  { \__exp_e_put:nn } { \__exp_e:nn } {#2}
\}

(End definition for \__exp_e_noexpand:Nnn.)

\__exp_e_unexpanded:Nnn
\__exp_e_unexpanded:nn
\__exp_e_unexpanded:N
\__exp_e_unexpanded:N

The \unexpanded primitive expands and ignores any space, \scan_stop:, or token affected by \exp_not:N, then expects a brace group. Since we only support brace-balanced token lists it is impossible to support the case where the argument of \unexpanded starts with an implicit brace. Even though we want to expand and ignore spaces we cannot blindly f-expand because tokens affected by \exp_not:N should discarded without being expanded further.

As usual distinguish four cases: brace group (the normal case, where we just put the item in the result), space (just f-expand to remove the space), empty (an error), or N-type \texttt{⟨token⟩}. In the last case call \__exp_e_unexpanded:Nn triggered by an f-expansion. Having a non-expandable \texttt{⟨token⟩} after \unexpanded is an error (we recover by passing \{\} to \unexpanded: this is different from \TeX{} because the error recovery of \unexpanded changes the balance of braces), unless that \texttt{⟨token⟩} is \scan_stop: or a space (recall that we don’t implement the case of an implicit begin-group token). An expandable \texttt{⟨token⟩} is instead expanded, unless it is \noexpand. The latter primitive can be followed by an expandable \texttt{N}-type token (removed), by a non-expandable one (kept and later causing an error), by a space (removed by f-expansion), or by a brace group or nothing (later causing an error).
\cs_new:Npn \__exp_e_unexpanded:Nnn \#1 \{ \__exp_e_unexpanded:nn \}
\cs_new:Npn \__exp_e_unexpanded:nn \#1
\{
  \tl_if_head_is_N_type:nTF \#1
  { \exp_args:Nf \__exp_e_unexpanded:nn
    \{ \__exp_e_unexpanded:N \#1 \} \#1 }
  {
    \tl_if_empty:nTF \#1
    \msg_expandable_error:nnn
    \{ kernel \} \{ e-type \}
    \{ \unexpanded missing-brace \}
  }
\}
\exp_e_end:nn

\exp_args:Nf \exp_e_unexpanded:nn

\{#1\}

\cs_new:Npn \exp_e_unexpanded:nN #1#2
\{ #1\exp_after:wN \if_meaning:w \exp_not:N #2 #2\exp_after:wN \use_i:nn
\else:\exp_after:wN \use_ii:nn
\fi: #1\}

\token_if_eq_catcode:NNTF #2 \c_space_token
\exp_stop_f: #1\}
\exp_after:wN \if_meaning:w \exp_not:N #1 #1\exp_after:wN \use_i:nn
\fi: #1\}

(End definition for \exp_e_unexpanded:Nn and others.)

\__exp_e_the:Nnn
\__exp_e_the:N
\__exp_e_the_toks_reg:N

Finally implement \the. Followed by anything other than an \N-type ⟨token⟩ this causes
an error (we just let \LaTeX make one), otherwise we test the ⟨token⟩. If the ⟨token⟩
is expandable, expand it. Otherwise it could be any kind of register, or things like
\numexpr, so there is no way to deal with all cases. Thankfully, only \toks data needs
to be protected from expansion since everything else gives a string of characters. If the
⟨token⟩ is \toks we find a number and unpack using the the_toks functions. If it is a
token register we unpack it in a brace group and call \exp_e_put:nn to move it to
the result. Otherwise we unpack and continue expanding (useless but safe) since it is basically impossible to have a handle on where the result of \the ends.

\cs_new:Npn \_\_exp_e_the:Nnn \#1 \#2
\exp_after:wN \if_meaning:w \exp_not:N \#1 \#1
\exp_after:wN \use_i:nn
\else:\exp_after:wN \use_ii:nn \fi:
\exp_after:wN \__exp_e_the:N
\exp_after:wN ?\exp_after:wN {\exp:w \if_false: } \fi:
\exp_after:wN \exp_end: \#1
\cs_new:Npn \_\_exp_e_the:N \#1
\exp_after:wN \if_meaning:w \exp_not:N \#1 \#1
\exp_after:wN \__exp_e_the_toks:n
\else:\__exp_e_if_toks_register:NTF \#1
{ \exp_after:wN \_\_exp_e_the_toks_reg:N }
\exp_after:wN \_\_exp_e:nn \exp_after:wN { \tex_the:D \if_false: } \fi:
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1
\exp_after:wN \_\_exp_e_the_toks_reg:N \#1

(End definition for \_\_exp_e_the:Nnn, \_\_exp_e_the:N, and \_\_exp_e_the_toks_reg:N)

\_\_exp_e_the_toks:wnn \_\_exp_e_the_toks:n \_\_exp_e_the_toks:N

The calling function has applied \int_value:w so we collect digits with \_\_exp_e_the_the_toks:n (which gets the token list as an argument) and \_\_exp_e_the_toks:N (which gets the first token in case it is N-type). The digits are themselves collected into an \int_value:w argument to \_\_exp_e_the_toks:wnn. Then that function unpacks the \toks{number} into the result. We include ? because \_\_exp_e_put:nnn removes one item from its second argument. Note that our approach is rather crude: in cases like \the\toks12-34 the first \int_value:w removes the space and we will incorrectly unpack the \the\toks1234.
\cs_new:Npn \__exp_e_the_toks:wnn #1; #2
{
\exp_args:No \__exp_e_put:nnn
{ \tex_the:D \tex_toks:D #1 } { ? #2 }
}
\cs_new:Npn \__exp_e_the_toks:n #1
{
\tl_if_head_is_N_type:nTF {#1}
{ \exp_after:wN \__exp_e_the_toks:N \if_false: { \fi: #1 } }
{ ; {#1} }
}
\cs_new:Npn \__exp_e_the_toks:N #1
{
\if_int_compare:w 10 < 9 \token_to_str:N #1 \exp_stop_f:
\exp_after:wN \use_i:nn
\else:
\exp_after:wN \use_ii:nn
\fi:
{ #1
\exp_after:wN \__exp_e_the_toks:n
\exp_after:wN { \if_false: } \fi:
}
{ \exp_after:wN ;
\exp_after:wN { \if_false: } \fi: #1
}
)

(End definition for \__exp_e_the_toks:wnn, \__exp_e_the_toks:n, and \__exp_e_the_toks:N.)

\exp_if_toks_register:NTF \__exp_e_the_XeTeXinterchartoks:
\__exp_e_the_errhelp:
\__exp_e_the_everycr:
\__exp_e_the_everydisplay:
\__exp_e_the_everyeof:
\__exp_e_the_everyjob:
\__exp_e_the_everymath:
\__exp_e_the_everypar:
\__exp_e_the_pdfpageattr:
\__exp_e_the_pdfpagesattr:
\__exp_e_the_pdfpkmode:

We need to detect both \toks registers like \toks0 in \LaTeX2ε and parameters such as \everypar, as the result of unpacking the register should not expand further. Registers are found by \token_if_toks_register:NTF by inspecting the meaning. The list of parameters is finite so we just use a \cs_if_exist:cTF test to look up in a table. We abuse \cs_to_str:N's ability to remove a leading escape character whatever it is.

\prg_new_conditional:Npnn \__exp_e_if_toks_register:N #1 { TF }
{
\token_if_toks_register:NTF #1 { \prg_return_true: }
{ \cs_if_exist:cTF
{ \__exp_e_the_
\exp_after:wN \cs_to_str:N
\token_to_meaning:N #1
: }
{ \prg_return_true: } { \prg_return_false: }
}
}

\cs_new_eq:NN \__exp_e_the_XeTeXinterchartoks: ?
\cs_new_eq:NN \__exp_e_the_errhelp: ?
\cs_new_eq:NN \__exp_e_the_everycr: ?
\cs_new_eq:NN \__exp_e_the_everydisplay: ?
\cs_new_eq:NN \__exp_e_the_everyjob: ?
\cs_new_eq:NN \__exp_e_the_everymath: ?
\cs_new_eq:NN \__exp_e_the_everypar: ?
\cs_new_eq:NN \__exp_e_the_pdfpageattr: ?
\cs_new_eq:NN \__exp_e_the_pdfpagesattr: ?
\cs_new_eq:NN \__exp_e_the_pdfpkmode: ?
\cs_new_eq:NN \_\_exp_e_the_everyhbox: ?
\cs_new_eq:NN \_\_exp_e_the_everyjob: ?
\cs_new_eq:NN \_\_exp_e_the_everymath: ?
\cs_new_eq:NN \_\_exp_e_the_everypar: ?
\cs_new_eq:NN \_\_exp_e_the_everyvbox: ?
\cs_new_eq:NN \_\_exp_e_the_output: ?
\cs_new_eq:NN \_\_exp_e_the_pdfpageattr: ?
\cs_new_eq:NN \_\_exp_e_the_pdfpageresources: ?
\cs_new_eq:NN \_\_exp_e_the_pdfpagesattr: ?
\cs_new_eq:NN \_\_exp_e_the_pdfpkmode: ?

(End definition for \_\_exp_e_if_toks_register:NTF and others.)

We are done emulating e-type argument expansion when \texttt{\_\_expanded} is unavailable.

42.7 Defining function variants

\\s__cs_mark
\s__cs_stop

Internal scan marks. No \texttt{l3quark} yet, so do things by hand.
\cs_new_eq:NN \s__cs_mark \scan_stop:
\cs_new_eq:NN \s__cs_stop \scan_stop:

(End definition for \s__cs_mark and \s__cs_stop.)

\q__cs_recursion_stop

Internal recursion quarks. No \texttt{l3quark} yet, so do things by hand.
\cs_new:Npn \q__cs_recursion_stop { \q__cs_recursion_stop }

(End definition for \q__cs_recursion_stop.)

\__cs_use_none_delimit_by_s_stop:w
\__cs_use_i_delimit_by_s_stop:nw
\__cs_use_none_delimit_by_q_recursion_stop:w

Internal scan marks.
\cs_new:Npn \__cs_use_none_delimit_by_s_stop:w #1 \s__cs_stop { }
\cs_new:Npn \__cs_use_i_delimit_by_s_stop:nw #1 #2 \s__cs_stop {#1}
\cs_new:Npn \__cs_use_none_delimit_by_q_recursion_stop:w
#1 \q__cs_recursion_stop { }

(End definition for \__cs_use_none_delimit_by_s_stop:w, \__cs_use_i_delimit_by_s_stop:nw, and \__cs_use_none_delimit_by_q_recursion_stop:w.)

\cs_generate_variant:Nn
\cs_generate_variant:cn

#1: Base form of a function; \textit{e.g.}, \texttt{tl_set:Nn}
#2: One or more variant argument specifiers; \textit{e.g.}, \{Nx,c,cx\}

After making sure that the base form exists, test whether it is protected or not and define \texttt{\_\_cs_tmp:w} as either \texttt{\cs_new:Npx} or \texttt{\cs_new_protected:Npx}, which is then used to define all the variants (except those involving \texttt{x}-expansion, always protected). Split up the original base function only once, to grab its name and signature. Then we wish to iterate through the comma list of variant argument specifiers, which we first convert to a string: the reason is explained later.
\cs_new_protected:Npn \cs_generate_variant:Nn #1#2
\{ 
\__cs_generate_variant:N #1
\use:x
\{ 
\__cs_generate_variant:nnNN
\cs_split_function:N #1

397
The goal here is to pick up protected parent functions. There are four cases: the parent function can be a primitive or a macro, and can be expandable or not. For non-expandable primitives, all variants should be protected; skipping the \else: branch is safe because non-expandable primitives cannot be \TeX{} conditionals.

The other case where variants should be protected is when the parent function is a protected macro: then \texttt{protected} appears in the meaning before the first occurrence of \texttt{macro}. The \texttt{ww} auxiliary removes everything in the meaning string after the first \texttt{ma}. We use \texttt{ma} rather than the full \texttt{macro} because the meaning of the \texttt{\firstmark} primitive (and four others) can contain an arbitrary string after a leading \texttt{\firstmark}: Then, look for \texttt{pr} in the part we extracted: no need to look for anything longer: the only strings we can have are an empty string, \texttt{\long}, \texttt{\protected}, \texttt{\protected\long}, \texttt{\first}, \texttt{\top}, \texttt{\bot}, \texttt{\splittop}, or \texttt{\splitbot}, with \ replaced by the appropriate escape character. If \texttt{pr} appears in the part before \texttt{ma}, the first \texttt{\s__cs_mark} is taken as an argument of the \texttt{\__cs_generate_variant:wwNw} auxiliary, and \#3 is \texttt{\cs_new_protected:Npx}, otherwise it is \texttt{\cs_new_protected:Np}. }

\begin{verbatim}
\exp_not:N \exp_after:wN \exp_not:N \if_meaning:w
\exp_not:N \exp_not:N #1 #1
\cs_set_eq:NN \exp_not:N \__cs_tmp:w \cs_new_protected:Npx
\exp_not:N \else:
\exp_not:N \exp_after:wN \exp_not:N \__cs_generate_variant:ww
\exp_not:N \token_to_meaning:N #1 \tl_to_str:n { ma }
\s__cs_mark
\s__cs_mark \cs_new_protected:Npx
\tl_to_str:n { pr }
\s__cs_mark \cs_new:Npx
\s__cs_stop
\exp_not:N \fi:
\exp_last_unbraced:NNNNo
\exp_last_unbraced:NNNNo
#1 \tl_to_str:n { ma } \#2 \s__cs_mark
\__cs_generate_variant:wwNw \#1
\exp_last_unbraced:NNNNo
\exp_last_unbraced:NNNNo
#1 \tl_to_str:n { pr } \#2 \s__cs_mark \#3 \s__cs_stop
\cs_set_eq:NN \cs_tmp:w \#3
\end{verbatim}
If the boolean is \texttt{\c_false_bool}, the base function has no colon and we abort with an error; otherwise, set off a loop through the desired variant forms. The original function is retained as \#4 for efficiency.

2926 \texttt{\cs_new_protected:Npn \__cs_generate_variant:nnNN #1#2#3#4}
2927 { 
2928 \if_meaning:w \c_false_bool #3 
2929 \msg_error:nnx { kernel } { missing-colon } 
2930 \{ \token_to_str:c {#1} \} 
2931 \exp_after:wN \__cs_use_none_delimit_by_q_recursion_stop:w 
2932 \fi: 
2933 \__cs_generate_variant:Nnnw #4 {#1}{#2} 
2934 }

\texttt{\__cs_generate_variant:Nnnw}

\#1 : Base function.
\#2 : Base name.
\#3 : Base signature.
\#4 : Beginning of variant signature.

First check whether to terminate the loop over variant forms. Then, for each variant form, construct a new function name using the original base name, the variant signature consisting of \( l \) letters and the last \( k - l \) letters of the base signature (of length \( k \)). For example, for a base function \texttt{\prop_put:Nnn} which needs a \texttt{cV} variant form, we want the new signature to be \texttt{cVn}.

There are further subtleties:

- In \texttt{\cs_generate_variant:Nn\foo:nnTF}{xxTF}, we must define \texttt{\foo:xxTF} using \texttt{\exp_args:Nxx}, rather than a hypothetical \texttt{\exp_args:NxxTF}. Thus, we wish to trim a common trailing part from the base signature and the variant signature.

- In \texttt{\cs_generate_variant:Nn\foo:on}{ox}, the function \texttt{\foo:ox} must be defined using \texttt{\exp_args:Nnx}, not \texttt{\exp_args:Nox}, to avoid double \texttt{ox} expansion.

- Lastly, \texttt{\cs_generate_variant:Nn \foo:on}{xn} must trigger an error, because we do not have a means to replace \texttt{ox} by \texttt{x}. More generally, we can only convert \texttt{n} to \texttt{c}, or convert \texttt{n} to \texttt{V}, \texttt{v}, \texttt{o}, \texttt{x}.

All this boils down to a few rules. Only \texttt{n} and \texttt{N}-type arguments can be replaced by \texttt{\cs_generate_variant:Nn}. Other argument types are allowed to be passed unchanged from the base form to the variant: in the process they are changed to \texttt{n} except for \texttt{N} and \texttt{p}-type arguments. A common trailing part is ignored.

We compare the base and variant signatures one character at a time within \texttt{x}\-expansion. The result is given to \texttt{\__cs_generate_variant:wwNN} (defined later) in the form \texttt{\s__cs_mark \{errors\} \s__cs_stop \{base function\} \{new function\}}. If all went well, \texttt{\{errors\}} is empty; otherwise, it is a kernel error message and some clean-up code.

Note the space after \#3 and after the following brace group. Those are ignored by \texttt{TPeX} when fetching the last argument for \texttt{\__cs_generate_variant_loop:nNwN}, but can be used as a delimiter for \texttt{\__cs_generate_variant_loop_end:nwNNNNn}.
\if_meaning:w \scan_stop: #4
  \exp_after:wN \__cs_use_none_delimit_by_q_recursion_stop:w
  \fi:
  \use:x
  { \exp_not:N \__cs_generate_variant:wwNN \__cs_generate_variant_loop:nNwN \s__cs_mark
    #3 -
    { - { } \fi: \__cs_generate_variant_loop_long:WNnNn } -
    { }
    \s__cs_stop
    \exp_not:N #1 {#2} {#4}
  }
\__cs_generate_variant:Nnnw #1 {#2} {#3}
\end{definition}

\__cs_generate_variant_loop:nNwN
\__cs_generate_variant_loop_base:N
\__cs_generate_variant_loop_same:w
\__cs_generate_variant_loop_end:nwwwNNnn
\__cs_generate_variant_loop_long:wNNnn
\__cs_generate_variant_loop_invalid:NNwNNnn
\__cs_generate_variant_loop_special:NNwNNnn

#1: Last few consecutive letters common between the base and variant (more precisely, \__cs_generate_variant_same:N ⟨letter⟩ for each letter).
#2: Next variant letter.
#3: Remainder of variant form.
#4: Next base letter.

The first argument is populated by \__cs_generate_variant_loop_same:w when a variant letter and a base letter match. It is flushed into the input stream whenever the two letters are different: if the loop ends before, the argument is dropped, which means that trailing common letters are ignored.

The case where the two letters are different is only allowed if the base is \texttt{N} and the variant is \texttt{c}, or when the base is \texttt{n} and the variant is \texttt{o, V, v, f} or \texttt{x}. Otherwise, call \__cs_generate_variant_loop_invalid:WNnNn to remove the end of the loop, get arguments at the end of the loop, and place an appropriate error message as a second argument of \__cs_generate_variant:wwNN.

If the letters are distinct and the base letter is indeed \texttt{n} or \texttt{N}, leave in the input stream whatever argument \#1 was collected, and the next variant letter \#2, then loop by calling \__cs_generate_variant_loop:nWNN.

The loop can stop in three ways.

- If the end of the variant form is encountered first, \#2 is \__cs_generate_variant_loop_end:WNnNnn (expanded by the conditional \texttt{\if:w}), which inserts some tokens to end the conditional; grabs the ⟨base name⟩ as \#7, the ⟨variant signature⟩ \#8, the ⟨next base letter⟩ \#1 and the part \#3 of the base signature that wasn’t read yet; and combines those into the ⟨new function⟩ to be defined.

- If the end of the base form is encountered first, \#4 is \texttt{-{}\fi:} which ends the conditional (with an empty expansion), followed by \__cs_generate_variant_loop_long:WNnNn, which places an error as the second argument of \__cs_generate_variant:wwNN.

- The loop can be interrupted early if the requested expansion is unavailable, namely when the variant and base letters differ and the base is not the right one (\texttt{n} or \texttt{N} to
support the variant). In that case too an error is placed as the second argument of \_\_cs_generate_variant:wwNN.

Note that if the variant form has the same length as the base form, \#2 is as described in the first point, and \#4 as described in the second point above. The \_\_cs_generate_variant_loop_end:nwwNNnn breaking function takes the empty brace group in \#4 as its first argument: this empty brace group produces the correct signature for the full variant.

\begin{verbatim}
\cs_new:Npn \_\_cs_generate_variant_loop:nNwN #1#2#3 \s__cs_mark #4
When the base and variant letters are identical, don’t do any expansion. For most argument types, we can use the n-type no-expansion, but the N and p types require a slightly different behaviour with respect to braces. For V-type this function could output N to avoid adding useless braces but that is not a problem.
\cs_new:Npn \__cs_generate_variant_same:N #1 
{
    \if:w N #1 \else:
    \if:w p #1 \else:
        \token_to_str:N n
        \if:w n #1 \else:
            \__cs_generate_variant_loop_special:NwNNnn #1#1
        \fi:
        \fi:
        \fi:
    \fi:
}

(End definition for \__cs_generate_variant_same:N.)

\__cs_generate_variant:wwNN

If the variant form has already been defined, log its existence (provided log-functions
is active). Otherwise, make sure that the \exp_args:N \#3 form is defined, and if it
contains x, change \__cs_tmp:w locally to \cs_new_protected:Npx. Then define the
variant by combining the \exp_args:N \#3 variant and the base function.

\cs_new_protected:Npm \__cs_generate_variant:wwNN
\#1 \s__cs_mark \s__cs_stop \#3#4
{
    \cs_if_free:NT \#4
    {
        \group_begin:
        \__cs_generate_internal_variant:n {#1}
        \__cs_tmp:w \#4 { \exp_not:c { \exp_args:N #1 } \exp_not:N \#3 }
        \group_end:
    }
}

(End definition for \__cs_generate_variant:wwNN.)

\__cs_generate_internal_variant:n
\__cs_generate_internal_variant_loop:n

First test for the presence of x (this is where working with strings makes our lives
easier), as the result should be protected, and the next variant to be defined us-
ing that internal variant should be protected (done by setting \__cs_tmp:w). Then
\__cs_generate_internal_variant:NNn with arguments \cs_new_protected:cpn
\use:x (for protected) or \cs_new:cpn \tex_expanded:D (expandable) and the signa-
ture. If p appears in the signature, or if the function to be defined is expandable and the
primitive \expanded is not available, or if there are more than 8 arguments, call some
fall-back code that just puts the appropriate \:: commands. Otherwise, call \__cs_gen-
erate_internal_one_go:NNn to construct the \exp_args:N... function as a macro
taking up to 9 arguments and expanding them using \use:x or \tex_expanded:D.

\cs_new_protected:Npm \__cs_generate_internal_variant:n \#1
{
    \exp_not:N \__cs_generate_internal_variant:wwNN\#1 #1 \s__cs_mark
    \{ \cs_set_eq:NN \exp_not:N \__cs_tmp:w \cs_new_protected:cpn
    \use:x
    \token_to_str:N x \s__cs_mark
    \}
}

\cs_new:cpn
\exp_not:N \tex_expanded:D
\s__cs_stop
{#1}
\exp_last_unbraced:NNNNo
\cs_new_protected:Npn \__cs_generate_internal_variant:wwnNwn #1
{ \token_to_str:N x } #2 \s__cs_mark #3#4#5#6 \s__cs_stop #7
{ #3
\cs_if_free:cT { exp_args:N #7 }
{ \__cs_generate_internal_variant:NNn #4 #5 {#7} }
}
\cs_set_protected:Npn \__cs_tmp:w #1
{ \cs_new_protected:Npn \__cs_generate_internal_variant:NNn ##1##2##3
{ \if_catcode:w X \use_none:nnnnnnn ##3
\prg_do_nothing: \prg_do_nothing: \prg_do_nothing:
\prg_do_nothing: \prg_do_nothing: \prg_do_nothing:
\prg_do_nothing: \prg_do_nothing: X
\exp_after:wN \__cs_generate_internal_test:Nw \exp_after:wN ##2
\else:
\exp_after:wN \__cs_generate_internal_test_aux:w \exp_after:wN #1
\fi:
##3
\s__cs_mark
{ \use:x
{ ##1 { exp_args:N ##3 }
{ \__cs_generate_internal_variant_loop:n ##3 { : \use_i:nn } }
}

}{ \__cs_generate_internal_one_go:NNn ##1 ##2 {##3} }

}{ \__cs_generate_internal_test:Nw \__cs_generate_internal_test_aux:w
#1 #2 \s__cs_mark #3 #4 \s__cs_stop {#3}
\cs_if_exist:NTF \tex_expanded:D
{ \cs_new_eq:NN \__cs_generate_internal_test:Nw \__cs_generate_internal_test Aux:w
\__cs_generate_internal_test Aux:w
}
{ \cs_new_protected:Npn \__cs_generate_internal_test:Nw #1
{ \if_meaning:w \tex_expanded:D #1
\exp_after:wN \__cs_generate_internal_test_aux:w \exp_after:wN #1
\else:
\exp_after:wN \__cs_generate_internal_test_aux:w
\fi:
}
This command grabs char by char outputting \: #1 (not expanded further). We avoid tests by putting a trailing : \use_i:nn, which leaves \cs_end: and removes the looping
macro. The colon is in fact also turned into `::` so that the required structure for `\exp_args:N...` commands is correctly terminated.

```latex
\cs_new:Npn \__cs_generate_internal_variant_loop:n #1
\{\exp_after:wN \exp_not:N \cs:w :: #1 \cs_end:\}
\__cs_generate_internal_variant_loop:n

(End definition for `\__cs_generate_internal_variant:n` and `\__cs_generate_internal_variant_loop:n`.)
```

```latex
\prg_generate_conditional_variant:Nnn \__cs_generate_variant:nnNnn
\__cs_generate_variant:w \__cs_generate_variant:n
\__cs_generate_variant_p_form:nnn
\__cs_generate_variant_T_form:nnn
\__cs_generate_variant_F_form:nnn
\__cs_generate_variant_TF_form:nnn
\cs_new_protected:Npn \prg_generate_conditional_variant:Nnn #1
\{\use:x\{\__cs_generate_variant:nnNnn \cs_split_function:N #1\}\}
\cs_new_protected:Npn \__cs_generate_variant:nnNnn #1#2#3#4#5
\{\if_meaning:w \c_false_bool #3 \msg_error:nnx { kernel } { missing-colon } \{ \token_to_str:c {#1} \} \fi:\exp_after:wN \__cs_generate_variant:w \tl_to_str:n {#5} , \scan_stop: , \text\cs_recursion_stop \__cs_use_none_delimit_by_s_stop:w \s__cs_mark {#1} {#2} {#4} \s__cs_stop\}
\cs_new_protected:Npn \__cs_generate_variant:w #1 , #2 \s__cs_mark #3#4#5
\{\if_meaning:w \scan_stop: #1 \scan_stop:\if_meaning:w \q\cs_nil #1 \q\cs_nil \use_i:nnn \fi:\exp_after:wN \__cs_use_none_delimit_by_q_recursion_stop:w \else:\cs_if_exist_use:cTF { \__cs_generate_variant_TF_form:nnn } { \token_to_str:c { \#3 : \#4 } \#5 } \{ \msg_error:nxx { kernel } { conditional-form-unknown } \{ \#1} \{ \token_to_str:c { \#3 : \#4 } \} \}
\fi:\\__cs_generate_variant:w #2 \s__cs_mark \{#3 \} \{#4 \} \{#5\}
\}
\cs_new_protected:Npn \__cs_generate_variant:n \__cs_generate_variant_p_form:nnn
\cs_new_protected:Npn \__cs_generate_variant_T_form:nnn
\cs_new_protected:Npn \__cs_generate_variant_F_form:nnn
\cs_new_protected:Npn \__cs_generate_variant_TF_form:nnn
\cs_new_protected:Npn \prg_generate_conditional_variant:Nnn #1
\{\use:x\{\__cs_generate_variant:nnNnn \cs_split_function:N #1\}\}
\cs_new_protected:Npn \__cs_generate_variant:nnNnn #1#2#3#4#5
\{\if_meaning:w \c_false_bool #3 \msg_error:nnx { kernel } { missing-colon } \{ \token_to_str:c {#1} \} \fi:\exp_after:wN \__cs_generate_variant:w \tl_to_str:n {#5} , \scan_stop: , \text\cs_recursion_stop \__cs_use_none_delimit_by_s_stop:w \s__cs_mark {#1} {#2} {#4} \s__cs_stop\}
\cs_new_protected:Npn \__cs_generate_variant:w #1 , #2 \s__cs_mark #3#4#5
\{\if_meaning:w \scan_stop: #1 \scan_stop:\if_meaning:w \q\cs_nil #1 \q\cs_nil \use_i:nnn \fi:\exp_after:wN \__cs_generate_variant:w \tl_to_str:n {#5} , \scan_stop: , \text\cs_recursion_stop \else:\cs_if_exist_use:cTF { \__cs_generate_variant_TF_form:nnn } { \token_to_str:c { \#3 : \#4 } \#5 } \{ \msg_error:nxx { kernel } { conditional-form-unknown } \{ \#1} \{ \token_to_str:c { \#3 : \#4 } \} \}
\fi:\\__cs_generate_variant:w #2 \s__cs_mark \{#3 \} \{#4 \} \{#5\}
\}
```

406
42.8 Definitions with the automated technique

Some of these could be done more efficiently, but the complexity of coding then becomes an issue. Notice that the auto-generated functions actually take no arguments themselves.

\begin{verbatim}
\cs_set_protected:Npn \__cs_tmp:w #1
  {\group_begin:\__cs_generate_internal_variant:n {##1}\group_end:}
\__cs_tmp:w { nc }
\__cs_tmp:w { no }
\__cs_tmp:w { nV }
\__cs_tmp:w { nv }
\__cs_tmp:w { ne }
\__cs_tmp:w { nf }
\__cs_tmp:w { oc }
\__cs_tmp:w { oo }
\end{verbatim}

Here are the actual function definitions, using the helper functions above. The group is used because \__cs_generate_internal_variant:n redefines \cs_tmp:w locally.
End definition for \exp_args:Nnc and others. These functions are documented on page 36.

(End definition for \exp_args:Nnc and others. These functions are documented on page 37.)

\exp_args:NNcf
\exp_args:NNno
\exp_args:NNnV
\exp_args:NNoo
\exp_args:NNVV
\exp_args:Ncno
\exp_args:NcnV
\exp_args:Ncoo
\exp_args:NcVV
\exp_args:Nncnc
\exp_args:Nnno
\exp_args:Nnnf
\exp_args:Nnff
\exp_args:Nooo
\exp_args:Nnfo
\exp_args:Neee
\exp_args:Nffo
\exp_args:NNNx
\exp_args:NNnx
\exp_args:NNox
\exp_args:Nccx
\exp_args:Nncx
\exp_args:Nnxx
\exp_args:Nox
\exp_args:Nnx
\exp_args:Nxx

\exp_args:Noox

(End definition for \exp_args:NNcf and others. These functions are documented on page 37.)
Chapter 43

l3sort implementation

43.1 Variables

\__sort_length_int The sequence has \__sort_length_int items and is stored from \__sort_min_int to \__sort_top_int – 1. While reading the sequence in memory, we check that \__sort_top_int remains at most \__sort_max_int, precomputed by \__sort_compute_range:. That bound is such that the merge sort only uses \toks registers less than \__sort_true_max_int, namely those that have not been allocated for use in other code: the user’s comparison code could alter these.

\__sort_block_int Merge sort is done in several passes. In each pass, blocks of size \__sort_block_int are merged in pairs. The block size starts at 1, and, for a length in the range \(2^k + 1, 2^{k+1}\), reaches \(2^k\) in the last pass.

(End definition for \__sort_length_int and others.)

(End definition for \__sort_block_int.)
When merging two blocks, \texttt{l\_sort\_begin\_int} marks the lowest index in the two blocks, and \texttt{l\_sort\_end\_int} marks the highest index, plus 1.

\begin{verbatim}
\int_new:N \l__sort_begin_int
\int_new:N \l__sort_end_int
(End definition for \texttt{l\_sort\_begin\_int} and \texttt{l\_sort\_end\_int}.)
\end{verbatim}

When merging two blocks (whose end-points are \texttt{beg} and \texttt{end}), \texttt{A} starts from the high end of the low block, and decreases until reaching \texttt{beg}. The index \texttt{B} starts from the top of the range and marks the register in which a sorted item should be put. Finally, \texttt{C} points to the copy of the high block in the interval of registers starting at \texttt{l\_sort\_length\_int}, upwards. \texttt{C} starts from the upper limit of that range.

\begin{verbatim}
\int_new:N \l__sort_A_int
\int_new:N \l__sort_B_int
\int_new:N \l__sort_C_int
(End definition for \texttt{l\_sort\_A\_int}, \texttt{l\_sort\_B\_int}, and \texttt{l\_sort\_C\_int}.)
\end{verbatim}

When merging two blocks, \texttt{s\_sort\_mark} marks the lowest index in the two blocks, and \texttt{s\_sort\_stop} marks the highest index, plus 1.

\begin{verbatim}
\scan_new:N \s__sort_mark
\scan_new:N \s__sort_stop
(End definition for \texttt{s\_sort\_mark} and \texttt{s\_sort\_stop}.)
\end{verbatim}

### 43.2 Finding available \texttt{toks} registers

\texttt{l\_sort\_shrink\_range}: After \texttt{l\_sort\_compute\_range}: (defined below) determines that \texttt{toks} registers between \texttt{l\_sort\_min\_int} (included) and \texttt{l\_sort\_true\_max\_int} (excluded) have not yet been assigned. \texttt{l\_sort\_shrink\_range}: computes \texttt{l\_sort\_max\_int} to reflect the need for a buffer when merging blocks in the merge sort. Given \(2^n \leq A \leq 2^n + 2^n - 1\) registers we can sort \(\lceil A/2 \rceil + 2^{n-2}\) items while if we have \(2^n + 2^n - 1 \leq A \leq 2^{n+1}\) registers we can sort \(A - 2^{n-1}\) items. We first find out a power \(2^n\) such that \(2^n \leq A \leq 2^{n+1}\) by repeatedly halving \texttt{l\_sort\_block\_int}, starting at \(2^{15}\) or \(2^{14}\) namely half the total number of registers, then we use the formulas and set \texttt{l\_sort\_max\_int}.

\begin{verbatim}
\cs_new_protected:Npn \__sort_shrink_range: {
  \int_set:Nn \l__sort_A_int { \l__sort_true_max_int - \l__sort_min_int + 1 }
  \int_set:Nn \l__sort_block_int { \c_max_register_int / 2 }
  \__sort_shrink_range_loop:
  \int_set:Nn \l__sort_max_int \l__sort_block_int
  \{ \int_compare:nNnTF { \l__sort_block_int * 3 / 2 } > \l__sort_A_int \{
    \{ \l__sort_block_int * 3 / 2 } > \l__sort_A_int \{
      \l__sort_min_int
      + \l__sort_A_int - 1 ) / 2
      + \l__sort_block_int / 4
      + 1
      \}
  \}
  \{ \l__sort_true_max_int - \l__sort_block_int / 2 }
\}
\end{verbatim}
First find out what \toks have not yet been assigned. There are many cases. In \LaTeX2ε with no package, available \toks range from \count15 + 1 to \c_max_register_int included (this was not altered despite the 2015 changes). When \loctoks is defined, namely in plain \(\epsilon\)\TeX, or when the package etex is loaded in \LaTeX2ε, redefine \_sort\_compute_range: to use the range \count265 to \count275 − 1. The elocalloc package also defines \loctoks but uses yet another number for the upper bound, namely \e@alloc@top (minus one). We must check for \loctoks every time a sorting function is called, as etex or elocalloc could be loaded.

In Con\TeX MkJIV the range is from \c_syst_last_allocated_toks+1 to \c_max_register_int, and in MkJII it is from \lastallocatedtoks+1 to \c_max_register_int. In all these cases, call \_sort\_shrink_range:.

\begin{verbatim}
\cs_new_protected:Npn \_sort_shrink_range_loop:
\{ \if_int_compare:w \l__sort_A_int < \l__sort_block_int \tex_divide:D \l__sort_block_int 2 \exp_stop_f: \exp_after:wN \_sort_shrink_range_loop: \fi: \fi: \}
\end{verbatim}
\cs_if_exist:NT \loctoks \__sort_redefine_compute_range:
\tl_map_inline:nn { \lastallocatedtoks \c_syst_last_allocated_toks }
{
\cs_if_exist:NT #1
{
\cs_gset_protected:Npn \__sort_compute_range:
{
\int_set:Nn \l__sort_min_int { #1 + 1 }
\int_set:Nn \l__sort_true_max_int { \c_max_register_int + 1 }
\__sort_shrink_range:
}
}
}
(End definition for \__sort_compute_range:, \__sort_redefine_compute_range:, and \c__sort_max_length_int.)

43.3 Protected user commands
\__sort_main:NNNn

Sorting happens in three steps. First store items in \toks registers ranging from \l__sort_min_int to \l__sort_top_int − 1, while checking that the list is not too long. If we reach the maximum length, that’s an error; exit the group. Secondly, sort the array of \toks registers, using the user-defined sorting function: \__sort_level: calls \__sort_compare:nn as needed. Finally, unpack the \toks registers (now sorted) into the target \tl, or into \g__sort_internal_seq for seq and clist. This is done by \__sort_seq:NNNNn and \__sort_tl:NNn.
\cs_new_protected:Npn \__sort_main:NNNn #1#2#3#4
{
\__sort_disable_toksdef:
\__sort_compute_range:
\int_set_eq:NN \l__sort_top_int \l__sort_min_int #1 #3
{\if_int_compare:w \l__sort_top_int = \l__sort_max_int
 \__sort_too_long_error:NNw #2 #3
 \fi:
 \tex_toks:D \l__sort_top_int {##1}
 \int_incr:N \l__sort_top_int
}
\int_set:Nn \l__sort_length_int
{\l__sort_top_int - \l__sort_min_int }
\cs_set:Npn \__sort_compare:nn #1 #2 (#4)
\int_set:Nn \l__sort_block_int { 1 }
\__sort_level:
}
(End definition for \__sort_main:NNNn.)
\tl_sort:Nn
\tl_sort:cn
\tl_gsort:Nn
\tl_gsort:cn
\__sort_tl:NNn
\__sort_tl_toks:w

Call the main sorting function then unpack \toks registers outside the group into the target token list. The unpacking is done by \__sort_tl_toks:w; registers are numbered from \l__sort_min_int to \l__sort_top_int − 1. For expansion behaviour we need
a couple of primitives. The \tl_gclear:N reduces memory usage. The \prg_break_point: is used by \__sort_main:NNNn when the list is too long.

\begin{verbatim}
\cs_new_protected:Npn \tl_sort:Nn { \__sort_tl:NNn \tl_set_eq:NN }
\cs_generate_variant:Nn \tl_sort:Nn { c }
\cs_new_protected:Npn \tl_gsort:Nn { \__sort_tl:NNn \tl_gset_eq:NN }
\cs_generate_variant:Nn \tl_gsort:Nn { c }
\cs_new_protected:Npn \__sort_tl:NNn #1#2#3
\group_begin:
\__sort_main:NNNn #1 #2 #4 {#5}
\__kernel_tl_gset:Nx \g__sort_internal_tl
\__sort_tl_toks:w \l__sort_min_int ;
\group_end:

\prg_break_point:
\cs_new:Npn \__sort_tl_toks:w #1 ;
\if_int_compare:w #1 < \l__sort_top_int
\tex_the:D \tex_toks:D #1 }
\exp_after:wN \__sort_tl_toks:w
\int_value:w \int_eval:n { #1 + 1 } \exp_after:wN ;
\fi:
\end{verbatim}

(End definition for \tl_sort:Nn and others. These functions are documented on page 115.)

\seq_sort:Nn \seq_sort:cn \seq_gsort:Nn \seq_gsort:cn \clist_sort:Nn \clist_sort:cn \clist_gsort:Nn \clist_gsort:cn \__sort_seq:NNNNn
Use the same general framework for seq and clist. Apply the general sorting code, then unpack \toks into \g__sort_internal_seq. Outside the group copy or convert (for clist) the data to the target variable. The \seq_gclear:N reduces memory usage. The \prg_break_point: is used by \__sort_main:NNNn when the list is too long.

\begin{verbatim}
\cs_new_protected:Npn \seq_sort:Nn { \__sort_seq:NNNNn \seq_map_inline:Nn \seq_map_break:n \seq_set_eq:NN }
\cs_generate_variant:Nn \seq_sort:Nn { c }
\cs_new_protected:Npn \seq_gsort:Nn { \__sort_seq:NNNNn \seq_map_inline:Nn \seq_map_break:n \seq_gset_eq:NN }
\cs_generate_variant:Nn \seq_gsort:Nn { c }
\cs_new_protected:Npn \clist_sort:Nn { \__sort_seq:NNNNn \clist_map_inline:Nn \clist_map_break:n \clist_set_from_seq:NN }
\cs_generate_variant:Nn \clist_sort:Nn { c }
\cs_new_protected:Npn \clist_gsort:Nn { \__sort_seq:NNNNn \clist_map_inline:Nn \clist_map_break:n \clist_gset_from_seq:NN }
\cs_generate_variant:Nn \clist_gsort:Nn { c }
\cs_new_protected:Npn \__sort_seq:NNNNn #1#2#3#4#5
\group_begin:
\__sort_main:NNNn #1 #2 #4 #5
\end{verbatim}

413
3458 \seq_gset_from_inline_x:Nnn \g__sort_internal_seq
3459 { \int_step_function:nnN
3460 \l__sort_min_int \l__sort_top_int - 1
3461 } { \tex_the:D \tex_toks:D ##1 }
3462 \group_end:
3463 \seq_gclear:N \g__sort_internal_seq
3464 \prg_break_point:

(End definition for \seq_sort:Nn and others. These functions are documented on page 146.)

43.4 Merge sort

\_\_sort_level: This function is called once blocks of size \_\_sort_block_int (initially 1) are each sorted. If the whole list fits in one block, then we are done (this also takes care of the case of an empty list or a list with one item). Otherwise, go through pairs of blocks starting from 0, then double the block size, and repeat.

3469 \cs_new_protected:Npn \_\_sort_level:
3470 { \if_int_compare:w \l__sort_block_int < \l__sort_length_int
3471 \_\_sort_end_int \_\_sort_min_int
3472 \_\_sort_merge_blocks:
3473 \tex_advance:D \_\_sort_block_int \_\_sort_block_int
3474 \exp_after:wN \_\_sort_level:
3475 \fi:
3476 }

(End definition for \_\_sort_level.)

\_\_sort_merge_blocks: This function is called to merge a pair of blocks, starting at the last value of \_\_sort_end_int (end-point of the previous pair of blocks). If shifting by one block to the right we reach the end of the list, then this pass has ended: the end of the list is sorted already. Otherwise, store the result of that shift in A, which indexes the first block starting from the top end. Then locate the end-point (maximum) of the second block: shift end upwards by one more block, but keeping it \leq top. Copy this upper block of \toks registers in registers above length, indexed by C: this is covered by \_\_sort_copy_block:. Once this is done we are ready to do the actual merger using \_\_sort_merge_blocks_aux:, after shifting A, B and C so that they point to the largest index in their respective ranges rather than pointing just beyond those ranges. Of course, once that pair of blocks is merged, move on to the next pair.

3479 \cs_new_protected:Npn \_\_sort_merge_blocks:
3480 { \_\_sort_begin_int \_\_sort_end_int
3481 \tex_advance:D \_\_sort_end_int \_\_sort_block_int
3482 \if_int_compare:w \_\_sort_end_int < \_\_sort_top_int
3483 \_\_sort_A_int \_\_sort_end_int
3484 \tex_advance:D \_\_sort_end_int \_\_sort_block_int
3485 \if_int_compare:w \_\_sort_end_int > \_\_sort_top_int
3486 \_\_sort_end_int \_\_sort_top_int
3487 \_\_sort_end_int \_\_sort_top_int

414
\_\_sort\_copy\_block: We wish to store a copy of the “upper” block of $\text{toks}$ registers, ranging between the initial value of $\l_{\_\_sort\_B\_int}$ (included) and $\l_{\_\_sort\_end\_int}$ (excluded) into a new range starting at the initial value of $\l_{\_\_sort\_C\_int}$, namely $\l_{\_\_sort\_top\_int}$.

\begin{verbatim}
\cs_new_protected:Npn \_\_sort\_copy\_block:
  { 
  \tex_toks:D \l_{\_\_sort\_C\_int} \tex_toks:D \l_{\_\_sort\_B\_int} 
  \int_incr:N \l_{\_\_sort\_C\_int} 
  \int_incr:N \l_{\_\_sort\_B\_int} 
  \if_int_compare:w \l_{\_\_sort\_B\_int} = \l_{\_\_sort\_end\_int} 
  \use_i:nn 
  \fi:
  \_\_sort\_copy\_block:
  }
\end{verbatim}

(End definition for $\_\_sort\_copy\_block$:)

\_\_sort\_merge\_blocks\_aux: At this stage, the first block starts at $\l_{\_\_sort\_begin\_int}$, and ends at $\l_{\_\_sort\_A\_int}$, and the second block starts at $\l_{\_\_sort\_top\_int}$ and ends at $\l_{\_\_sort\_C\_int}$. The result of the merger is stored at positions indexed by $\l_{\_\_sort\_B\_int}$, which starts at $\l_{\_\_sort\_end\_int} - 1$ and decreases down to $\l_{\_\_sort\_begin\_int}$, covering the full range of the two blocks. In other words, we are building the merger starting with the largest values. The comparison function is defined to return either swapped or same. Of course, this means the arguments need to be given in the order they appear originally in the list.

\begin{verbatim}
\cs_new_protected:Npn \_\_sort\_merge\_blocks\_aux:
  { 
  \exp_after:wN \_\_sort\_compare:nn \exp_after:wN 
  \{ \tex_the:D \tex_toks:D \exp_after:wN \l_{\_\_sort\_A\_int} \exp_after:wN \} 
  \exp_after:wN \{ \tex_the:D \tex_toks:D \l_{\_\_sort\_C\_int} \} 
  \prg_do_nothing:
  \_\_sort\_return\_mark:w 
  \_\_sort\_return\_mark:w 
  \_\_sort\_mark 
  \_\_sort\_return\_none\_error:
  }
\end{verbatim}

(End definition for $\_\_sort\_merge\_blocks\_aux$:)
Each comparison should call `\sort_return_same:` or `\sort_return_swapped:` exactly once. If neither is called, `\__sort_return_none_error:` is called, since the `\return_mark` removes tokens until `\s__sort_mark`. If one is called, the `\return_mark` auxiliary removes everything except `\__sort_return_same:w` (or its `\swapped` analogue) followed by `\__sort_return_none_error:`. Finally if two or more are called, `\__sort_return_two_error:` ends up before any `\__sort_return_mark:w`, so that it produces an error.

```latex
\cs_new_protected:Npn \sort_return_same:
  \ifcsname __sort_return_swapped: \else \fi \sort_return_same: #1 \__sort_return_mark:w #2 \s__sort_mark
{\twelvetexthe:D \twelvetoks:D \l__sort_B_int \twelvetexthe:D \twelvetoks:D \l__sort_C_int \int_decr:N \l__sort_B_int}
\cs_new_protected:Npn \sort_return_swapped:
  \ifcsname __sort_return_swapped: \else \fi \sort_return_swapped: #1 \__sort_return_mark:w #2 \s__sort_mark
{\twelvetexthe:D \twelvetoks:D \l__sort_B_int \twelvetexthe:D \twelvetoks:D \l__sort_C_int \int_decr:N \l__sort_B_int}
\cs_new_protected:Npn \__sort_return_mark:w #1 \s__sort_mark { \ifcsname __sort_return_same:w \fi}
\cs_new_protected:Npn \__sort_return_same:w #1 \__sort_return_none_error:
{\twelvetexthe:D \twelvetoks:D \l__sort_B_int \twelvetexthe:D \twelvetoks:D \l__sort_C_int \int_decr:N \l__sort_B_int}
\cs_new_protected:Npn \__sort_return_two_error:
{\msg_error:nnxx { sort } { return-two }
 {\twelvetexthe:D \twelvetoks:D \l__sort_B_int }
 {\twelvetexthe:D \twelvetoks:D \l__sort_C_int }
}
(End definition for `\sort_return_same:` and others. These functions are documented on page 44.)
```

If the comparison function returns `same`, then the second argument fed to `\__sort_compare:nn` should remain to the right of the other one. Since we build the merger starting from the right, we copy that `\toks` register into the allotted range, then shift the pointers `B` and `C`, and go on to do one more step in the merger, unless the second block has been exhausted: then the remainder of the first block is already in the correct registers and we are done with merging those two blocks.

```
\cs_new_protected:Npn \__sort_return_same:w #1 \__sort_return_none_error:
{\twelvetexthe:D \twelvetoks:D \l__sort_B_int \twelvetexthe:D \twelvetoks:D \l__sort_C_int \int_decr:N \l__sort_B_int}

416
\_\_sort\_merge\_blocks\_aux:

\( \text{End definition for } \_\_\_sort\_return\_same:w. \)

\_\_sort\_return\_swapped:w

If the comparison function returns \texttt{swapped}, then the next item to add to the merger is the first argument, contents of the \texttt{toks} register \( A \). Then shift the pointers \( A \) and \( B \) to the left, and go for one more step for the merger, unless the left block was exhausted (\( A \) goes below the threshold). In that case, all remaining \texttt{toks} registers in the second block, indexed by \( C \), are copied to the merger by \_\_\_sort\_merge\_blocks\_end:.

\( \_\_\_sort\_merge\_blocks\_end: \)

This function’s task is to copy the \texttt{toks} registers in the block indexed by \( C \) to the merger indexed by \( B \). The end can equally be detected by checking when \( B \) reaches the threshold \( \texttt{begin} \), or when \( C \) reaches \( \texttt{top} \).

\( \_\_\_sort\_merge\_blocks\_end: \)

(End definition for \_\_\_sort\_merge\_blocks\_end:.)

43.5 Expandable sorting

Sorting expandably is very different from sorting and assigning to a variable. Since tokens cannot be stored, they must remain in the input stream, and be read through at every step. It is thus necessarily much slower (at best \( O(n^2 \ln n) \)) than non-expandable sorting functions (\( O(n \ln n) \)).

A prototypical version of expandable quicksort is as follows. If the argument has no item, return nothing, otherwise partition, using the first item as a pivot (argument \#4 of \_\_\_sort:nnNnn). The arguments of \_\_\_sort:nnNnn are 1. items less than \#4, 2. items greater or equal to \#4, 3. comparison, 4. pivot, 5. next item to test. If \#5 is the tail of
the list, call \texttt{\tl_sort:nN} on \#1 and on \#2, placing \#4 in between; \texttt{\use:ff} expands the parts to make \texttt{\tl_sort:nN} \texttt{f}-expandable. Otherwise, compare \#4 and \#5 using \#3. If they are ordered, place \#5 amongst the “greater” items, otherwise amongst the “lesser” items, and continue partitioning.

\begin{verbatim}
\cs_new:Npn \tl_sort:nN #1#2
{ \tl_if_blank:nF {#1} {
 \__sort:nnNnn { } { } #2
 #1 \q__sort_recursion_tail \q__sort_recursion_stop
}
}
\end{verbatim}

There are quite a few optimizations available here: the code below is less legible, but more than twice as fast.

In the simple version of the code, \texttt{\__sort:nnNnn} is called $O(n \log n)$ times on average (the number of comparisons required by the quicksort algorithm). Hence most of our focus is on optimizing that function.

The first speed up is to avoid testing for the end of the list at every call to \texttt{\__sort:nnNnn}. For this, the list is prepared by changing each \texttt{⟨item⟩} of the original token list into \texttt{⟨command⟩{⟨item⟩}}, just like sequences are stored. We arrange things such that the \texttt{⟨command⟩} is the \texttt{⟨conditional⟩} provided by the user: the loop over the \texttt{⟨prepared tokens⟩} then looks like

\begin{verbatim}
\cs_new:Npn \__sort_loop:wNn ... #6#7
{ #6 {⟨pivot⟩} {#7} ⟨loop big⟩ ⟨loop small⟩ ⟨extra arguments⟩ }
\__sort_loop:wNn ... ⟨prepared tokens⟩
⟨end-loop⟩ { }
\end{verbatim}

In this example, which matches the structure of \texttt{\__sort_quick_split_i:NnnnNn} and a few other functions below, the \texttt{\__sort_loop:wNn} auxiliary normally receives the user’s \texttt{⟨conditional⟩} as \#6 and an \texttt{⟨item⟩} as \#7. This is compared to the \texttt{⟨pivot⟩} (the argument \#5, not shown here), and the \texttt{⟨conditional⟩} leaves the \texttt{⟨loop big⟩} or \texttt{⟨loop small⟩} auxiliary, which both have the same form as \texttt{\__sort_loop:wNn}, receiving the next pair \texttt{⟨conditional⟩} \texttt{⟨item⟩} as \#6 and \#7. At the end, \#6 is the \texttt{⟨end-loop⟩} function, which terminates the loop.

The second speed up is to minimize the duplicated tokens between the \texttt{true} and \texttt{false} branches of the conditional. For this, we introduce two versions of \texttt{\__sort:nnNn},
which receive the new item as \#1 and place it either into the list \#2 of items less than the pivot \#4 or into the list \#3 of items greater or equal to the pivot.

\cs_new:Npn \__sort_i:nnnnNn \#1\#2\#3\#4\#5\#6
{\#5 \{\#4\} \{\#6\} \__sort_ii:nnnnNn \__sort_i:nnnnNn\{\#6\} \{\#2\} \{\#3\} \{\#4\}}
\cs_new:Npn \__sort_ii:nnnnNn \#1\#2\#3\#4\#5\#6
{\#5 \{\#4\} \__sort_ii:nnnnNn \__sort_i:nnnnNn\{\#6\} \{\#2\} \{\#3\} \{\#1\} \{\#4\}}

Note that the two functions have the form of \__sort_loop:wNn above, receiving as \#5 the conditional or a function to end the loop. In fact, the lists \#2 and \#3 must be made of pairs \langle conditional \rangle \{\langle item \rangle \}, so we have to replace \{\#6\} above by \{ \#5 \{\#6\} \}, and \{\#1\} by \#1. The actual functions have one more argument, so all argument numbers are shifted compared to this code.

The third speed up is to avoid \use:ff using a continuation-passing style: \__sort_quick_split:Nnnn expects a list followed by \s__sort_mark \{\langle code \rangle\}, and expands to \langle code \rangle \{sorted list\}. Sorting the two parts of the list around the pivot is done with

\__sort_quick_split:Nnnn \#2 ... \s__sort_mark
{\__sort_quick_split:Nnnn \#1 ... \s__sort_mark \{\langle code \rangle\}
{\langle pivot \rangle}}

Items which are larger than the \langle pivot \rangle are sorted, then placed after code that sorts the smaller items, and after the (braced) \langle pivot \rangle.

The fourth speed up is avoid the recursive call to \tl_sort:nN with an empty first argument. For this, we introduce functions similar to the \__sort_i:nnnnNn of the last example, but aware of whether the list of \langle conditional \rangle \{\langle item \rangle\} read so far that are less than the pivot, and the list of those greater or equal, are empty or not: see \__sort_quick_split:Nnnn and functions defined below. Knowing whether the lists are empty or not is useless if we do not use distinct ending codes as appropriate. The splitting auxiliaries communicate to the \langle end-loop \rangle function (that is initially placed after the “prepared” list) by placing a specific ending function, ignored when looping, but useful at the end. In fact, the \langle end-loop \rangle function does nothing but place the appropriate ending function in front of all its arguments. The ending functions take care of sorting non-empty sublists, placing the pivot in between, and the continuation before.

The final change in fact slows down the code a little, but is required to avoid memory issues: schematically, when \TeX encounters
\use:n \{ \use:n \{ \use:n \{ ... \} ... \} ... \}
the argument of the first \use:n is not completely read by the second \use:n, hence must remain in memory; then the argument of the second \use:n is not completely read when grabbing the argument of the third \use:n, hence must remain in memory, and so on. The memory consumption grows quadratically with the number of nested \use:n. In
practice, this means that we must read everything until a trailing \texttt{\_sort_stop} once in a while, otherwise sorting lists of more than a few thousand items would exhaust a typical \TeX's memory.

The code within the \texttt{\_sort_stop} sorts the list, leaving in most cases a leading \texttt{\_sort_stop}, which stops the expansion, letting the result be return within \texttt{\_sort_stop}. We filter out the case of a list with no item, which would otherwise cause problems. Then prepare the token list \texttt{\_sort_prepare:Nnnn} by inserting the conditional \texttt{\_sort_prepare_end:NNNnw} before each item. The \texttt{\_sort_prepare} auxiliary receives the conditional as \texttt{\_sort_prepare}, the prepared token list so far as \texttt{\_sort_prepare_end}, the next prepared item as \texttt{\_sort_prepare}, and the item after that as \texttt{\_sort_prepare}. The loop ends when \texttt{\_sort_prepare} contains \texttt{\_sort_end:nnTFNn}, then the \texttt{\_sort_end} auxiliary finds the prepared token list as \texttt{\_sort_end}. The scene is then set up for \texttt{\_sort_split:NnNn}, which sorts the prepared list and perform the post action placed after \texttt{\_sort_mark}, namely removing the trailing \texttt{\_sort_stop} and \texttt{\_sort_stop} and leaving \texttt{\_sort_stop} to stop f-expansion.

\begin{verbatim}
\cs_new:Npn \tl_sort:nN #1#2
\exp_not:f
{\tl_if_blank:nF {#1}
 { \_sort_prepare:Nnnn #2 { } { } \_sort_prepare #1 { \_sort_prepare_end:NNNnw } \_sort_stop}
}
\cs_new:Npn \_sort_prepare:Nnnn #1#2#3#4
{\prg_break: #4 \prg_break_point: \_sort_prepare_end:NNNnw #1 #2 #3 #4 \_sort_stop}
\cs_new:Npn \_sort_prepare_end:NNNnw #1#2#3#4#5 \_sort_stop
{ \_sort_split:NnNn \_sort_only_i:Nnnnn \_sort_split_ii:Nnnnn \_sort_split_i:Nnnnn \_sort_split_ii:Nnnnn }
\cs_new:Npn \_sort_split:NnNn \_sort_only_i:Nnnnn \_sort_split_ii:Nnnnn \_sort_split_i:Nnnnn \_sort_split_ii:Nnnnn
\cs_new:Npn \_sort_only_i:Nnnnn \_sort_split_ii:Nnnnn \_sort_split_i:Nnnnn \_sort_split_ii:Nnnnn
\end{verbatim}

\texttt{\_sort_split} splits the list, leaving in most cases a leading \texttt{\_sort_split}, which stops the expansion, letting the result be return within \texttt{\_sort_split}. We filter out the case of a list with no item, which would otherwise cause problems. Then prepare the token list \texttt{\_sort_prepare} by inserting the conditional \texttt{\_sort_prepare_end} before each item. The \texttt{\_sort_prepare} auxiliary receives the conditional as \texttt{\_sort_prepare}, the prepared token list so far as \texttt{\_sort_prepare_end}, the next prepared item as \texttt{\_sort_prepare}, and the item after that as \texttt{\_sort_prepare}. The loop ends when \texttt{\_sort_prepare} contains \texttt{\_sort_end}, then the \texttt{\_sort_end} auxiliary finds the prepared token list as \texttt{\_sort_end}. The scene is then set up for \texttt{\_sort_split}, which sorts the prepared list and perform the post action placed after \texttt{\_sort_mark}, namely removing the trailing \texttt{\_sort_stop} and \texttt{\_sort_stop} and leaving \texttt{\_sort_stop} to stop f-expansion.

\texttt{\_sort_only_i,\_sort_only_ii,\_sort_split_i and \_sort_split_ii} auxiliaries receive a useless first argument, the new item \texttt{\_sort_prepare} (that they append to either one of the next two arguments), the list \texttt{\_sort_prepare} of items less than the pivot, bigger items \texttt{\_sort_prepare}, the pivot \texttt{\_sort_prepare}, a \texttt{\_sort_end}, and a \texttt{\_sort_mark}. The \texttt{\_sort_split} is the user's \texttt{\_sort_message} except at the end of the list where it is \texttt{\_sort_end}. The comparison is applied to the \texttt{\_sort_split} and the \texttt{\_sort_message}, and calls the \texttt{\_sort_only_i} or \texttt{\_sort_split_i} auxiliaries if the \texttt{\_sort_message} is smaller, and the \texttt{\_sort_only_ii} or \texttt{\_sort_split_ii} auxiliaries otherwise. In both cases, the next auxiliary goes to work right away, with no intermediate expansion that would slow down operations. Note that the argument \texttt{\_sort_prepare} left for the next call has the form \texttt{(item)}, so that the lists \texttt{\_sort_prepare} and \texttt{\_sort_prepare} keep the right form to be fed to the next sorting function. The \texttt{\_sort_split} auxiliary

\end{verbatim}
differs from these in that it is missing three of the arguments, which would be empty, and its first argument is always the user’s (conditional) rather than an ending function.

\cs_new:Npn \__sort_quick_split:NnNn #1#2#3#4
\begin{verbatim}
{ #3 {#2} {#4} \__sort_quick_only_i:NnnnnNn
\__sort_quick_only_i:NnnnnNn
\__sort_quick_single_end:nnnvw
{ #3 {#4} } { } { } {#2} }
\end{verbatim}
\cs_new:Npn \__sort_quick_only_i:NnnnnNn #1#2#3#4#5#6#7
\begin{verbatim}
{ #6 {#5} {#7} \__sort_quick_split_ii:NnnnnNn
\__sort_quick_split_i:NnnnnNn
\__sort_quick_only_i_end:nnnvw
{ #6 {#7} } { #3 #2 } { } {#5} }
\end{verbatim}
\cs_new:Npn \__sort_quick_split_i:NnnnnNn #1#2#3#4#5#6#7
\begin{verbatim}
{ #6 {#5} {#7} \__sort_quick_split_ii:NnnnnNn
\__sort_quick_split_i:NnnnnNn
\__sort_quick_only_i_end:nnnvw
{ #6 {#7} } { } { #4 #2 } {#5} }
\end{verbatim}
\cs_new:Npn \__sort_quick_split_ii:NnnnnNn #1#2#3#4#5#6#7
\begin{verbatim}
{ #6 {#5} {#7} \__sort_quick_split_ii:NnnnnNn
\__sort_quick_split_i:NnnnnNn
\__sort_quick_split_end:nnnvw
{ #6 {#7} } {#3} { #4 #2 } {#5} }
\end{verbatim}
\cs_new:Npn \__sort_quick_split_ii:NnnnnNn #1#2#3#4#5#6#7
\begin{verbatim}
{ #6 {#5} {#7} \__sort_quick_split_ii:NnnnnNn
\__sort_quick_split_i:NnnnnNn
\__sort_quick_split_end:nnnvw
{ #6 {#7} } {#3} { #4 #2 } {#5} }
\end{verbatim}
\end{verbatim}
(End definition for \__sort_quick_split:NnNn and others.)
\begin{verbatim}
\__sort_quick_split_end:nnnFWn
\__sort_single_end:nnnvw
\__sort_only_i:nnnvw
\__sort_only_ii:nnnvw
\__sort_split_end:nnnvw
\end{verbatim}
The \__sort_quick_end:nnTFNn appears instead of the user’s conditional, and receives as its arguments the pivot #1, a fake item #2, a true and a false branches #3 and #4, followed by an ending function #5 (one of the four auxiliaries here) and another copy #6 of the fake item. All those are discarded except the function #5. This function receives lists #1 and #2 of items less than or greater than the pivot #3, then a continuation code #5 just after \s__sort_mark. To avoid a memory problem described earlier, all of the ending functions read #6 until \s__sort_stop and place #6 back into the input stream. When the lists #1 and #2 are empty, the single auxiliary simply places the continuation #5 before the pivot {#3}. When #2 is empty, #1 is sorted and placed before the pivot {#3}, taking care to feed the continuation #5 as a continuation for the function sorting #1. When #1 is empty, #2 is sorted, and the continuation argument is used to place the continuation #5 and the pivot {#3} before the sorted result. Finally, when both
lists are non-empty, items larger than the pivot are sorted, then items less than the pivot,
and the continuations are done in such a way to place the pivot in between.

43.6 Messages

\_\_sort\_error:
Bailing out of the sorting code is a bit tricky. It may not be safe to use a delimited
argument, so instead we redefine many l3sort commands to be trivial, with \_\_sort\_level:
jumping to the break point. This error recovery won’t work in a group.

While sorting, \texttt{\textbackslash tok}def is locally disabled to prevent users from using \texttt{\textbackslash newtok} or
similar commands in their comparison code: the \texttt{\textbackslash tok} registers that would be assigned
are in use by l3sort. In format mode, none of this is needed since there is no \texttt{\textbackslash tok} allocator.
\msg_error:nxx { sort } { toksdef }
\{ \token_to_str:N \ #1 \}
\__sort_error:
\tex_toksdef:D \ #1
\}
\msg_new:nnnn { sort } { toksdef }
\{ Allocation-of-\iow_char:N\toks-registers-impossible-while-sorting. \}
\{ The-comparison-code-used-for-sorting-a-list-has-attempted-to-
\ define-\ #1 as-a-new-\iow_char:N\toks-register-using-
\iow_char:N\newtoks-
\ or-a-similar-command.-The-list-will-not-be-sorted. \}
\}
(\ End definition for \_\_sort_disable_toksdef: and \_\_sort_disabled_toksdef:n. )

\__sort_too_long_error:NNw
When there are too many items in a sequence, this is an error, and we clean up properly
the mapping over items in the list: break using the type-specific breaking function \#1.
\cs_new_protected:Npn \__sort_too_long_error:NNw \ #1\ #2 \fi:
\{ \fi:
\msg_error:nxxx { sort } { too-large }
\{ \token_to_str:N \ #2 \}
\{ \int_eval:n \{ \l__sort_true_max_int - \l__sort_min_int \} \}
\{ \int_eval:n \{ \l__sort_top_int - \l__sort_min_int \} \}
\#1 \__sort_error:
\}
\msg_new:nnnn { sort } { too-large }
\{ The-list-\#1-is-too-long-to-be-sorted-by-TeX. \}
\{ TeX-has-\#2-toks-registers-still-available:-
\ this-only-allows-to-sort-with-up-to-\#3-
\ items.-The-list-will-not-be-sorted. \}
\}
(\ End definition for \_\_sort_too_long_error:NNw. )
\msg_new:nnnn { sort } { return-none }
\{ The-comparison-code-did-not-return. \}
\{ When-sorting-a-list,-the-code-to-compare-items-\#1-and-\#2-
did-not-call-
\iow_char:N\sort_return_same: -nor-
\iow_char:N\sort_return_swapped: -.
\ Exactly-one-of-these-should-be-called. \}
\msg_new:nnnn { sort } { return-two }
\{ The-comparison-code-returned-multiple-times. \}
\{ When-sorting-a-list,-the-code-to-compare-items-\#1-and-\#2-called-
\iow_char:N\sort_return_same: -or-
\iow_char:N\sort_return_swapped: -multiple-times.-
\ Exactly-one-of-these-should-be-called. \}
\prop_gput:Nnn \g_msg_module_name_prop { sort } { LaTeX3 }
\prop_gput:Nnn \g_msg_module_type_prop { sort } { }
423
3777 ⟨/package⟩
Chapter 44

l3tl-analysis implementation

44.1 Internal functions

\s__tl  The format used to store token lists internally uses the scan mark \s__tl as a delimiter.

(End definition for \s__tl.)

44.2 Internal format

The task of the l3tl-analysis module is to convert token lists to an internal format which allows us to extract all the relevant information about individual tokens (category code, character code), as well as reconstruct the token list quickly. This internal format is used in l3regex where we need to support arbitrary tokens, and it is used in conversion functions in l3str-convert, where we wish to support clusters of characters instead of single tokens.

We thus need a way to encode any \langle token\rangle (even begin-group and end-group character tokens) in a way amenable to manipulating tokens individually. The best we can do is to find \langle tokens\rangle which both \o-expand and \x-expand to the given \langle token\rangle. Collecting more information about the category code and character code is also useful for regular expressions, since most regexes are catcode-agnostic. The internal format thus takes the form of a succession of items of the form

\langle tokens\rangle \s__tl \langle catcode\rangle \langle char code\rangle \s__tl

The \langle tokens\rangle \o- and \x-expand to the original token in the token list or to the cluster of tokens corresponding to one Unicode character in the given encoding (for l3str-convert). The \langle catcode\rangle is given as a single hexadecimal digit, 0 for control sequences. The \langle char code\rangle is given as a decimal number, −1 for control sequences.

Using delimited arguments lets us build the \langle tokens\rangle progressively when doing an encoding conversion in l3str-convert. On the other hand, the delimiter \s__tl may not appear unbraced in \langle tokens\rangle. This is not a problem because we are careful to wrap control sequences in braces (as an argument to \exp_not:n) when converting from a general token list to the internal format.

The current rule for converting a \langle token\rangle to a balanced set of \langle tokens\rangle which both \o-expands and \x-expands to it is the following.
A control sequence `{\cs}` becomes \exp_not:n { \cs } \s__tl 0 -1 \s__tl.

A begin-group character `{` becomes \exp_after:wN { \if_false: } \fi: \s__tl 1 ⟨char code⟩ \s__tl.

An end-group character `}` becomes \if_false: { \fi: } \s__tl 2 ⟨char code⟩ \s__tl.

A character with any other category code becomes \exp_not:n ⟨character⟩ \s__tl ⟨hex catcode⟩ ⟨char code⟩ \s__tl.

### 44.3 Variables and helper functions

\s__tl

The scan mark \s__tl is used as a delimiter in the internal format. This is more practical than using a quark, because we would then need to control expansion much more carefully: compare \int_value:w '#1 \s__tl with \int_value:w '#1 \exp_stop_f: \exp_not:N \q_mark to extract a character code followed by the delimiter in an x-expansion.

\l__tl_analysis_token \l__tl_analysis_char_token \l__tl_analysis_next_token

The tokens in the token list are probed with the \TeX primitive \futurelet. We use \l__tl_analysis_token in that construction. In some cases, we convert the following token to a string before probing it: then the token variable used is \l__tl_analysis_char_token. When getting tokens from the input stream we may need to look two tokens ahead, for which we use \l__tl_analysis_next_token.

\l__tl_peek_code_tl

Holds some code to be run once the next token has been fully analysed in \peek_analysis_map_inline:n.

\c__tl_peek_catcodes_tl

A token list containing the character number 32 (space) with all possible category codes except 1 and 2 (begin-group and end-group). Why 32? Because some Lua\TeX versions only allow creation of catcode 10 (space) tokens with this character code, and because even in other engines it is much easier to produce since \char_generate:nn refuses to produce spaces.

\c__tl_peek_catcodes_tl

\c__tl_peek_catcodes_tl

\c__tl_peek_catcodes_tl
\__tl_analysis_normal_int  The number of normal (N-type argument) tokens since the last special token.
\int_new:N \l__tl_analysis_normal_int
(End definition for \l__tl_analysis_normal_int.)

\l__tl_analysis_index_int  During the first pass, this is the index in the array being built. During the second pass, it is equal to the maximum index in the array from the first pass.
\int_new:N \l__tl_analysis_index_int
(End definition for \l__tl_analysis_index_int.)

\l__tl_analysis_nesting_int Nesting depth of explicit begin-group and end-group characters during the first pass. This lets us detect the end of the token list without a reserved end-marker.
\int_new:N \l__tl_analysis_nesting_int
(End definition for \l__tl_analysis_nesting_int.)

\l__tl_analysis_type_int  When encountering special characters, we record their “type” in this integer.
\int_new:N \l__tl_analysis_type_int
(End definition for \l__tl_analysis_type_int.)

\g__tl_analysis_result_tl  The result of the conversion is stored in this token list, with a succession of items of the form
\langle tokens \rangle \langle tokens \rangle \langle catcode \rangle \langle char code \rangle
\tl_new:N \g__tl_analysis_result_tl
(End definition for \g__tl_analysis_result_tl.)

\__tl_analysis_extract_charcode: \__tl_analysis_extract_charcode_aux:w
Extracting the character code from the meaning of \l__tl_analysis_token. This has no error checking, and should only be assumed to work for begin-group and end-group character tokens. It produces a number in the form ‘\langle char \rangle’.
\cs_new:Npn \__tl_analysis_extract_charcode:
\{ \exp_after:wN \__tl_analysis_extract_charcode_aux:w \token_to_meaning:N \l__tl_analysis_token \}
\cs_new:Npn \__tl_analysis_extract_charcode_aux:w #1 - #2 - { ‘
(End definition for \__tl_analysis_extract_charcode: and \__tl_analysis_extract_charcode_aux:w.)
Counts the number of spaces in the string representation of its second argument, as well as the number of characters following the last space in that representation, and feeds the two numbers as semicolon-delimited arguments to the first argument. When this function is used, the escape character is printable and non-space.

\cs_new:Npn \__tl_analysis_cs_space_count:NN \#1 \#2
\{ \exp_after:wN \#1 \int_value:w \int_eval:w 0 \exp_after:wN \__tl_analysis_cs_space_count:w \token_to_str:N \#2 \fi: \__tl_analysis_cs_space_count_end:w ; - ! \}
\cs_new:Npn \__tl_analysis_cs_space_count:w \#1 ~
\{ \if_f\_false: \#1 \#1 \fi: + 1 \__tl_analysis_cs_space_count:w \}
\cs_new:Npn \__tl_analysis_cs_space_count_end:w \#1 \fi: \#2 !
\{ \exp_after:wN \int_value:w \str_count_ignore_spaces:n {\#1} \}

(End definition for \__tl_analysis_cs_space_count:NN, \__tl_analysis_cs_space_count:w, and \__tl_analysis_cs_space_count_end:w.)

44.4 Plan of attack

Our goal is to produce a token list of the form roughly

\langle token 1 \rangle \s__tl \langle catcode 1 \rangle \langle char code 1 \rangle \s__tl
\langle token 2 \rangle \s__tl \langle catcode 2 \rangle \langle char code 2 \rangle \s__tl
... \langle token \ N \rangle \s__tl \langle catcode \ N \rangle \langle char code \ N \rangle \s__tl

Most but not all tokens can be grabbed as an undelimited (N-type) argument by \TeX. The plan is to have a two pass system. In the first pass, locate special tokens, and store them in various \toks registers. In the second pass, which is done within an x-expanding assignment, normal tokens are taken in as N-type arguments, and special tokens are retrieved from the \toks registers, and removed from the input stream by some means. The whole process takes linear time, because we avoid building the result one item at a time.

We make the escape character printable (backslash, but this later oscillates between slash and backslash): this allows us to distinguish characters from control sequences.

A token has two characteristics: its \texttt{meaning}, and what it looks like for \TeX when it is in scanning mode (e.g., when capturing parameters for a macro). For our purposes, we distinguish the following meanings:

- begin-group token (category code 1), either space (character code 32), or non-space;
- end-group token (category code 2), either space (character code 32), or non-space;
- space token (category code 10, character code 32);
- anything else (then the token is always an N-type argument).

The token itself can “look like” one of the following
• a non-active character, in which case its meaning is automatically that associated
to its character code and category code, we call it “true” character;
• an active character;
• a control sequence.

The only tokens which are not valid N-type arguments are true begin-group characters,
true end-group characters, and true spaces. We detect those characters by scanning
ahead with \futurelet, then distinguishing true characters from control sequences set
equal to them using the \string representation.

The second pass is a simple exercise in expandable loops.

\_\_tl_analysis:n

Everything is done within a group, and all definitions are local. We use \group_align_-
safe_begin/end: to avoid problems in case \_\_tl_analysis:n is used within an align-
ment and its argument contains alignment tab tokens.

\cs_new_protected:Npn \_\_tl_analysis:n #1
\group_begin:
\group_align_safe_begin:
\_\_tl_analysis_a:n {#1}
\_\_tl_analysis_b:n {#1}
\group_align_safe_end:
\group_end:

(End definition for \_\_tl_analysis:n.)

### 44.5 Disabling active characters

\_\_tl_analysis_disable:n

Active characters can cause problems later on in the processing, so we provide a way to
disable them, by setting them to undefined. Since Unicode contains too many characters
to loop over all of them, we instead do this whenever we encounter a character. For \TeX\xspace
and \up\TeX\xspace we skip characters beyond \([0,255]\) because \lccode only allows those values.

\cs_new_protected:Npn \_\_tl_analysis_disable:n #1
\group_begin:
\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N
\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N \^^@\char_set_catcode_active:N
\cs_new_protected:Npn \_\_tl_analysis_disable:n #1
\group_begin:
\group_align_safe_begin:
\_\_tl_analysis_a:n {#1}
\_\_tl_analysis_b:n {#1}
\group_align_safe_end:
\group_end:

(End definition for \_\_tl_analysis_disable:n.)

\cs_gset_protected:Npn \_\_tl_analysis_disable:n #1
\if_int_compare:w 256 > #1 \exp_stop_f:
\tex_lccode:D 0 = #1 \exp_stop_f:
\tex_lowercase:D { \tex_let:D \tex_let:D \tex_let:D 0 } \tex_undefined:D
\fi:
\group_end:
44.6 First pass

The goal of this pass is to detect special (non-N-type) tokens, and count how many N-type tokens lie between special tokens. Also, we wish to store some representation of each special token in a \toks register.

We have 11 types of tokens:

1. a true non-space begin-group character;
2. a true space begin-group character;
3. a true non-space end-group character;
4. a true space end-group character;
5. a true space blank space character;
6. an active character;
7. any other true character;
8. a control sequence equal to a begin-group token (category code 1);
9. a control sequence equal to an end-group token (category code 2);
10. a control sequence equal to a space token (character code 32, category code 10);
11. any other control sequence.

Our first tool is \futurelet. This cannot distinguish case 8 from 1 or 2, nor case 9 from 3 or 4, nor case 10 from case 5. Those cases are later distinguished by applying the \string primitive to the following token, after possibly changing the escape character to ensure that a control sequence’s string representation cannot be mistaken for the true character.

In cases 6, 7, and 11, the following token is a valid N-type argument, so we grab it and distinguish the case of a character from a control sequence: in the latter case, \str_tail:n \{\langle token\rangle\} is non-empty, because the escape character is printable.

We read tokens one by one using \futurelet. While performing the loop, we keep track of the number of true begin-group characters minus the number of true end-group characters in \_\_tl_analysis_nesting_int. This reaches −1 when we read the closing brace.

---

(End definition for \_\_tl_analysis_disable:n.)
Read one character and check its type.

\begin{verbatim}
\cs_new_protected:Npn \_\_tl_analysis_a_loop:w
{ \tex_futurelet:D \_\_tl_analysis_token \_\_tl_analysis_a_type:w }
\end{verbatim}

At this point, \_\_tl_analysis_token holds the meaning of the following token. We store in \_\_tl_analysis_type_int information about the meaning of the token ahead:

- 0 space token;
- 1 begin-group token;
- -1 end-group token;
- 2 other.

The values 0, 1, -1 correspond to how much a true such character changes the nesting level (2 is used only here, and is irrelevant later). Then call the auxiliary for each case. Note that nesting conditionals here is safe because we only skip over \_\_tl_analysis_token if it matches with one of the character tokens (hence is not a primitive conditional).

\begin{verbatim}
\cs_new_protected:Npn \_\_tl_analysis_a_type:w
{ \l__tl_analysis_type_int = \if_meaning:w \l__tl_analysis_token \c_space_token 0 \else: \if_catcode:w \exp_not:N \l__tl_analysis_token \c_group_begin_token 1 \else: \if_catcode:w \exp_not:N \l__tl_analysis_token \c_group_end_token - 1 \else: \fi: \fi: \exp_stop_f: \if_case:w \l__tl_analysis_type_int \exp_after:wN \_\_tl_analysis_a_space:w \or: \exp_after:wN \_\_tl_analysis_a_bgroup:w \or: \exp_after:wN \_\_tl_analysis_a_safe:N \else: \exp_after:wN \_\_tl_analysis_a_egroup:w \fi: \fi: \fi: \exp_stop_f: \if_case:w \l__tl_analysis_type_int \exp_after:wN \_\_tl_analysis_a_space:w \or: \exp_after:wN \_\_tl_analysis_a_bgroup:w \or: \exp_after:wN \_\_tl_analysis_a_safe:N \else: \exp_after:wN \_\_tl_analysis_a_egroup:w \fi: \exp_stop_f: }
\end{verbatim}

In this branch, the following token's meaning is a blank space. Apply \texttt{string} to that token: a true blank space gives a space, a control sequence gives a result starting with the escape character, an active character gives something else than a space since we disabled the space. We grab as \_\_tl_analysis_char_token the first character of the string representation then test it in \_\_tl_analysis_a_space_test:w. Also, since
\_tl_analysis_a_store: expects the special token to be stored in the relevant \toks register, we do that. The extra \exp_not:n is unnecessary of course, but it makes the treatment of all tokens more homogeneous. If we discover that the next token was actually a control sequence or an active character instead of a true space, then we step the counter of normal tokens. We now have in front of us the whole string representation of the control sequence, including potential spaces; those will appear to be true spaces later in this pass. Hence, all other branches of the code in this first pass need to consider the string representation, so that the second pass does not need to test the meaning of tokens, only strings.

\cs_new_protected:Npn \_\_tl_analysis_a_space:w
\{ \tex_afterassignment:D \_\_tl_analysis_a_space_test:w
\exp_after:wN \cs_set_eq:NN \exp_after:wN \l__tl_analysis_char_token
\token_to_str:N 
\}
\cs_new_protected:Npn \_\_tl_analysis_a_space_test:w
\{ \if_meaning:w \l__tl_analysis_char_token \c_space_token
\tex_toks:D \l__tl_analysis_index_int \{ \exp_not:n {~} \}
\_\_tl_analysis_a_store:
\else:
\int_incr:N \l__tl_analysis_normal_int
\fi:
\_\_tl_analysis_a_loop:w
\}

(End definition for \_\_tl_analysis_a_space:w and \_\_tl_analysis_a_space_test:w.)

\_\_tl_analysis_a_bgroup:w
\_\_tl_analysis_a_egroup:w
\_\_tl_analysis_a_group:nw
\_\_tl_analysis_a_group_aux:w
\_\_tl_analysis_a_group_auxii:w
\_\_tl_analysis_a_group_test:w

The token is most likely a true character token with catcode 1 or 2, but it might be a control sequence, or an active character. Optimizing for the first case, we store in a toks register some code that expands to that token. Since we will turn what follows into a string, we make sure the escape character is different from the current character code (by switching between solidus and backslash). To detect the special case of an active character let to the catcode 1 or 2 character with the same character code, we disable the active character with that character code and re-test: if the following token has become undefined we can in fact safely grab it. We are finally ready to turn what follows to a string and test it. This is one place where we need \l__\_tl_analysis_char_token to be a separate control sequence from \l__\_tl_analysis_token, to compare them.

\group_begin:
\char_set_catcode_group_begin:N \^^@ % \{ \_\_tl_analysis_a_bgroup:w
\cs_new_protected:Npn \_\_tl_analysis_a_bgroup:w
\{ \_\_tl_analysis_a_group:nw \{ \exp_after:wN \^^@ \if_false: \fi: \} \fi: \} \}
\char_set_catcode_group_end:N \^^@ \{ \_\_tl_analysis_a_group:nw \{ \if_false: \{ \fi: \^^@ \} \} \%
\group_end:
\cs_new_protected:Npn \_\_tl_analysis_a_group:nw \#1
\{ \tex_lccode:D 0 = \_\_tl_analysis_extract_charcode: \scan_stop:
\tex_lowercase:D \{ \tex_toks:D \_\_tl_analysis_index_int \{#1\} \}
\if_int_compare:w \tex_lccode:D 0 = \tex_escapechar:D \int_set:Nn \tex_escapechar:D \{ 139 - \tex_escapechar:D \}

432
\__tl_analysis_disable:n { \tex_lccode:D 0 }
\tex_futurelet:D \l__tl_analysis_token \__tl_analysis_a_group_aux:w
\cs_new_protected:Npn \__tl_analysis_a_group_aux:w
{\if_meaning:w \l__tl_analysis_token \tex_undefined:D
\exp_after:wN \__tl_analysis_a_safe:N
\else:
\exp_after:wN \__tl_analysis_a_group_auxii:w
\fi:
}\cs_new_protected:Npn \__tl_analysis_a_group_auxii:w
{\tex_afterassignment:D \__tl_analysis_a_group_test:w
\exp_after:wN \cs_set_eq:NN \exp_after:wN \l__tl_analysis_char_token
\token_to_str:N
}\cs_new_protected:Npn \__tl_analysis_a_group_test:w
{\if_charcode:w \l__tl_analysis_token \l__tl_analysis_char_token
\__tl_analysis_a_store: \else:
\int_incr:N \l__tl_analysis_normal_int
\fi:
\__tl_analysis_a_loop:w
}\__tl_analysis_a_store: This function is called each time we meet a special token; at this point, the \toks register \l__tl_analysis_index_int holds a token list which expands to the given special token. Also, the value of \l__tl_analysis_type_int indicates which case we are in:

-1 end-group character;
0 space character;
1 begin-group character.

We need to distinguish further the case of a space character (code 32) from other character codes, because those behave differently in the second pass. Namely, after testing the \lccode of 0 (which holds the present character code) we change the cases above to

-2 space end-group character;
-1 non-space end-group character;
0 space blank space character;
1 non-space begin-group character;
2 space begin-group character.

(End definition for \__tl_analysis_a_bgroup:w and others.)
This has the property that non-space characters correspond to odd values of \l__tl_analysis_type_int. The number of normal tokens until here and the type of special token are packed into a \skip register. Finally, we check whether we reached the last closing brace, in which case we stop by disabling the looping function (locally).

\cs_new_protected:Npn \__tl_analysis_a_store:
\{ \tex_advance:D \l__tl_analysis_nesting_int \l__tl_analysis_type_int
\if_int_compare:w \tex_lccode:D 0 = \char_32 \exp_stop_f:
\tex_advance:D \l__tl_analysis_type_int \l__tl_analysis_type_int
\fi:
\tex_skip:D \l__tl_analysis_index_int
= \l__tl_analysis_normal_int sp
\plus \l__tl_analysis_type_int sp \scan_stop:
\int_incr:N \l__tl_analysis_index_int
\int_zero:N \l__tl_analysis_normal_int
\if_int_compare:w \l__tl_analysis_nesting_int = - \c_one_int
\cs_set_eq:NN \__tl_analysis_a_loop:w \scan_stop:
\fi:
\}
\(\text{End definition for \__tl_analysis_a_store:}\)

\__tl_analysis_a_safe:N \__tl_analysis_a_cs:ww
This should be the simplest case: since the upcoming token is safe, we can simply grab it in a second pass. If the token is a single character (including space), the \if_charcode:w test yields true; we disable a potentially active character (that could otherwise masquerade as the true character in the next pass) and we count one “normal” token. On the other hand, if the token is a control sequence, we should replace it by its string representation for compatibility with other code branches. Instead of slowly looping through the characters with the main code, we use the knowledge of how the second pass works: if the control sequence name contains no space, count that token as a number of normal tokens equal to its string length. If the control sequence contains spaces, they should be registered as special characters by increasing \l__tl_analysis_index_int (no need to carefully count character between each space), and all characters after the last space should be counted in the following sequence of “normal” tokens.

\cs_new_protected:Npn \__tl_analysis_a_safe:N \__tl_analysis_a_cs:ww
\{ \if_charcode:w \scan_stop:
\exp_after:wN \use_none:n \token_to_str:N \__tl_analysis_a_safe:N \prg_do_nothing:
\scan_stop:
\exp_after:wN \use_i:nn \else:
\exp_after:wN \use_ii:nn \fi:
\}
\{ \__tl_analysis_disable:n \{ #1 \} \int_incr:N \l__tl_analysis_normal_int \}
\{ \__tl_analysis_cs_space_count:NN \__tl_analysis_a_cs:ww #1 \}
\__tl_analysis_a_loop:w
\}
\cs_new_protected:Npn \__tl_analysis_a_cs:ww #1; #2;
\}
44.7 Second pass

The second pass is an exercise in expandable loops. All the necessary information is stored in \skip and \toks registers.

```latex
\cs_new_protected:Npn \__tl_analysis_b:n #1
\__kernel_tl_gset:Nx \g__tl_analysis_result_tl
\__tl_analysis_b_loop:w 0; #1
\prg_break_point:
\endgroup
\cs_new:Npn \__tl_analysis_b_loop:w #1;
\exp_after:wN \__tl_analysis_b_normals:ww
\int_value:w \tex_skip:D #1 ; #1 ;
\exp_not:n { \exp_not:n { #3 } } \s__tl
```

(End definition for \_\_tl_analysis_a_safe:N and \_\_tl_analysis_a_cs:ww.)

The first argument is the number of normal tokens which remain to be read, and the second argument is the index in the array produced in the first step. A character’s string representation is always one character long, while a control sequence is always longer (we have set the escape character to a printable value). In both cases, we leave \exp_not:n \{\token\} \s__tl in the input stream (after x-expansion). Here, \exp_not:n is used rather than \exp_not:N because \#3 could be a macro parameter character or could be \s__tl (which must be hidden behind braces in the result).

```
\cs_new:Npn \__tl_analysis_b_normals:ww #1;
\exp_after:wN \__tl_analysis_b_normal:wwN
\int_value:w \tex_skip:D #1 ; #1 ;
\exp_not:n { \exp_not:n { #3 } } \s__tl
\endgroup
```

(End definition for \_\_tl_analysis_b:n and \_\_tl_analysis_b_loop:w.)
\if_charcode:w
\scan_stop:
 \exp_after:wN \use_none:n \token_to_str:N #3 \prg_do_nothing:
 \scan_stop:
 \exp_after:wN \__tl_analysis_b_normals:ww
\else:
 \exp_after:wN \__tl_analysis_b_normal:wwN
\fi:
 #3 #1; #2;
\}

(End definition for \__tl_analysis_b_normals:ww and \__tl_analysis_b_normal:ww.)

\__tl_analysis_b_char:Nww

If the normal token we grab is a character, leave \langle catcode \rangle \langle charcode \rangle followed by \s\__tl in the input stream, and call \__tl_analysis_b_normals:ww with its first argument decremented.

\cs_new:Npx \__tl_analysis_b_char:Nww #1
{
 \exp_not:N \if_meaning:w #1 \exp_not:N \tex_undefined:D
 \token_to_str:N D \exp_not:N \else:
 \exp_not:N \if_catcode:w #1 \c_catcode_other_token
 \token_to_str:N C \exp_not:N \else:
 \exp_not:N \if_catcode:w #1 \c_catcode_letter_token
 \token_to_str:N B \exp_not:N \else:
 \exp_not:N \if_catcode:w #1 \c_math_toggle_token
 \token_to_str:N 3 \exp_not:N \else:
 \exp_not:N \if_catcode:w #1 \c_alignment_token
 \token_to_str:N 4 \exp_not:N \else:
 \exp_not:N \if_catcode:w #1 \c_math_superscript_token
 \token_to_str:N 7 \exp_not:N \else:
 \exp_not:N \if_catcode:w #1 \c_math_subscript_token
 \token_to_str:N 8 \exp_not:N \else:
 \exp_not:N \if_catcode:w #1 \c_space_token
 \token_to_str:N A \exp_not:N \else:
 6
 \exp_not:n { \fi: \fi: \fi: \fi: \fi: \fi: \fi: }
 \exp_not:N \int_value:w '#1 \s\__tl
 \__tl_analysis_cs_space_count:NN \__tl_analysis_b_cs_test:ww #1
 \exp_not:N \int_value:w \exp_not:N \int_eval:w - 1 +
\}

(End definition for \__tl_analysis_b_char:Nww.)

\__tl_analysis_b_cs:Nww
\__tl_analysis_b_cs_test:ww

If the token we grab is a control sequence, leave 0 -1 (as category code and character code) in the input stream, followed by \s\__tl, and call \__tl_analysis_b_normals:ww with updated arguments.

\cs_new:Npn \__tl_analysis_b_cs:Nww #1
{
 0 -1 \s\__tl
 \__tl_analysis_cs_space_count:NN \__tl_analysis_b_cs_test:ww #1
}
\cs_new:Npn \__tl_analysis_b_cs_test:ww #1 ; #2 ; #3 ; #4 ;
{
 \exp_after:wN \__tl_analysis_b_normals:ww
\}

436
Here, \texttt{#1} is the current index in the array built in the first pass. Check now whether we reached the end (we shouldn't keep the trailing end-group character that marked the end of the token list in the first pass). Unpack the \texttt{toks} register: when \texttt{x}-expanding again, we will get the special token. Then leave the category code in the input stream, followed by the character code, and call \texttt{\_\_tl_analysis_b_loop:w} with the next index.

\begin{verbatim}
\__tl_analysis_b_special:w
\_tl_analysis_b_special_char:w
\_tl_analysis_b_special_space:w

\if_int_compare:w \#1 = \c_zero_int
\else:
  \tex_skip:D \int_eval:n { \#4 + \#1 } \exp_stop_f:
\fi:
- \#2
\exp_after:wN ;
\int_value:w \int_eval:n { \#4 + \#1 } ;
\}
\end{verbatim}

(End definition for \texttt{\_\_tl_analysis_b_cs:Nww} and \texttt{\_\_tl_analysis_b_cs_test:ww}.)

\begin{verbatim}
\__tl_analysis_b_special:w
\_tl_analysis_b_special_char:w
\_tl_analysis_b_special_space:w

\if_int_compare:w \#1 = \l__tl_analysis_index_int
\exp_after:wN \prg_break:
\fi:
\tex_the:D \tex_toks:D \#1 \s__tl
\if_case:w \tex_gluestretch:D \tex_skip:D \#1 \exp_stop_f:
\token_to_str:N A
\or: 1
\or: 1
\else: 2
\fi:
\if_int_odd:w \tex_gluestretch:D \tex_skip:D \#1 \exp_stop_f:
\exp_after:wN \__tl_analysis_b_special_char:wN \int_value:w
\else:
  \exp_after:wN \__tl_analysis_b_special_space:w \int_value:w
\fi:
\int_eval:n { 1 + \#1 } \exp_after:wN ;
\token_to_str:N
\}
\end{verbatim}

\begin{verbatim}
\__tl_analysis_b_special_char:wN \#1 ; \#2
\__tl_analysis_b_special_space:w \#1 ;
\}
\end{verbatim}

\begin{verbatim}
\char_set_catcode_other:N A
\cs_new:Npn \__tl_analysis_b_special:w
\fi: \__tl_analysis_b_normal:wwN 0 ; \#1 ;
{ }
\fi:
\if_int_compare:w \#1 = \l__tl_analysis_index_int
\exp_after:wN \prg_break:
\fi:
\tex_the:D \tex_toks:D \#1 \s__tl
\if_case:w \tex_gluestretch:D \tex_skip:D \#1 \exp_stop_f:
\token_to_str:N A
\or: 1
\or: 1
\else: 2
\fi:
\if_int_odd:w \tex_gluestretch:D \tex_skip:D \#1 \exp_stop_f:
\exp_after:wN \__tl_analysis_b_special_char:wN \int_value:w
\else:
  \exp_after:wN \__tl_analysis_b_special_space:w \int_value:w
\fi:
\int_eval:n { 1 + \#1 } \exp_after:wN ;
\token_to_str:N
\}
\end{verbatim}

\begin{verbatim}
\__tl_analysis_b_special_space:w \#1 ; ~
\}
\cs_new:Npn \__tl_analysis_b_special_char:wN \#1 ; \#2
{ \int_value:w \#2 \s__tl
  \__tl_analysis_b_loop:w \#1 ;
}
\cs_new:Npn \__tl_analysis_b_special_char:wN \#1 ;
{ 32 \s__tl
  \__tl_analysis_b_loop:w \#1 ;
}
\end{verbatim}

\begin{verbatim}
\group_begin:
\char_set_catcode_other:N A
\cs_new:Npn \__tl_analysis_b_special:w
\fi: \__tl_analysis_b_normal:wwN 0 ; \#1 ;
{ }
\fi:
\if_int_compare:w \#1 = \l__tl_analysis_index_int
\exp_after:wN \prg_break:
\fi:
\tex_the:D \tex_toks:D \#1 \s__tl
\if_case:w \tex_gluestretch:D \tex_skip:D \#1 \exp_stop_f:
\token_to_str:N A
\or: 1
\or: 1
\else: 2
\fi:
\if_int_odd:w \tex_gluestretch:D \tex_skip:D \#1 \exp_stop_f:
\exp_after:wN \__tl_analysis_b_special_char:wN \int_value:w
\else:
  \exp_after:wN \__tl_analysis_b_special_space:w \int_value:w
\fi:
\int_eval:n { 1 + \#1 } \exp_after:wN ;
\token_to_str:N
\}
\end{verbatim}

\begin{verbatim}
\__tl_analysis_b_special_char:wN \#1 ; \#2
\__tl_analysis_b_special_space:w \#1 ;
\}
\cs_new:Npn \__tl_analysis_b_special_char:wN \#1 ; \#2
{ \int_value:w \#2 \s__tl
  \__tl_analysis_b_loop:w \#1 ;
}
\cs_new:Npn \__tl_analysis_b_special_space:w \#1 ; ~
{ 32 \s__tl
  \__tl_analysis_b_loop:w \#1 ;
}
\end{verbatim}

437
44.8 Mapping through the analysis

First obtain the analysis of the token list into \texttt{\_\_tl_analysis_result_tl}. To allow nested mappings, increase the nesting depth \texttt{\_\_kernel_prg_map_int} (shared between all modules), then define the looping macro, which has a name specific to that nesting depth. That looping grabs the \texttt{\langle tokens\rangle}, \texttt{\langle catcode\rangle} and \texttt{\langle char code\rangle}; it checks for the end of the loop with \texttt{\use_none:n ##2}, normally empty, but which becomes \texttt{\tl_map_break:} at the end; it then performs the user’s code \texttt{#2}, and loops by calling itself. When the loop ends, remember to decrease the nesting depth.

```
\cs_new_protected:Npn \tl_analysis_map_inline:nn #1  
  { \__tl_analysis:n {#1} \int_gincr:N \g__kernel_prg_map_int  
    \exp_args:Nc \__tl_analysis_map_inline_aux:Nn 
    { \__tl_analysis_map_inline_ \int_use:N \g__kernel_prg_map_int :wNw }  
    \cs_new_protected:Npn \tl_analysis_map_inline:Nn #1  
    { \exp_args:No \tl_analysis_map_inline:nn #1 }  
    \cs_new_protected:Npn \__tl_analysis_map_inline_aux:Nn #1#2#3  
    { \cs_gset_protected:Npn #1 ##1 \s__tl ##2 ##3 \s__tl  
      { \use_none:n ##2  
        \__tl_analysis_map_inline_aux:nnn {##1} {##3} {##2} }  
      \cs_gset_protected:Npn \__tl_analysis_map_inline_aux:nnn ##1##2##3  
      { #2 #1 }  
      \exp_after:wN #1  
    }  
  }  
```

(End definition for \texttt{\_\_tl_analysis_map_inline:nn} and others. These functions are documented on page 45.)

44.9 Showing the results

Add to \texttt{\_\_tl_analysis:n} a third pass to display tokens to the terminal. If the token list variable is not defined, throw the same error as \texttt{\tl_show:N} by simply calling that function.

```
\cs_new_protected:Npn \tl_analysis_show:N  
  { \__tl_analysis_show:NNN \msg_show:nnxxxx  \tl_show:N }  
\cs_new_protected:Npn \tl_analysis_log:N  
```

(End definition for \texttt{\_\_tl_analysis_map_inline:nn} and others. These functions are documented on page 45.)
\tl_analysis_show:N
\tl_analysis_log:N
\__tl_analysis_show:NNN

Here, \#1 o- and x-expands to the token; \#2 is the category code (one uppercase hexadecimal digit), 0 for control sequences; \#3 is the character code, which we ignore. In the cases of control sequences and active characters, the meaning may overflow one line, and we want to truncate it. Those cases are thus separated out.

(End definition for \tl_analysis_show:N, \tl_analysis_log:N, and \__tl_analysis_show:NNN. These functions are documented on page 45.)

\__tl_analysis_show:
\__tl_analysis_show_loop:wNw

\tl_analysis_show:n
\tl_analysis_log:n
\__tl_analysis_show:Nn

No existence test needed here.
\__tl_analysis_show_loop:wNw \\
(End definition for \__tl_analysis_show: and \__tl_analysis_show_loop:wNw.)

\__tl_analysis_show_normal:n Non-active characters are a simple matter of printing the character, and its meaning. Our test suite checks that begin-group and end-group characters do not mess up \TeX{}’s alignment status.

\cs_new:Npn \__tl_analysis_show_normal:n #1 \\
{ \exp_after:wN \token_to_str:N #1 ~ \\
( \exp_after:wN \token_to_meaning:N #1 ) }

(End definition for \__tl_analysis_show_normal:n.)

\__tl_analysis_show_value:N This expands to the value of \#1 if it has any.

\cs_new:Npn \__tl_analysis_show_value:N #1 \\
{ \token_if_expandable:NF #1 \\
{ \token_if_chardef:NTF #1 \prg_break: { } \\
\token_if_mathchardef:NTF #1 \prg_break: { } \\
\token_if_dim_register:NTF #1 \prg_break: { } \\
\token_if_int_register:NTF #1 \prg_break: { } \\
\token_if_skip_register:NTF #1 \prg_break: { } \\
\token_if_toks_register:NTF #1 \prg_break: { } \\
\use_none:nnn \\
\prg_break_point: \\
\use:n { \exp_after:wN = \tex_the:D #1 } \\
} }

(End definition for \__tl_analysis_show_value:N.)

\__tl_analysis_show_long:nn Control sequences and active characters are printed in the same way, making sure not to go beyond the \l_iow_line_count_int. In case of an overflow, we replace the last characters by \c__tl_analysis_show_etc_str.

\cs_new:Npn \__tl_analysis_show_cs:n #1 \\
{ \exp_args:No \__tl_analysis_show_long:nn {#1} { control~sequence= } }

\cs_new:Npn \__tl_analysis_show_active:n #1 \\
{ \exp_args:No \__tl_analysis_show_long:nn {#1} { active~character= } }

\cs_new:Npn \__tl_analysis_show_long:nn #1 \\
{ \__tl_analysis_show_long_aux:oofn \\
{ \token_to_str:N #1 } \\
{ \token_to_meaning:N #1 } \\
{ \__tl_analysis_show_value:N #1 } }

\cs_new:Npn \__tl_analysis_show_long_aux:nnnn #1#2#3#4 \\
{ \int_compare:nNnTF \\
{ \str_count:n { #1 ~ ( #4 #2 #3 ) } } \\
> { \l_iow_line_count_int - 3 } \\
> { \l_iow_line_count_int - 3 } }
44.10 Peeking ahead

The break statements use the general \prg_map_break:Nn.
\peek_analysis_map_break:
\peek_analysis_map_break:n

Save the user’s code in a control sequence that is suitable for nested maps. We may wish to pass to this function an \outer control sequence or active character; for this we will undefine potentially-\outer tokens within a group, closed after the function receives its arguments. This user’s code function also calls the loop auxiliary, and includes the trailing \prg_break_point:Nn for when the user wants to stop the loop. The loop auxiliary must remove that break point because it must look at the input stream.

\peek_analysis_map_inline:n
\peek_analysis_map_loop:NNn
\peek_analysis_map_test:
\peek_analysis_map_normal:N
\peek_analysis_map_cs:
\peek_analysis_map_char:N
\peek_analysis_map_special:
\peek_analysis_map_retest:
\peek_analysis_map_next:
\peek_analysis_map_str:
\peek_analysis_map_str:w
\peek_analysis_map_str:n
\peek_analysis_map_active_str:n
\peek_analysis_map_explicit:n
\peek_analysis_map_escape:
\peek_analysis_map_collect:w
\peek_analysis_map_collect:n
\peek_analysis_map_collect_loop:
\peek_analysis_map_collect_test:
The loop starts a group (closed by the user-code function defined above) with a normalized escape character, and checks if the next token is special or N-type.

Normal tokens are not too hard, but can be \outer, hence the \exp_not:N in the code above. If the token is expandable then it might be an \outer or a \TeX conditional, so to be safe we set it to \scan_stop: (the assignment is local and stopped by the \group_end: upon calling the user’s code). Then distinguish characters (including active ones and macro parameter characters) from control sequences (whose string representation is more than one character because the escape character is printable). For a control sequence call the user code with suitable arguments.
For special characters the idea is to eventually act with \token_to_str:N, then pick up one by one the characters of this string representation until hitting the token that follows. First determine the character code of (the meaning of) the ⟨token⟩ (which we know is a special token), make sure the escape character is different from it, normalize the meanings of two active characters and the empty control sequence, and filter out these cases in \_\_tl_peek_analysis_retest::
\cs_new_protected:Npn \_\_tl_peek_analysis_special:
{ \tex_let:D \l\_tl_analysis_token = ~ \l_peek_token \\
\int_set:Nn \l\_tl_peek_charcode_int
{ \_\_tl_analysis_extract_charcode: } \\
\if_int_compare:w \l\_tl_peek_charcode_int = \tex_escapechar:D
\int_set:Nn \tex_escapechar:D { '/' } \\
\fi:
\char_set_active_eq:nN { \l\_tl_peek_charcode_int } \scan_stop:
\char_set_active_eq:nN { \tex_escapechar:D } \scan_stop:
\cs_set_eq:cN { } \scan_stop:
\tex_futurelet:D \l\_tl_analysis_token \_\_tl_peek_analysis_retest:
\cs_new_protected:Npn \_\_tl_peek_analysis_retest:
{ \if_meaning:w \l\_tl_analysis_token \scan_stop:
\exp_after:wN \_\_tl_peek_analysis_normal:N
\else:
\exp_after:wN \_\_tl_peek_analysis_next:
\fi:
}
token, either a space (32, 10) or a begin-group or end-group token (catcode 1 or 2), and we excluded a few cases that would be difficult later (empty control sequence, active character with the same character code as its meaning or as the escape character). Now look at the (next token) following it using a combination of \afterassignment and \futurelet. The syntax of this primitive is \futurelet \langle peek token \rangle \langle first token \rangle \langle next token \rangle, and it sets \langle peek token \rangle equal to \langle next token \rangle. Traditionally, one takes \langle first token \rangle to be some macro that regains control of the code and, e.g., analyses \langle peek token \rangle. Here, both \langle first token \rangle and \langle next token \rangle are mostly unknown tokens in the input stream (but we know the \langle first token \rangle has catcode 1, 2 or 10), where \langle first token \rangle was already stored as \l_p_t_t_e_k_t_c_k_e_n, and we regain control using \afterassignment, which inserts its argument after the assignment, hence after \langle peek token \rangle but before \langle first token \rangle.

We then hit the \langle first token \rangle with \token_to_str:N and grab characters until finding \l_p_t_t_e_k_t_c_k_e_n. More precisely, by looking at the first character in the string representation of the \langle first token \rangle we distinguish three cases: a stringified control sequence starts with the escape character; for an explicit character we find that same character; for an explicit character we find anything else (we made sure to exclude the case of an active character whose string representation coincides with the other two cases).

When \#1 is a stringified active character we pass appropriate arguments to the user’s code; thankfully \char_generate:nn can make active characters.
When \#1 matches the character we had extracted from the meaning of \l_peek_token, the token was an explicit character, which can be a standard space, or a begin-group or end-group character with some character code. In the latter two cases we call \char_generate:nn with suitable arguments and put suitable \if_false: \fi: constructions to make the result balanced and such that \alpha:expanding or \x:expanding gives back a single (unbalanced) begin-group or end-group character.

Finally there is the case of a special token whose string representation starts with an escape character, namely the token was a control sequence. In that case we could have grabbed the token directly as an \N-type argument, but of course we couldn’t know that until we had run all the various tests including stringifying the token. We are thus left with the hard work of picking up one by one the characters in the csname (being careful about spaces), until finding a token that matches the \langle next token \rangle picked up earlier (which was not stringified), such that the control sequence that we found so far indeed has the expected meaning \l_peek_token. This comparison with \l_peek_token catches a reasonably common case like \c_group_begin_token \_ in which the trailing \_ has category code other: without comparison of the constructed csname with
\l_peek_token collection would stop at \c, which is wrong.

\cs_new_protected:Npn \__tl_peek_analysis_escape:
\tl_clear:N \l__tl_internal_a_tl
\tex_futurelet:D \l__tl_analysis_token
\__tl_peek_analysis_collect:w
\}
\cs_new_protected:Npn \__tl_peek_analysis_collect:w
{ \__tl_analysis_char_arg:Nw \__tl_peek_analysis_collect:n }
\cs_new_protected:Npn \__tl_peek_analysis_collect:n #1
{ \tl_put_right:Nn \l__tl_internal_a_tl {#1}
\__tl_peek_analysis_collect_loop:
}
\cs_new_protected:Npn \__tl_peek_analysis_collect_loop:
{ \tex_futurelet:D \l__tl_analysis_token
\__tl_peek_analysis_collect_test:
}
\cs_new_protected:Npn \__tl_peek_analysis_collect_test:
{ \if_meaning:w \l__tl_analysis_token \l__tl_analysis_next_token
\exp_after:wN \if_meaning:w \cs:w \l__tl_internal_a_tl \cs_end: \l_peek_token
\__tl_peek_analysis_collect_end:NNN
\fi:
\fi:
\__tl_peek_analysis_collect:w
}

(End definition for \peek_analysis_map_inline:n and others. This function is documented on page 193.)

\__tl_analysis_show_etc_str
When a control sequence (or active character) and its meaning are too long to fit in one line of the terminal, the end is replaced by this token list.
\tl_const:Nx \c__tl_analysis_show_etc_str % (\token_to_str:N \ETC.)

44.11 Messages

\c__tl_analysis_show_etc_str

(End definition for \peek_analysis_map_inline:n and others. This function is documented on page 193.)
(End definition for \c__ti_analysis_show_etc_str.)

\msg_new:nnn { tl } { show-analysis }
{ The-token-list- \tl_if_empty:nF {#1} { #1 ~ }
\tl_if_empty:nTF {#2}
{ is-empty }
{ contains-the-tokens: #2 }
}

\package
Chapter 45

l3regex implementation

45.1 Plan of attack

Most regex engines use backtracking. This allows to provide very powerful features (back-references come to mind first), but it is costly, and raises the problem of catastrophic backtracking. Since \TeX is not first and foremost a programming language, complicated code tends to run slowly, and we must use faster, albeit slightly more restrictive, techniques, coming from automata theory.

Given a regular expression of \( n \) characters, we do the following:

- (Compiling.) Analyse the regex, finding invalid input, and convert it to an internal representation.

- (Building.) Convert the compiled regex to a non-deterministic finite automaton (NFA) with \( O(n) \) states which accepts precisely token lists matching that regex.

- (Matching.) Loop through the query token list one token (one “position”) at a time, exploring in parallel every possible path (“active thread”) through the NFA, considering active threads in an order determined by the quantifiers’ greediness.

We use the following vocabulary in the code comments (and in variable names).

- **Group**: index of the capturing group, \(-1\) for non-capturing groups.

- **Position**: each token in the query is labelled by an integer \( \langle \text{position} \rangle \), with \( \text{min_pos} - 1 \leq \langle \text{position} \rangle \leq \text{max_pos} \). The lowest and highest positions \( \text{min_pos} - 1 \) and \( \text{max_pos} \) correspond to imaginary begin and end markers (with non-existent category code and character code). \( \text{max_pos} \) is only set quite late in the processing.

- **Query**: the token list to which we apply the regular expression.

- **State**: each state of the NFA is labelled by an integer \( \langle \text{state} \rangle \) with \( \text{min_state} \leq \langle \text{state} \rangle < \text{max_state} \).

- **Active thread**: state of the NFA that is reached when reading the query token list for the matching. Those threads are ordered according to the greediness of quantifiers.
• Step: used when matching, starts at 0, incremented every time a character is read, and is not reset when searching for repeated matches. The integer \l__regex_step_int is a unique id for all the steps of the matching algorithm.

We use \texttt{l3intarray} to manipulate arrays of integers. We also abuse TeX’s \texttt{toks} registers, by accessing them directly by number rather than tying them to control sequence using the \texttt{nevtoks} allocation functions. Specifically, these arrays and \texttt{toks} are used as follows. When building, \texttt{toks(state)} holds the tests and actions to perform in the \langle state \rangle of the NFA. When matching,

• \texttt{\g__regex_state_active_intarray} holds the last \langle step \rangle in which each \langle state \rangle was active.

• \texttt{\g__regex_thread_info_intarray} consists of blocks for each \langle thread \rangle (with min_thread \leq \langle thread \rangle < max_thread). Each block has 1+2\l__regex_capturing_group_int entries: the \langle state \rangle in which the \langle thread \rangle currently is, followed by the beginnings of all submatches, and then the ends of all submatches. The \langle threads \rangle are ordered starting from the best to the least preferred.

• \texttt{\g__regex_submatch_prev_intarray}, \texttt{\g__regex_submatch_begin_intarray} and \texttt{\g__regex_submatch_end_intarray} hold, for each submatch (as would be extracted by \texttt{regex_extract_all:nnN}), the place where the submatch started to be looked for and its two end-points. For historical reasons, the minimum index is twice \texttt{max_state}, and the used registers go up to \l__regex_submatch_int. They are organized in blocks of \l__regex_capturing_group_int entries, each block corresponding to one match with all its submatches stored in consecutive entries.

When actually building the result,

• \texttt{toks\langle position \rangle} holds \langle tokens \rangle which o- and x-expand to the \langle position\rangle-th token in the query.

• \texttt{\g__regex_balance_intarray} holds the balance of begin-group and end-group character tokens which appear before that point in the token list.

The code is structured as follows. Variables are introduced in the relevant section. First we present some generic helper functions. Then comes the code for compiling a regular expression, and for showing the result of the compilation. The building phase converts a compiled regex to NFA states, and the automaton is run by the code in the following section. The only remaining brick is parsing the replacement text and performing the replacement. We are then ready for all the user functions. Finally, messages, and a little bit of tracing code.

\subsection{ Helpers }

\texttt{\_\_regex_int_eval:w} Access the primitive: performance is key here, so we do not use the slower route via \texttt{\int_eval:n}.

\begin{verbatim}
\cs_new_eq:NN \_\_regex_int_eval:w \tex_numexpr:D
\end{verbatim}

(End definition for \texttt{\_\_regex_int_eval:w}.)

\texttt{\_\_regex_standard_escapechar:} Make the \texttt{\\escapechar} into the standard backslash.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex_standard_escapechar:
\int_set:Nn \tex_escapechar:D { \texttt{\textbackslash} }
\end{verbatim}

449
\_\_regex_toks_use:w \ 
Unpack a \toks given its number.

\begin{verbatim}
cs_new:Npn \_\_regex_toks_use:w { \tex_the:D \tex_toks:D }
\end{verbatim}

\_\_regex_toks_clear:Nn \_\_regex_toks_set:Nn \_\_regex_toks_set:N
\ 
Empty a \toks or set it to a value, given its number.

\begin{verbatim}
cs_new_protected:Npn \_\_regex_toks_clear:N #1
{ \_\_regex_toks_set:Nn #1 { } }
cs_new_eq:NN \_\_regex_toks_set:Nn \_\_regex_toks_set:Nn \tex_toks:D
cs_new_protected:Npn \_\_regex_toks_set:No #1
{ \tex_toks:D #1 \exp_after:wN }
\end{verbatim}

\_\_regex_toks_memcpy:NNn \ 
Copy \#3 \toks registers from \#2 onwards to \#1 onwards, like C’s memcpy.

\begin{verbatim}
cs_new_protected:Npn \_\_regex_toks_memcpy:NNn #1#2#3
{ prg_replicate:nn {#3}
  { \tex_toks:D #1 = \tex_toks:D #2
    \int_incr:N #1
    \int_incr:N #2 }
  }
\end{verbatim}

\_\_regex_toks_put_left:Nx \_\_regex_toks_put_right:Nx \_\_regex_toks_put_right:Nn
\ 
During the building phase we wish to add x-expanded material to \toks, either to the left or to the right. The expansion is done “by hand” for optimization (these operations are used quite a lot). The \texttt{Nn} version of \_\_regex_toks_put_right:Nx is provided because it is more efficient than \texttt{x-expanding with \exp_not:n.}

\begin{verbatim}
cs_new_protected:Npn \_\_regex_toks_put_left:Nx #1#2
{ \cs_set_nopar:Npx \_\_regex_tmp:w { #2 }
    \tex_toks:D #1 \exp_after:wN \exp_after:wN \exp_after:wN
    { \exp_after:wN \_\_regex_tmp:w \tex_the:D \tex_toks:D #1 }
  }
cs_new_protected:Npn \_\_regex_toks_put_right:Nx #1#2
{ \cs_set_nopar:Npx \_\_regex_tmp:w {#2}
    \tex_toks:D #1 \exp_after:wN
    { \tex_the:D \tex_toks:D \exp_after:wN #1 \_\_regex_tmp:w }
  }
cs_new_protected:Npn \_\_regex_toks_put_right:Nn \_\_regex_toks_put_right:N
{ \tex_toks:D #1 \exp_after:wN { \tex_the:D \tex_toks:D #1 \_\_regex_tmp:w } }
\end{verbatim}

\_\_regex_curr_cs_to_str: \ 
Expands to the string representation of the token (known to be a control sequence) at the current position \_\_regex_curr_pos_int. It should only be used in \texttt{x}-expansion to avoid losing a leading space.

\begin{verbatim}
cs_new:Npn \_\_regex_curr_cs_to_str:
\end{verbatim}

450
\__regex_intarray_item:NnF \__regex_intarray_item_aux:nNF
Item of intarray, with a default value.
\cs_new:Npn \__regex_intarray_item:NnF \#1\#2
{ \exp_args:Nf \__regex_intarray_item_aux:nNF \{ \int_eval:n \{\#2\} \} \#1 }
\cs_new:Npn \__regex_intarray_item_aux:nNF \#1\#2
{ \if_int_compare:w \#1 > \c_zero_int

\exp_after:wN \use_i:nn
\else:
\exp_after:wN \use_ii:nn
\fi:
{ \__kernel_intarray_item:Nn \#2 \{\#1\} }
}(End definition for \__regex_intarray_item:NnF and \__regex_intarray_item_aux:nNF.)
\__regex_maplike_break:
Analogous to \tl_map_break:, this correctly exits \tl_map_inline:nn and similar constructions and jumps to the matching \prg_break_point:Nn \__regex_maplike_break: { }.
\cs_new:Npn \__regex_maplike_break:
{ \prg_map_break:Nn \__regex_maplike_break: { } }
(End definition for \__regex_maplike_break:)
This regular expression matches nothing, but is still a valid regular expression. We could use a failing assertion, but I went for an empty class. It is used as the initial value for regular expressions declared using \regex_new:N.

\tl_const:Nn \c__regex_no_match_regex
{ \__regex_branch:n
{ \__regex_class:NnnnN \c_true_bool { } { 1 } { 0 } \c_true_bool }
}

(End definition for \c__regex_no_match_regex.)

\l__regex_balance_int
During this phase, \l__regex_balance_int counts the balance of begin-group and end-group character tokens which appear before a given point in the token list. This variable is also used to keep track of the balance in the replacement text.

\int_new:N \l__regex_balance_int
(End definition for \l__regex_balance_int.)

### 45.2.2 Testing characters

\c__regex_ascii_min_int
\c__regex_ascii_max_control_int
\c__regex_ascii_max_int
\int_const:Nn \c__regex_ascii_min_int { 0 }
\int_const:Nn \c__regex_ascii_max_control_int { 31 }
\int_const:Nn \c__regex_ascii_max_int { 127 }

(End definition for \c__regex_ascii_min_int, \c__regex_ascii_max_control_int, and \c__regex_ascii_max_int.)

\c__regex_ascii_lower_int
\int_const:Nn \c__regex_ascii_lower_int { 'a - 'A }

(End definition for \c__regex_ascii_lower_int.)

### 45.2.3 Internal auxiliaries

\q__regex_recursion_stop
Internal recursion quarks.

\quark_new:N \q__regex_recursion_stop
(End definition for \q__regex_recursion_stop.)

\regex_use_none_delimit_by_q_recursion_stop:w
\regex_use_i_delimit_by_q_recursion_stop:nw
\cs_new:Npn \__regex_use_none_delimit_by_q_recursion_stop:w
#1 \q__regex_recursion_stop {#1}
\cs_new:Npn \__regex_use_i_delimit_by_q_recursion_stop:nw
#1 #2 \q__regex_recursion_stop (#1)

(End definition for \__regex_use_none_delimit_by_q_recursion_stop:w and \__regex_use_i_delimit_by_q_recursion_stop:nw.)

\q__regex_nil
Internal quarks.

\quark_new:N \q__regex_nil
(End definition for \q__regex_nil.)

452
Branching quark conditional.

When testing whether a character of the query token list matches a given character class in the regular expression, we often have to test it against several ranges of characters, checking if any one of those matches. This is done with a structure like

\[
\langle \text{test1} \rangle \ldots \langle \text{test}_n \rangle \\
\__regex_break_point:TF \{ \langle \text{true code} \rangle \} \{ \langle \text{false code} \rangle \}
\]

If any of the tests succeeds, it calls \__regex_break_true:w, which cleans up and leaves \langle \text{true code} \rangle in the input stream. Otherwise, \__regex_break_point:TF leaves the \langle \text{false code} \rangle in the input stream.

This function makes showing regular expressions easier, and lets us define \D in terms of \d for instance. There is a subtlety: the end of the query is marked by \-, and thus matches \D and other negated properties; this case is caught by another part of the code.

Simple comparisons triggering \__regex_break_true:w when true.

(End definition for \__regex_item_reverse:n and \__regex_item_caseful_range:nn.)
For caseless matching, we perform the test both on the `curr_char` and on the `case_changed_char`. Before doing the second set of tests, we make sure that `case_changed_char` has been computed.

\begin{verbatim}
\cs_new_protected:Npn \__regex_item_caseless_equal:n #1 \#1
\begin{verbatim}
{ \if_int_compare:w #1 = \l__regex_curr_char_int \exp_after:wN \__regex_break_true:w \fi:
\__regex_maybe_compute_ccc:
\if_int_compare:w #1 = \l__regex_case_changed_char_int \exp_after:wN \__regex_break_true:w \fi:
}\end{verbatim}
\end{verbatim}
\begin{verbatim}
\cs_new_protected:Npn \__regex_item_caseless_range:nn #1 #2 \#1 \#2
\begin{verbatim}
\reverse_if:N \if_int_compare:w #1 > \l__regex_curr_char_int \reverse_if:N \if_int_compare:w #2 < \l__regex_curr_char_int \exp_after:wN \exp_after:wN \exp_after:wN \__regex_break_true:w \fi:
\fi:
\__regex_maybe_compute_ccc:
\reverse_if:N \if_int_compare:w #1 > \l__regex_case_changed_char_int \reverse_if:N \if_int_compare:w #2 < \l__regex_case_changed_char_int \exp_after:wN \exp_after:wN \exp_after:wN \__regex_break_true:w \fi:
\fi:
\end{verbatim}
\end{verbatim}
(End definition for \__regex_item_caseless_equal:n and \__regex_item_caseless_range:nn.)
\begin{verbatim}
\cs_new_protected:Npn \__regex_compute_case_changed_char:
\begin{verbatim}
\int_set_eq:NN \l__regex_case_changed_char_int \l__regex_curr_char_int \if_int_compare:w \l__regex_curr_char_int > 'Z \exp_stop_f: \if_int_compare:w \l__regex_curr_char_int > 'z \exp_stop_f: \else: \if_int_compare:w \l__regex_curr_char_int < 'a \exp_stop_f: \else: \int_sub:Nn \l__regex_case_changed_char_int \c__regex_ascii_lower_int \fi: \fi: \else: \if_int_compare:w \l__regex_curr_char_int < 'A \exp_stop_f: \else: \int_add:Nn \l__regex_case_changed_char_int \c__regex_ascii_lower_int \fi: \fi: \cs_set_eq:NN \__regex_maybe_compute_ccc: \prg_do_nothing: \end{verbatim}
\end{verbatim}
(End definition for \__regex_compute_case_changed_char.)
Those must always be defined to expand to a \texttt{caseful} (default) or \texttt{caseless} version, and not be protected: they must expand when compiling, to hard-code which tests are caseless or caseful.

\cs_new_eq:NN \__regex_item_equal:n ?
\cs_new_eq:NN \__regex_item_range:nn ?

(End definition for \__regex_item_equal:n and \__regex_item_range:nn.)

\__regex_item_catcode:nT
\__regex_item_catcode_reverse:nT
\__regex_item_catcode:

The argument is a sum of powers of 4 with exponents given by the allowed category codes (between 0 and 13). Dividing by a given power of 4 gives an odd result if and only if that category code is allowed. If the catcode does not match, then skip the character code tests which follow.

\cs_new_protected:Npn \__regex_item_catcode:nT #1
\cs_new_protected:Npn \__regex_item_catcode_reverse:nT #1#2
{ \__regex_item_catcode:nT {#1} { \__regex_item_reverse:n {#2} } }

(End definition for \__regex_item_catcode:nT, \__regex_item_catcode_reverse:nT, and \__regex_item_catcode:.)

\__regex_item_exact:nn
\__regex_item_exact_cs:n

This matches an exact \texttt{(category)-(character code)} pair, or an exact control sequence, more precisely one of several possible control sequences, separated by \texttt{\scan_stop:}.

\cs_new_protected:Npn \__regex_item_exact:nn #1#2
{ \if_int_compare:w #1 = \l__regex_curr_catcode_int
\if_int_compare:w #2 = \l__regex_curr_char_int
\exp_after:wN \exp_after:wN \exp_after:wN \__regex_break_true:w
\fi:
\fi:
}

\cs_new_protected:Npn \__regex_item_exact_cs:n #1
{ \int_compare:nNnTF \l__regex_curr_catcode_int = 0
\__kernel_tl_set:Nx \l__regex_internal_a_tl
\tl_if_in:noTF \scan_stop: \__regex_curr_cs_to_str: \scan_stop:
\tl_if_in:noTF \scan_stop: \scan_stop: }

\__regex_item_exact:nn \__regex_item_exact_cs:n #1
{ \int_compare:nNnTF \l__regex_curr_catcode_int = 0
\__kernel_tl_set:Nx \l__regex_internal_a_tl

\__regex_item_cs:n  \hspace{1em}

Match a control sequence (the argument is a compiled regex). First test the catcode of
the current token to be zero. Then perform the matching test, and break if the csname
indeed matches.

\cs_new_protected:Npn \__regex_item_cs:n #1
\{
\int_compare:nNnT \l__regex_curr_catcode_int = 0
\group_begin:
\__regex_single_match:
\__regex_disable_submatches:
\__regex_build_for_cs:n {#1}
\bool_set_eq:NN \l__regex_saved_success_bool
\exp_args:Nx \__regex_match_cs:n { \__regex_curr_cs_to_str: }
\if_meaning:w \c_true_bool \g__regex_success_bool
\group_insert_after:N \__regex_break_true:w
\fi:
\bool_gset_eq:NN \g__regex_success_bool
\l__regex_saved_success_bool
\group_end:
\}

(End definition for \__regex_item_cs:n.)

\__regex_prop_d: \hspace{1em}
\__regex_prop_h: \hspace{1em}
\__regex_prop_s: \hspace{1em}
\__regex_prop_v: \hspace{1em}
\__regex_prop_w: \hspace{1em}
\__regex_prop_N: \hspace{1em}

\cs_new_protected:Npm \__regex_prop_d:
\{ \__regex_item_caseful_range:nn \{ '0 \} \{ '9 \} \}
\cs_new_protected:Npm \__regex_prop_h:
\{ \__regex_item_caseful_equal:n \{ '\} \}
\cs_new_protected:Npm \__regex_prop_s:
\{ \__regex_item_caseful_equal:n \{ '\^I \} \}
\cs_new_protected:Npm \__regex_prop_v:
\{ \__regex_item_caseful_equal:n \{ '\^M \} \}
\cs_new_protected:Npm \__regex_prop_w:
\{ \__regex_item_caseful_equal:n \{ '\^L \} \}
\cs_new_protected:Npm \__regex_prop_N:
\{ \__regex_item_caseful_equal:n \{ '\^J \} \}

(End definition for \__regex_prop_d.)

45.2.4 Character property tests

Character property tests for \d, \W, etc. These character properties are not affected
by the (?! option. The characters recognized by each one are as follows:
\d=[0-9],
\w=[0-9A-Z_a-z], \s=[\h\^I-\^M], \h=[\^I-\^M], \v=[\^J-\^M], and the
upper case counterparts match anything that the lower case does not match. The order
in which the various tests appear is optimized for usual mostly lower case letter text.
\cs_new_protected:Npn \__regex_prop_v:
  { \__regex_item_caseful_range:nn { '\^\^J} { '\^\^M} } % lf, vtab, ff, cr
\cs_new_protected:Npn \__regex_prop_w:
  {
    \__regex_item_caseful_range:nn { 'a} { 'z}
    \__regex_item_caseful_range:nn { 'A} { 'Z}
    \__regex_item_caseful_range:nn { '0} { '9}
    \__regex_item_caseful_equal:n { '_'}
  }
\cs_new_protected:Npn \__regex_prop_N:
  {
    \__regex_item_reverse:n
    { \__regex_item_caseful_equal:n { '\^\^J}}
  }
(End definition for \__regex_prop_d: and others.)

\__regex_posix_alnum: POSIX properties.  No surprise.
\__regex_posix_alpha:
\__regex_posix_ascii:
\__regex_posix_blank:
\__regex_posix_cntrl:
\__regex_posix_digit:
\__regex_posix_lower:
\__regex_posix_print:
\__regex_posix_punct:
\__regex_posix_space:
\__regex_posix_word:
\__regex_posix_xdigit:
45.2.5 Simple character escape

Before actually parsing the regular expression or the replacement text, we go through them once, converting `\n` to the character `10`, `etc. In this pass, we also convert any special character (`*`, `?`, `{`, `etc.`) or escaped alphanumerics into a marker indicating that this was a special sequence, and replace escaped special characters and non-escaped alphanumerics by markers indicating that those were “raw” characters. The rest of the code can then avoid caring about escaping issues (those can become quite complex to handle in combination with ranges in character classes).

Usage: `\__regex_escape_use:nnnn ⟨inline 1⟩⟨inline 2⟩⟨inline 3⟩{⟨token list⟩}`
The ⟨token list⟩ is converted to a string, then read from left to right, interpreting backslashes as escaping the next character. Unescaped characters are fed to the function ⟨inline 1⟩, and escaped characters are fed to the function ⟨inline 2⟩ within an x-expansion context (typically those functions perform some tests on their argument to decide how to output them). The escape sequences `\a`, `\e`, `\f`, `\n`, `\r`, `\t` and `\x` are recognized, and those are replaced by the corresponding character, then fed to ⟨inline 3⟩. The result is then left in the input stream. Spaces are ignored unless escaped.

The conversion is done within an x-expanding assignment.

\__regex_escape_use:nnnn

The result is built in \l__regex_internal_a_tl, which is then left in the input stream. Tracing code is added as appropriate inside this token list. Go through #4 once, applying #1, #2, or #3 as relevant to each character (after de-escaping it).
\__regex_escape_loop:N\__regex_escape_\:w
\__regex_escape_loop:N reads one character: if it is special (space, backslash, or end-marker), perform the associated action, otherwise it is simply an unescaped character. After a backslash, the same is done, but unknown characters are “escaped”.

\cs_new:Npn \__regex_escape_loop:N #1
\{\cs_if_exist_use:cF \{ \__regex_escape_\token_to_str:N #1:w \}
\{ \__regex_escape_unescaped:N #1 \}
\} \__regex_escape_loop:N
\cs_new:cpn \{ \__regex_escape_ \c_backslash_str :w \} \__regex_escape_loop:N
\{ \cs_if_exist_use:cF \{ \__regex_escape_/\token_to_str:N #1:w \}
\{ \__regex_escape_escaped:N #1 \}
\} \__regex_escape_loop:N
\cs_new_eq:NN \__regex_escape_unescaped:N \c_question_mark
\cs_new_eq:NN \__regex_escape_escaped:N \c_double_question_mark
\cs_new_eq:NN \__regex_escape_raw:N \c_double_question_mark
Those functions are never called before being given a new meaning, so their definitions here don’t matter.

\cs_new_eq:cN \{ \__regex_escape_ \iow_char:N\scan_stop: :w \} \prg_break:
\cs_new:cpn \{ \__regex_escape_/ \iow_char:N\scan_stop: :w \}
\{ \msg_expandable_error:nn { regex } { trailing-backslash } \prg_break:\}
\cs_new:cpx \{ \__regex_escape_/a:w \}
\{ \exp_not:N \__regex_escape_raw:N \iow_char:N \^^G \}
\cs_new:cpx \{ \__regex_escape_/t:w \}
\{ \exp_not:N \__regex_escape_raw:N \iow_char:N \^^I \}
\cs_new:cpx \{ \__regex_escape_/r:w \}
\{ \exp_not:N \__regex_escape_raw:N \iow_char:N \^^M \}
\cs_new:cpx \{ \__regex_escape_/e:w \}
\{ \exp_not:N \__regex_escape_raw:N \iow_char:N \^^[ \}
The loop is ended upon seeing the end-marker “break”, with an error if the string ended
in a backslash. Spaces are ignored, and \a, \e, \f, \n, \r, \t take their meaning here.
When \texttt{x} is encountered, \texttt{\__regex_escape_x_test:N} is responsible for grabbing some hexadecimal digits, and feeding the result to \texttt{\__regex_escape_x_end:w}. If the number is too big interrupt the assignment and produce an error, otherwise call \texttt{\__regex_escape_raw:N} on the corresponding character token.

\begin{verbatim}
\cs_new:cpn { \__regex_escape_/x:w } \__regex_escape_loop:N
{ \exp_after:wN \__regex_escape_x_end:w \\
  \int_value:w "0 \__regex_escape_x_test:N \\
}; \cs_new:Npn \__regex_escape_x_end:w #1 ;
{ \int_compare:nNnTF {#1} > \c_max_char_int \\
  \msg_error:nnn { \regex } { x-overflow } {#1} \int_to_Hex:n {#1} \\
} \exp_last_unbraced:Nf \__regex_escape_raw:N \\
{ \char_generate:nn {#1} { 12 } }
}
\end{verbatim}

(End definition for \texttt{\__regex_escape_/x:w}, \texttt{\__regex_escape_x_end:w}, and \texttt{\__regex_escape_x_large:n}.)

\texttt{\__regex_escape_x_test:N}
\texttt{\__regex_escape_x_testii:N}

Find out whether the first character is a left brace (allowing any number of hexadecimal digits), or not (allowing up to two hexadecimal digits). We need to check for the end-of-string marker. Eventually, call either \texttt{\__regex_escape_x_loop:N} or \texttt{\__regex_escape_x:N}.

\begin{verbatim}
\cs_new:Npn \__regex_escape_x_test:N #1
{ \if_meaning:w \scan_stop: #1 \exp_after:wN \use_i:nnn \\
  \fi: \use:n { \if_charcode:w \c_space_token #1 \\
  \exp_after:wN \__regex_escape_x_test:N \\
  \else: \exp_after:wN \__regex_escape_x_testii:N \\
  \exp_after:wN #1 \fi: }
}
\cs_new:Npn \__regex_escape_x_testii:N #1
{ \if_charcode:w \c_left_brace_str #1 \\
  \exp_after:wN \__regex_escape_x_loop:N \\
  \else: \__regex_hexadecimal_use:NTF #1 \\
  \exp_after:wN \__regex_escape_x:N \\
  \exp_after:wN #1 \fi: }
\end{verbatim}

460
\__regex_escape_x:N
This looks for the second digit in the unbraced case.
\__regex_escape_x_loop:N
Grab hexadecimal digits, skip spaces, and at the end, check that there is a right brace, otherwise raise an error outside the assignment.
\__regex_hexadecimal_use:NTF
\TeX\ detects uppercase hexadecimal digits for us but not the lowercase letters, which we need to detect and replace by their uppercase counterpart.
These two tests are used in the first pass when parsing a regular expression. That pass is responsible for finding escaped and non-escaped characters, and recognizing which ones have special meanings and which should be interpreted as “raw” characters. Namely,

- alphanumerics are “raw” if they are not escaped, and may have a special meaning when escaped;
- non-alphanumeric printable ascii characters are “raw” if they are escaped, and may have a special meaning when not escaped;
- characters other than printable ascii are always “raw”.

The code is ugly, and highly based on magic numbers and the ascii codes of characters. This is mostly unavoidable for performance reasons. Maybe the tests can be optimized a little bit more. Here, “alphanumeric” means 0–9, A–Z, a–z; “special” character means non-alphanumeric but printable ascii, from space (hex 20) to del (hex 7E).
45.3 Compiling

A regular expression starts its life as a string of characters. In this section, we convert it to internal instructions, resulting in a “compiled” regular expression. This compiled expression is then turned into states of an automaton in the building phase. Compiled regular expressions consist of the following:

- \_\_regex\_class:NnnnN (boolean) \{\{tests\}\} \{\{min\}\} \{\{more\}\} \{lazyness\}

- \_\_regex\_group:nnnN \{\{branches\}\} \{\{min\}\} \{\{more\}\} \{lazyness\}, also \_\_regex\_group\_no\_capture:nnnN and \_\_regex\_group\_resetting:nnnN with the same syntax.

- \_\_regex\_branch:n \{\{contents\}\}

- \_\_regex\_command_K:

- \_\_regex\_assertion:Nn (boolean) \{\{assertion test\}\}, where the (assertion test) is \_\_regex\_b\_test: or \_\_regex\_Z\_test: or \_\_regex\_A\_test: or \_\_regex\_-G\_test:

Tests can be the following:

- \_\_regex\_item\_caseful\_equal:n \{\{char code\}\}

- \_\_regex\_item\_caseless\_equal:n \{\{char code\}\}

- \_\_regex\_item\_caseful\_range:nn \{\{min\}\} \{\{max\}\}
45.3.1 Variables used when compiling

We make sure to open the same number of groups as we close.

\begin{quote}
\verbatim
\int_new:N \l__regex_group_level_int
(End definition for \l__regex_group_level_int.)
\end{quote}

While compiling, ten modes are recognized, labelled $-63$, $-23$, $-6$, $-2$, $0$, $2$, $3$, $6$, $23$, $63$. See section 45.3.3. We only define some of these as constants.

\begin{quote}
\verbatim
\int_new:N \l__regex_mode_int
\int_const:Nn \c__regex_cs_in_class_mode_int { -6 }
\int_const:Nn \c__regex_cs_mode_int { -2 }
\int_const:Nn \c__regex_outer_mode_int { 0 }
\int_const:Nn \c__regex_catcode_mode_int { 2 }
\int_const:Nn \c__regex_class_mode_int { 3 }
\int_const:Nn \c__regex_catcode_in_class_mode_int { 6 }
(End definition for \l__regex_mode_int and others.)
\end{quote}

We wish to allow constructions such as \verb|\[^BE| ( .. \verb|\cL[a-z]..|), where the outer catcode test applies to the whole group, but is superseded by the inner catcode test. For this to work, we need to keep track of lists of allowed category codes: \l__regex_catcodes_int and \l__regex_default_catcodes_int are bitmaps, sums of $4^c$, for all allowed catcodes $c$. The latter is local to each capturing group, and we reset \l__regex_catcodes_int to that value after each character or class, changing it only when encountering a \verb|\c| escape. The boolean records whether the list of categories of a catcode test has to be inverted: compare \verb|\[^BE| and \verb|\c[BE].|

\begin{quote}
\verbatim
\int_new:N \l__regex_catcodes_int
\int_new:N \l__regex_default_catcodes_int
\bool_new:N \l__regex_catcodes_bool
(End definition for \l__regex_catcodes_int, \l__regex_default_catcodes_int, and \l__regex_catcodes_bool.)
\end{quote}

Constants: $4^c$ for each category, and the sum of all powers of $4$.

\begin{quote}
\verbatim
\int_const:Nn \c__regex_catcode_C_int { "1 }
\int_const:Nn \c__regex_catcode_B_int { "4 }
\int_const:Nn \c__regex_catcode_E_int { "10 }
\int_const:Nn \c__regex_catcode_M_int { "40 }
\int_const:Nn \c__regex_catcode_T_int { "100 }
\end{quote}
45.3.2 Generic helpers used when compiling

Used to compare pairs of things like \_\_regex_compile_special:N \ ? together. It's often inconvenient to get the catcodes of the character to match so we just compare the character code. Besides, the expanding behaviour of \if:w is very useful as that means we can use \c_left_brace_str and the like.

```
\prg_new_conditional:Npnn \_\_regex_two_if_eq:NNNN #1#2#3#4 { TF } {
    \if_meaning:w #1 #3
    \if:w #2 #4
    \prg_return_true:
    \else:
    \prg_return_false:
    \fi:
    \else:
    \prg_return_false:
    \fi:
}
```

(End definition for \_\_regex_two_if_eq:NNNNTF.)
If followed by some raw digits, collect them one by one in the integer variable \#1, and take the true branch. Otherwise, take the false branch.

```latex
\cs_new_protected:Npn \__regex_get_digits:NTFw #1#2#3#4#5
\__regex_if_raw_digit:NNTF #4 #5
{ #1 = #5 \__regex_get_digits_loop:nw {#2} }
{ #3 #4 #5 }
```

```latex
\cs_new:Npn \__regex_get_digits_loop:nw #1#2#3
\__regex_if_raw_digit:NNTF #2 #3
{ #3 \__regex_get_digits_loop:nw {#1} }
{ \scan_stop: #1 #2 #3 }
```

(End definition for \__regex_get_digits:NTFw and \__regex_get_digits_loop:w.)

\__regex_if_raw_digit:NNTF

Test used when grabbing digits for the \{m,n\} quantifier. It only accepts non-escaped digits.

```latex
\prg_new_conditional:Npnn \__regex_if_raw_digit:NN #1#2 { TF }
\if_meaning:w \__regex_compile_raw:N #1
\if_int_compare:w 1 < 1 #2 \exp_stop_f:
\prg_return_true:
\else:
\prg_return_false:
\fi:
\else:
\prg_return_false:
\fi:
```

(End definition for \__regex_if_raw_digit:NNTF.)

### 45.3.3 Mode

When compiling the NFA corresponding to a given regex string, we can be in ten distinct modes, which we label by some magic numbers:

-6 [\c{ [...]}] control sequence in a class,
-2 \c{ [...]} control sequence,
0 ... outer,
2 \c... catcode test,
6 [\c ...] catcode test in a class,
-63 [\c{[...]}) class inside mode −6,
-23 \c{[...]} class inside mode −2,
3 [...] class inside mode 0,
23 \c{[...] class inside mode 2,
This list is exhaustive, because \c escape sequences cannot be nested, and character classes cannot be nested directly. The choice of numbers is such as to optimize the most useful tests, and make transitions from one mode to another as simple as possible.

- Even modes mean that we are not directly in a character class. In this case, a left bracket appends 3 to the mode. In a character class, a right bracket changes the mode as \[ m \rightarrow (m - 15)/13 \], truncated.

- Grouping, assertion, and anchors are allowed in non-positive even modes (0, −2, −6), and do not change the mode. Otherwise, they trigger an error.

- A left bracket is special in even modes, appending 3 to the mode; in those modes, quantifiers and the dot are recognized, and the right bracket is normal. In odd modes (within classes), the left bracket is normal, but the right bracket ends the class, changing the mode from \( m \rightarrow (m - 15)/13 \), truncated; also, ranges are recognized.

- In non-negative modes, left and right braces are normal. In negative modes, however, left braces trigger a warning; right braces end the control sequence, going from −2 to 0 or −6 to 3, with error recovery for odd modes.

- Properties (such as the \d character class) can appear in any mode.

\__regex_if_in_class:TF\footnote{Test whether we are directly in a character class (at the innermost level of nesting).}  
There, many escape sequences are not recognized, and special characters are normal. Also, for every raw character, we must look ahead for a possible raw dash.

\__regex_if_in_cs:TF\footnote{Right braces are special only directly inside control sequences (at the inner-most level of nesting, not counting groups).}  

Assertions are only allowed in modes 0, −2, and −6, i.e., even, non-positive modes.

\cs_new:Npn \_regex_if_in_class_or_catcode:TF
\begin{verbatim}
{ \if_int_odd:w \l__regex_mode_int \exp_after:wN \use_i:nn
\else:
 \if_int_compare:w \l__regex_mode_int > \c__regex_outer_mode_int
 \exp_after:wN \exp_after:wN \exp_after:wN \use_i:nn
\else:
 \exp_after:wN \exp_after:wN \exp_after:wN \use_ii:nn
\fi:
\fi:}
\end{verbatim}

(End definition for \_regex_if_in_class_or_catcode:TF.)

\_regex_if_within_catcode:TF
This test takes the true branch if we are in a catcode test, either immediately following it (modes 2 and 6) or in a class on which it applies (modes 23 and 63). This is used to tweak how left brackets behave in modes 2 and 6.

\cs_new:Npn \_regex_if_within_catcode:TF
\begin{verbatim}
{ \if_int_compare:w \l__regex_mode_int > \c__regex_outer_mode_int
 \exp_after:wN \use_i:nn
\else:
 \exp_after:wN \use_ii:nn
\fi:}
\end{verbatim}

(End definition for \_regex_if_within_catcode:TF.)

\_regex_chk_c_allowed:T
The \texttt{c} escape sequence is only allowed in modes 0 and 3, i.e., not within any other \texttt{c} escape sequence.

\cs_new_protected:Npn \_regex_chk_c_allowed:T
\begin{verbatim}
{ \if_int_compare:w \l__regex_mode_int = \c__regex_outer_mode_int
 \exp_after:wN \use:n
\else:
 \if_int_compare:w \l__regex_mode_int = \c__regex_class_mode_int
 \exp_after:wN \exp_after:wN \exp_after:wN \use:n
\else:
 \msg_error:nn { regex } { c-bad-mode }
 \exp_after:wN \exp_after:wN \exp_after:wN \use_none:n
\fi:
\fi:}
\end{verbatim}

(End definition for \_regex_chk_c_allowed:T.)

\_regex_mode_quit_c:
This function changes the mode as it is needed just after a catcode test.

\cs_new_protected:Npn \_regex_mode_quit_c:
\begin{verbatim}
{ \if_int_compare:w \l__regex_mode_int = \c__regex_catcode_mode_int
 \int_set_eq:NN \l__regex_mode_int \c__regex_outer_mode_int
\else:
\end{verbatim}
\_\_regex_compile:w \_\_regex_compile_end:

Used when compiling a user regex or a regex for the \c{...} escape sequence within another regex. Start building a token list within a group (with x-expansion at the outset), and set a few variables (group level, catcodes), then start the first branch. At the end, make sure there are no dangling classes nor groups, close the last branch: we are done building \l\_\_regex_internal_regex.

\cs_new_protected:Npn \_\_regex_compile:w
\begin{verbatim}
{ \tl_build_begin:N \l__regex_build_tl
\int_zero:N \l__regex_group_level_int
\int_set_eq:NN \l__regex_default_catcodes_int \c__regex_all_catcodes_int
\int_set_eq:NN \l__regex_catcodes_int \l__regex_default_catcodes_int
\cs_set:Npn \__regex_item_equal:n { \__regex_item_caseful_equal:n }
\cs_set:Npn \__regex_item_range:nn { \__regex_item_caseful_range:nn }
\tl_build_put_right:Nn \l__regex_build_tl { \_\_regex_item_caseful:NN \fi: }
\end{verbatim}

\cs_new_protected:Npn \_\_regex_compile_end:
\begin{verbatim}
{ \_\_regex_if_in_class:TF
\begin{verbatim}
{ \_\_regex_if_in_class:TF
\msg_error:nn { regex } { missing-rbrack }
\use:c { \_\_regexCompile_]: }
\prg_do_nothing: \prg_do_nothing:
\end{verbatim}
\end{verbatim}
\if_int_compare:w \l__regex_group_level_int > \c_zero_int
\msg_error:nmx { regex } { missing-rparen }
{ \int_use:N \l__regex_group_level_int }
\prg_replicate:nn
{ \_\_regex_group_level_int }
\begin{verbatim}
{ \tl_build_put_right:Nn \l__regex_build_tl
\if_false: { \fi: }
\if_false: { \fi: } { 1 } { 0 } \c_true_bool
\end{verbatim}
\end{verbatim}
\tl_build_end:N \l__regex_build_tl
\exp_args:NNNo \_\_regex_build_tl
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{vértex}
\tl_build_put_right:Nn \l__regex_build_tl \{ \if_false: \{ \fi: \} \}
\tl_build_end:N \l__regex_build_tl
\exp_args:NNNX
\group_end:
\tl_set:Nn \l__regex_internal_regex \{ \l__regex_build_tl \}
}

(End definition for \_regex_compile:w and \_regex_compile_end:)

__regex_compile:n

The compilation is done between \_regex_compile:w and \_regex_compile_end:, starting in mode 0. Then \_regex_escape_use:nnnn distinguishes special characters, escaped alphanumerics, and raw characters, interpreting \a, \x and other sequences. The 4 trailing \prg_do_nothing: are needed because some functions defined later look up to 4 tokens ahead. Before ending, make sure that any \c{...} is properly closed. No need to check that brackets are closed properly since \_regex_compile_end: does that. However, catch the case of a trailing \cL construction.
\cs_new_protected:Npn \_regex_compile:n #1
\{ 
\_regex_compile:w
\_regex_standard_escapechar:
\int_set_eq:NN \l__regex_mode_int \c__regex_outer_mode_int
\_regex_escape_use:nnnn
\{
\_regex_char_if_special:NTF ##1 \_regex_compile_special:N \_regex_compile_raw:N ##1
\}
\{
\_regex_char_if_alphanumeric:NTF ##1 \_regex_compile_escaped:N \_regex_compile_raw:N ##1
\}
{ \_regex_compile_raw:N ##1 }
{ #1 }
\prg_do_nothing: \prg_do_nothing:
\prg_do_nothing: \prg_do_nothing:
\int_compare:nNnT \l__regex_mode_int = \c__regex_catcode_mode_int
{ \msg_error:nn { regex } { c-trailing } }
\int_compare:nNnT \l__regex_mode_int < \c__regex_outer_mode_int
{ \msg_error:nn { regex } { c-missing-rbrace }
\_regex_compile_end_cs:
\prg_do_nothing: \prg_do_nothing:
\prg_do_nothing: \prg_do_nothing:
}
\_regex_compile_end:
\}

(End definition for \_regex_compile:n.)

__regex_compile_escaped:N
__regex_compile_special:N

If the special character or escaped alphanumeric has a particular meaning in regexes, the corresponding function is used. Otherwise, it is interpreted as a raw character. We distinguish special characters from escaped alphanumeric characters because they behave differently when appearing as an end-point of a range.

470
\newcommand{\__regex_compile_special}{\cs_new_protected:Npn \__regex_compile_special:N #1}{\cs_if_exist_use:cF { \__regex_compile_#1: }{ \__regex_compile_raw:N #1 } }

\newcommand{\__regex_compile_escaped}{\cs_new_protected:Npn \__regex_compile_escaped:N #1}{\cs_if_exist_use:cF { \__regex_compile_/#1: }{ \__regex_compile_raw:N #1 } }

(End definition for \__regex_compile_escaped and \__regex_compile_special.)

\__regex_compile_one:n This is used after finding one “test”, such as \d, or a raw character. If that followed a catcode test (e.g., \cL), then restore the mode. If we are not in a class, then the test is “standalone”, and we need to add \__regex_class:NnnnN and search for quantifiers. In any case, insert the test, possibly together with a catcode test if appropriate.

\newcommand{\__regex_compile_one:n}{\cs_new_protected:Npn \__regex_compile_one:n #1}{\__regex_mode_quit_c: \__regex_if_in_class:TF {}
\tl_build_put_right:Nn \l__regex_build_tl \__regex_class:NnnnN \c_true_bool { \if_false: } \fi: }
\tl_build_put_right:Nx \l__regex_build_tl {\if_int_compare:w \l__regex_catcodes_int < \c__regex_all_catcodes_int \__regex_item_catcode:nT { \int_use:N \l__regex_catcodes_int }{ \exp_not:N \exp_not:n {#1} }\else: \exp_not:N \exp_not:n {#1} \fi: \int_set_eq:NN \l__regex_catcodes_int \l__regex_default_catcodes_int \__regex_if_in_class:TF { } { \__regex_compile_quantifier:w }
}

(End definition for \__regex_compile_one:n.)

\__regex_compile_abort_tokens:n This function places the collected tokens back in the input stream, each as a raw character. Spaces are not preserved.

\newcommand{\__regex_compile_abort_tokens:x}{\cs_new_protected:Npn \__regex_compile_abort_tokens:n #1}{\use:x {\exp_args:No \tl_map_function:nN { \tl_to_str:n {#1} } \__regex_compile_raw:N }{ \__regex_compile_abort_tokens:x }
}

(End definition for \__regex_compile_abort_tokens:n.)
45.3.5 Quantifiers

\__regex_compile_if_quantifier:TFw
This looks ahead and checks whether there are any quantifier (special character equal to
either of \?\*\{). This is useful for the \u and \ur escape sequences.
\cs_new_protected:Npn \__regex_compile_if_quantifier:TFw #1#2#3#4
{\token_if_eq_meaning:NNTF #3 \__regex_compile_special:N
 { \cs_if_exist:cTF { __regex_compile_quantifier_#4:w } }
 { \use_ii:nn }
 (#1) (#2) #3 #4}
(End definition for \__regex_compile_if_quantifier:TFw.)
\__regex_compile_quantifier:w
This looks ahead and finds any quantifier (special character equal to either of \?\*\{).
\cs_new_protected:Npn \__regex_compile_quantifier:w #1#2
{\token_if_eq_meaning:NNTF #1 \__regex_compile_special:N
 { \cs_if_exist_use:cF { __regex_compile_quantifier_#2:w } }
 { \__regex_compile_quantifier_none: #1 #2 }
}
(End definition for \__regex_compile_quantifier:w.)
\__regex_compile_quantifier_none:
\__regex_compile_quantifier_abort:xNN
Those functions are called whenever there is no quantifier, or a braced construction is
invalid (equivalent to no quantifier, and whatever characters were grabbed are left raw).
\cs_new_protected:Npn \__regex_compile_quantifier_none:
{\tl_build_put_right:Nn \l__regex_build_tl
 { \if_false: { \fi: } { 1 } { 0 } \c_false_bool }
}
\cs_new_protected:Npn \__regex_compile_quantifier_abort:xNN #1#2#3
{ \__regex_compile_quantifier_none:
 \msg_warning:nnnx { regex } { invalid-quantifier } {#1} {#3}
 \__regex_compile_abort_tokens:x {#1} #2 #3
}
(End definition for \__regex_compile_quantifier_none: and \__regex_compile_quantifier_abort:xNN.)
\__regex_compile_quantifier_lazyness:nnnN
Once the “main” quantifier (\?, \*, + or a braced construction) is found, we check whether it
is lazy (followed by a question mark). We then add to the compiled regex a closing brace
(ending \__regex_class:NnnnN and friends), the start-point of the range, its end-point,
and a boolean, true for lazy and false for greedy operators.
\cs_new_protected:Npn \__regex_compile_quantifier_lazyness:nnnN #1#2#3#4
{ \__regex_two_if_eq:NNNNTF #3 #4 \__regex_compile_special:N \?
 { \tl_build_put_right:Nn \l__regex_build_tl
 { \if_false: { \fi: } { #1 } { #2 } \c_true_bool } }
For each “basic” quantifier, ?, *, +, feed the correct arguments to \_\_regex_compile_quantifier_lazyness:nnNN. −1 means that there is no upper bound on the number of repetitions.

Three possible syntaxes: \{⟨int⟩\}, \{⟨int⟩,\}, or \{⟨int⟩,⟨int⟩\}. Any other syntax causes us to abort and put whatever we collected back in the input stream, as raw characters, including the opening brace. Grab a number into \_\_regex_internal_a_int. If the number is followed by a right brace, the range is \([a,a]\). If followed by a comma, grab one more number, and call the _ii or _iii auxiliary. Those auxiliaries check for a closing brace, leading to the range \([a,\infty]\) or \([a,b]\), encoded as \{⟨a⟩\}^{−1} and \{⟨a⟩,⟨b⟩−⟨a⟩\}.

(End definition for \_\_regex_compile_quantifier_lazyness:nnNN.)
5186 } #1 #2
5187 
5188 \cs_new_protected:Npn \_regex_compile_quantifier_braced_auxii:w \ #1#2
5189 \{ \_regex_two_if_eq:NNNNTF \#1 \#2 \_regex_compile_special:N \c_right_brace_str
5190 \{ \exp_args:No \_regex_compile_quantifier_lazyness:nnNN
5191 \{ \int_use:N \l__regex_internal_a_int \} \{-1\} \}
5192 \}
5193 \{ \_regex_compile_quantifier_abort:xNN
5194 \{ \c_left_brace_str \int_use:N \l__regex_internal_a_int \}
5195 \}
5196 \}
5197 \}
5198 \cs_new_protected:Npn \_regex_compile_quantifier_braced_auxiii:w \ #1#2
5199 \{ \_regex_two_if_eq:NNNNTF \#1 \#2 \_regex_compile_special:N \c_right_brace_str
5200 \{ \if_int_compare:w \l__regex_internal_a_int > \l__regex_internal_b_int
5201 \msg_error:nnx \{ regex \} \{ backwards-quantifier \}
5202 \{ \int_use:N \l__regex_internal_a_int \}
5203 \}
5204 \{ \int_use:N \l__regex_internal_b_int \}
5205 \int_zero:N \l__regex_internal_b_int
5206 \else:
5207 \int_sub:Nn \l__regex_internal_b_int \l__regex_internal_a_int
5208 \fi:
5209 \exp_args:Noo \_regex_compile_quantifier_lazyness:mmNN
5210 \{ \int_use:N \l__regex_internal_a_int \}
5211 \{ \int_use:N \l__regex_internal_b_int \}
5212 \}
5213 \{ \_regex_compile_quantifier_abort:xNN
5214 \{ \c_left_brace_str
5215 \int_use:N \l__regex_internal_a_int , \int_use:N \l__regex_internal_b_int
5216 \}
5217 \}
5218 \}
5219 \}
5220 \}
5221 \}
5222 \}
5223 \}
5224 \}
5225 \}
5226 \}
5227 \}
5228 \}

(End definition for \_regex_compile_quantifier_f:w and others.)

45.3.6 Raw characters

Within character classes, and following catcode tests, some escaped alphanumeric sequences such as \b do not have any meaning. They are replaced by a raw character, after spitting out an error.

\cs_new_protected:Npn \_regex_compile_raw_error:N #1
\{ \msg_error:nnx \{ regex \} \{ bad-escape \} \}

474
If we are in a character class and the next character is an unescaped dash, this denotes a range. Otherwise, the current character \texttt{#1} matches itself.

We have just read a raw character followed by a dash; this should be followed by an end-point for the range. Valid end-points are: any raw character; any special character, except a right bracket. In particular, escaped characters are forbidden.
45.3.7 Character properties

\_regex\_prop\_: In a class, the dot has no special meaning. Outside, insert \_regex\_prop\_:, which matches any character or control sequence, and refuses −2 (end-marker).

The constants \_regex\_prop\_d:, etc. hold a list of tests which match the corresponding character class, and jump to the \_regex\_break\_point:TF marker. As for a normal character, we check for quantifiers.
45.3.8 Anchoring and simple assertions

In modes where assertions are forbidden, anchors such as \A produce an error (\A is invalid in classes); otherwise they add an \___regex_assertion:Nn test as appropriate (the only negative assertion is \B). The test functions are defined later. The implementation for $ and $ is only different from \A etc because these are valid in a class.

End definition for \___regex_compile_/d: and others.)
45.3.9 Character classes

Outside a class, right brackets have no meaning. In a class, change the mode \((m \rightarrow (m - 15)/13\), truncated) to reflect the fact that we are leaving the class. Look for quantifiers, unless we are still in a class after leaving one (the case of \([\ldots\backslash cL[\ldots]\ldots]\)).

\cs_new_protected:cpn { __regex_compile_ }:

\{ \__regex_if_in_class:TF
\{
 \if_int_compare:w \l__regex_mode_int > \c__regex_catcode_in_class_mode_int
 \tl_build_put_right:Nn \l__regex_build_tl { \if_false: { \fi: } }
 \fi:
 \tex_advance:D \l__regex_mode_int - 15 \exp_stop_f:
 \tex_divide:D \l__regex_mode_int 13 \exp_stop_f:
 \if_int_odd:w \l__regex_mode_int \else:
 \exp_after:wN \__regex_compile_quantifier:w
 \fi:
\}
\{ \__regex_compile_raw:N ] \}

(End definition for __regex_compile_):

In a class, left brackets might introduce a POSIX character class, or mean nothing. Immediately following \c(category), we must insert the appropriate catcode test, then parse the class; we pre-expand the catcode as an optimization. Otherwise (modes 0, −2 and −6) just parse the class. The mode is updated later.

\cs_new_protected:cpn { __regex_compile_[ }:

\{ \__regex_if_in_class:TF
\{ \__regex_compile_class_posix_test:w \}
\{
 \__regex_if_within_catcode:TF
 \{
 \exp_after:wN \__regex_compile_class_catcode:w
 \int_use:N \l__regex_catcodes_int ;
 \}
 \{ \__regex_compile_class_normal:w \}
\}

(End definition for __regex_compile_[):}

\__regex_compile_class_normal:w

In the “normal” case, we insert \__regex_class:NNNN \langle boolean \rangle in the compiled code. The \langle boolean \rangle is true for positive classes, and false for negative classes, characterized by a leading ^\circ. The auxiliary \__regex_compile_class:TFNN also checks for a leading ) which has a special meaning.

\cs_new_protected:Npn \__regex_compile_class_normal:w
\{ \__regex_compile_class:TFNN
\{ \__regex_class:NNNN \c_true_bool \}
\{ \__regex_class:NNNN \c_false_bool \}
\}
This function is called for a left bracket in modes 2 or 6 (catcode test, and catcode test within a class). In mode 2 the whole construction needs to be put in a class (like single character). Then determine if the class is positive or negative, inserting \_\_regex\_item\_catcode:nT or the reverse variant as appropriate, each with the current catcodes bitmap #1 as an argument, and reset the catcodes.

\cs_new_protected:Npn \_\_regex_compile_class_catcode:w #1;
\__regex_compile_class:TFNN #1;#2

If the first character is ^, then the class is negative (use #2), otherwise it is positive (use #1). If the next character is a right bracket, then it should be changed to a raw one.

\cs_new_protected:Npn \_\_regex_compile_class_posix:\w #1#2#3#4
\_\_regex_compile_class_posix_test:w #1#2
\_\_regex_compile_class_posix_loop:w
\_\_regex_compile_class_posix_end:w

Here we check for a syntax such as [:alpha:]. We also detect [= and [, which have a meaning in POSIX regular expressions, but are not implemented in \l3regex. In case we see [:, grab raw characters until hopefully reaching :]. If that’s missing, or the POSIX class is unknown, abort. If all is right, add the test to the current class, with an extra \_\_regex\_item\_reverse:n for negative classes.
\str_case:nn { #2 }
  |
  : { \__regex_compile_class_posix:NNNNw } = {{
    \msg_warning:nnx { regex }
    { posix-unsupported } { = }
  }
  |
    \msg_warning:nnx { regex }
    { posix-unsupported } { . }
  }
}
\__regex_compile_raw:N [ #1 #2 }
\cs_new_protected:Npn \__regex_compile_class_posix:NNNNw #1#2#3#4#5#6
  \{\__regex_two_if_eq:NNNNTF #5 #6 \__regex_compile_special:N ^
  \{ \bool_set_false:N \l__regex_internal_bool
  \__kernel_tl_set:Nx \l__regex_internal_a_tl { \if_false: } \fi:
  \__regex_compile_class_posix_loop:w
  \}
  \{ \bool_set_true:N \l__regex_internal_bool
  \__kernel_tl_set:Nx \l__regex_internal_a_tl { \if_false: } \fi:
  \__regex_compile_class_posix_loop:w #5 #6
  
  }
\cs_new:Npn \__regex_compile_class_posix_loop:w #1#2
  \{ \token_if_eq_meaning:NNTF \__regex_compile_raw:N #1
  \{ #2 \__regex_compile_class_posix_loop:w } \{ \if_false: \{ \fi: \} \__regex_compile_class_posix_end:w #1 #2 
  
  \cs_new_protected:Npn \__regex_compile_class_posix_end:w #1#2#3#4
  \{ \__regex_two_if_eq:NNNNTF #1 #2 \__regex_compile_special:N : \__regex_compile_one:n
  \{ \bool_if:NF \l__regex_internal_bool \__regex_item_reverse:n \exp_not:c \{ \__regex_posix_ \l__regex_internal_a_tl : \}
  \}
  \{ \msg_warning:nnx { regex } \{ posix-unknown \}
  \l\__regex_internal_a_tl
  \__regex_compile_abort_tokens:x
  \}
  
  \str_case:nn { #2 }
  |
  : { \__regex_compile_class_posix:NNNNw } = {{
    \msg_warning:nnx { regex }
    { posix-unsupported } { = }
  }
  |
    \msg_warning:nnx { regex }
    { posix-unsupported } { . }
  }
}
\__regex_compile_raw:N [ #1 #2 }
\cs_new_protected:Npn \__regex_compile_class_posix:NNNNw #1#2#3#4#5#6
  \{\__regex_two_if_eq:NNNNTF #5 #6 \__regex_compile_special:N ^
  \{ \bool_set_false:N \l__regex_internal_bool
  \__kernel_tl_set:Nx \l__regex_internal_a_tl { \if_false: } \fi:
  \__regex_compile_class_posix_loop:w
  \}
  \{ \bool_set_true:N \l__regex_internal_bool
  \__kernel_tl_set:Nx \l__regex_internal_a_tl { \if_false: } \fi:
  \__regex_compile_class_posix_loop:w #5 #6
  
  }
\cs_new:Npn \__regex_compile_class_posix_loop:w #1#2
  \{ \token_if_eq_meaning:NNTF \__regex_compile_raw:N #1
  \{ #2 \__regex_compile_class_posix_loop:w } \{ \if_false: \{ \fi: \} \__regex_compile_class_posix_end:w #1 #2 
  
  \cs_new_protected:Npn \__regex_compile_class_posix_end:w #1#2#3#4
  \{ \__regex_two_if_eq:NNNNTF #1 #2 \__regex_compile_special:N : \__regex_compile_one:n
  \{ \bool_if:NF \l__regex_internal_bool \__regex_item_reverse:n \exp_not:c \{ \__regex_posix_ \l__regex_internal_a_tl : \}
  \}
  \{ \msg_warning:nnx { regex } \{ posix-unknown \}
  \l\__regex_internal_a_tl
  \__regex_compile_abort_tokens:x
  \}
  
  \str_case:nn { #2 }
  |
  : { \__regex_compile_class_posix:NNNNw } = {{
    \msg_warning:nnx { regex }
    { posix-unsupported } { = }
  }
  |
    \msg_warning:nnx { regex }
    { posix-unsupported } { . }
  }
}
45.3.10 Groups and alternations

The contents of a regex group are turned into compiled code in \l__regex_build-_tl, which ends up with items of the form \l__regex_branch:n \langle\text{concatenation}\rangle. This construction is done using \tl_build... functions within a \TeX group, which automatically makes sure that options (case-sensitivity and default catcode) are reset at the end of the group. The argument \#1 is \l__regex_group:nnnN or a variant thereof. A small subtlety to support \texttt{cL(abc)} as a shorthand for \texttt{(cLaLbLc)}: exit any pending catcode test, save the category code at the start of the group as the default catcode for that group, and make sure that the catcode is restored to the default outside the group.

\begin{verbatim}
\cs_new_protected:Npn \__regex_compile_group_begin:N #1
\{
\tl_build_put_right:Nn \l__regex_build_tl { #1 { \if_false: } \fi: }
\__regex_mode_quit_c:
\group_begin:
\tl_build_begin:N \l__regex_build_tl
\int_set_eq:NN \l__regex_default_catcodes_int \l__regex_catcodes_int
\int_incr:N \l__regex_group_level_int
\tl_build_put_right:Nn \l__regex_build_tl
{ \__regex_branch:n { \if_false: } \fi: }
\}
\cs_new_protected:Npm \__regex_compile_group_end:
\{
\if_int_compare:w \l__regex_group_level_int > \c_zero_int
\tl_build_put_right:Nn \l__regex_build_tl { \if_false: { \fi: } }
\tl_build_end:N \l__regex_build_tl
\exp_args:NNNx
\group_end:
\tl_build_put_right:Nn \l__regex_build_tl { \__regex_build_tl }
\int_set_eq:NN \l__regex_catcodes_int \l__regex_default_catcodes_int
\exp_after:wN \__regex_compile_quantifier:w
\else:
\msg_warning:nn { regex } { extra-rparen }
\exp_after:wN \__regex_compile_raw:N \exp_after:wN ) \fi:
\}
\end{verbatim}

(End definition for \l__regex_compile_class_posix_test:w and others.)
\_\_regex\_compile\_(:)  In a class, parentheses are not special. In a catcode test inside a class, a left parenthesis gives an error, to catch \[a\cL(bcd)e\]. Otherwise check for a ?, denoting special groups, and run the code for the corresponding special group.

\cs\_new\_protected:cpn { \_\_regex\_compile\_(:) }
\{ 
\_\_regex\_if\_in\_class:TF { \_\_regex\_compile\_raw:N ( } 
\{ 
\if_int_compare:w \l__regex\_mode\_int = \c__regex\_catcode\_in\_class\_mode\_int 
\msg\_error:nn { regex } { c-lparen\_in\_class } 
\exp\_after:wN \_\_regex\_compile\_raw:N \exp\_after:wN ( 
\else: 
\exp\_after:wN \_\_regex\_compile\_lparen:w 
\fi: 
\}
\}
\cs\_new\_protected:Npn \_\_regex\_compile\_lparen:w #1#2#3#4 
\{ 
\_\_regex\_two\_if\_eq:NNNNTF #1 #2 \_\_regex\_compile\_special:N ? 
\{ 
\cs\_if\_exist\_use:cF 
\{ \_\_regex\_compile\_special\_group:token\_to\_str:N #4 :w \} 
\{ 
\msg\_warning:nnx { regex } { special\_group\_unknown } 
\{ (? #4 \}
\_\_regex\_compile\_group\_begin:N \_\_regex\_group:nnnN 
\_\_regex\_compile\_raw:N ? #3 #4 
\}
\}
\}
\_\_regex\_compile\_group\_begin:N \_\_regex\_group:nnnN #1 #2 #3 #4 
\}
\}
(End definition for \_\_regex\_compile\_(:))

\_\_regex\_compile\_|:  In a class, the pipe is not special. Otherwise, end the current branch and open another one.

\cs\_new\_protected:cpn { \_\_regex\_compile\_|: }
\{ 
\_\_regex\_if\_in\_class:TF { \_\_regex\_compile\_raw:N | } 
\{ 
\tl\_build\_put\_right:Nn \l__regex\_build\_tl 
\{ \if\_false: { \fi: } \_\_regex\_branch:n \{ \if\_false: \} \fi: \} 
\}
\}
(End definition for \_\_regex\_compile\_:)

\_\_regex\_compile\_):  Within a class, parentheses are not special. Outside, close a group.

\cs\_new\_protected:cpn { \_\_regex\_compile\_: }
\{ 
\_\_regex\_if\_in\_class:TF { \_\_regex\_compile\_raw:N ) } 
\}

482
Non-capturing, and resetting groups are easy to take care of during compilation; for those
groups, the harder parts come when building.

\begin{verbatim}
\cs_new_protected:cpn { __regex_compile_special_group_::w }
\{ \__regex_compile_group_begin:N \__regex_group_no_capture:nnnN \}
\cs_new_protected:cpn { __regex_compile_special_group_|:w }
\{ \__regex_compile_group_begin:N \__regex_group_resetting:nnnN \}
\end{verbatim}

The match can be made case-insensitive by setting the option with \texttt{(?i)}; the original
behaviour is restored by \texttt{(?-i)}. This is the only supported option.

\begin{verbatim}
\cs_new_protected:Npn \__regex_compile_special_group_i:w #1#2
\{ \__regex_two_if_eq:NNNNTF #1 #2 \__regex_compile_special:N ) \}
\cs_set:Npn \__regex_item_equal:n { \__regex_item_caseless_equal:n }
\cs_set:Npn \__regex_item_range:nn { \__regex_item_caseless_range:nn }
\msg_warning:nnx { regex } { unknown-option } { (?i #2 }
\__regex_compile_raw:N ( \__regex_compile_raw:N ? \__regex_compile_raw:N i
\#1 #2 )
\cs_new_protected:cpn { __regex_compile_special_group_-:w } #1#2#3#4
\{ \__regex_two_if_eq:NNNNTF #1 #2 \__regex_compile_special:N i
\{ \__regex_two_if_eq:NNNNTF #3 #4 \__regex_compile_special:N ) \}
\use_ii:nn \}
\cs_set:Npn \__regex_item_equal:n { \__regex_item_caseful_equal:n }
\cs_set:Npn \__regex_item_range:nn { \__regex_item_caseful_range:nn }
\msg_warning:nnx { regex } { unknown-option } { (?-#2#4 }
\__regex_compile_raw:N ( \__regex_compile_raw:N ? \__regex_compile_raw:N -
\#1 #2 #3 \#4 )
\end{verbatim}

(End definition for \texttt{\_regex_compile_special_group_::w} and \texttt{\_regex_compile_special_group_|:w})

(End definition for \texttt{\_regex_compile_special_group_i:w} and \texttt{\_regex_compile_special_group_-:w})
45.3.11 Catcodes and csnames

The \c escape sequence can be followed by a capital letter representing a character category, by a left bracket which starts a list of categories, or by a brace group holding a regular expression for a control sequence name. Otherwise, raise an error.

\cs_new_protected:cpn { __regex_compile_/c: }
\{ \__regex.chk_c_allowed:T { \__regex_compile_c_test:NN } \}
\cs_new_protected:Nm \__regex_compile_c_test:NN #1#2
\{
  \token_if_eq_meaning:NNTF #1 \__regex_compile_raw:N
  \{
    \int_if_exist:cTF { c__regex_catcode_#2_int }
    \{
      \int_set_eq:Nc \l__regex_catcodes_int { c__regex_catcode_#2_int }
      \l__regex_mode_int = \if_case:w \l__regex_mode_int
      \c__regex_catcode_mode_int \else:
      \c__regex_catcode_in_class_mode_int
      \fi:
      \token_if_eq_charcode:NNT C #2 { \__regex_compile_c_C:NN }
    }
  \}
  \cs_if_exist_use:cF { __regex_compile_c_#2:w } \}
  \msg_error:nnx { regex } { c-missing-category } {#2} #1 #2
\}

(End definition for \__regex_compile_/c: and \__regex_compile_c_test:NN.)

\__regex_compile_c_C:NN

If \cC is not followed by . or (...) then complain because that construction cannot match anything, except in cases like \cC[^...], where it has no effect.

\cs_new_protected:Nm \__regex_compile_c:C:NN #1#2
\{
  \token_if_eq_meaning:NNTF \__regex_compile_special:N
  \{
    \token_if_eq_charcode:NNTF \__regex_compile_c:C:NN #2 .
    \{ \use_none:n \}
    \\msg_error:nn { regex } { c-C-invalid } {#2} #1 #2
  \}
  \use:n \}
\}

(End definition for \__regex_compile_c:C:NN.)

\__regex_compile_c_[:w
\__regex_compile_c_lbrack_loop:NN
\__regex_compile_c_lbrack_add:N
\__regex_compile_c_lbrack_end:

When encountering \c[, the task is to collect uppercase letters representing character categories. First check for ^ which negates the list of category codes.

\cs_new_protected:cpn { __regex_compile_c_[:w } #1#2
\{

484
\_\_regex_mode_int
    = \if_case:w \_\_regex_mode_int
        \c\_regex_catcode_mode_int
    \else:
        \c\_regex_catcode_in_class_mode_int
    \fi:
\int_zero:N \_\_regex_catcodes_int
\_\_regex_two_if_eq:NNNNTF \_\_compile_special:N ^
{ \bool_set_false:N \_\_regex_catcodes_bool
    \_\_compile_c_lbrack_loop:NN
}:
{ \bool_set_true:N \_\_regex_catcodes_bool
    \_\_compile_c_lbrack_loop:NN
    \_\_compile_c_lbrack_end:
        \_\_compile_c_lbrack_add:N
        \msg_error:nnx { regex } { c-missing-rbrack } {#2}
    \_\_compile_c_lbrack_end:
        \_\_compile_c_lbrack_end:
            \_\_compile_c_lbrack_end:
                \_\_compile_c_lbrack_end:
                    \_\_compile_c_lbrack_add:N
                \exp_stop_f:
            \else:
            \_\_compile_c_lbrack_end:
        \fi:
    \_\_compile_c_lbrack_end:
        \_\_compile_c_lbrack_end:
            \_\_compile_c_lbrack_end:
                \_\_compile_c_lbrack_end:
                    \_\_compile_c_lbrack_end:
                        \_\_compile_c_lbrack_end:
                            \_\_compile_c_lbrack_add:N
                        \exp_stop_f:
                    \else:
                    \_\_compile_c_lbrack_end:
                \fi:
            \_\_compile_c_lbrack_end:
        \else:
        \_\_compile_c_lbrack_end:
    \fi:
\cs_new_protected:Npn \_\_compile_c_lbrack_loop:NN #1 #2
{ \token_if_eq_meaning:NNTF \_\_compile_raw:N
    \_\_compile_c_lbrack_loop:NN
{ \int_if_exist:cTF \c\_regex_catcode_#2_int
    \_\_compile_c_lbrack_add:N
    \_\_compile_c_lbrack_loop:NN
}:
{ \token_if_eq_charcode:NNTF \_\_compile_c_lbrack_end:
    \msg_error:nnx { regex } { c-missing-rbrack } {#2}
    \_\_compile_c_lbrack_end:
        \_\_compile_c_lbrack_end:
            \_\_compile_c_lbrack_end:
                \_\_compile_c_lbrack_end:
                    \_\_compile_c_lbrack_end:
                        \_\_compile_c_lbrack_end:
                            \_\_compile_c_lbrack_add:N
                        \exp_stop_f:
                    \else:
                    \_\_compile_c_lbrack_end:
                \fi:
            \_\_compile_c_lbrack_end:
        \else:
        \_\_compile_c_lbrack_end:
    \fi:
\cs_new_protected:Npn \_\_compile_c_lbrack_add:N #1
{ \if_int_odd:w \int_eval:n { \_\_compile_catcodes_int / #1 } \exp_stop_f:
\else:
    \_\_compile_catcodes_int {#1}
\fi:
\cs_new_protected:Npn \_\_compile_c_lbrack_end:
{ \if_meaning:w \c\_false_bool \_\_compile_catcodes_bool
    \_\_compile_catcodes_int
    \c\_regex_all_catcodes_int - \_\_compile_catcodes_int
    \_\_compile_catcodes_int
    \_\_compile_catcodes_int
}:
(End definition for \_\_compile_c_[:w and others.)
\_\_regex\_compile\_c\_:

The case of a left brace is easy, based on what we have done so far: in a group, compile the
regular expression, after changing the mode to forbid nesting \c. Additionally, disable
submatch tracking since groups don’t escape the scope of \c\{\ldots\}.

\cs_new_protected:cpn { \_\_regex\_compile\_c\_ \c\_left\_brace\_str :w }
{ \_\_regex\_compile:w
    \_\_regex_disable\_submatches:
    \l__regex\_mode\_int
    = \if_case:w \l__regex\_mode\_int
        \c__regex\_cs\_mode\_int
    \else:
        \c__regex\_cs\_in\_class\_mode\_int
    \fi:
}

(End definition for \_\_regex\_compile\_c\_::)

\_\_regex\_compile\_:

We forbid unescaped left braces inside a \c\{\ldots\} escape because they otherwise lead
to the confusing question of whether the first right brace in \c\{x} should end \c or
whether one should match braces.

\cs_new_protected:cpn { \_\_regex\_compile\_ \c\_left\_brace\_str : }
{ \_\_regex\_if\_in\_cs:TF
    { \msg_error:nnn { regex } { cu-lbrace } { c } }
    { \exp_after:wN \_\_regex\_compile\_raw:N \c\_left\_brace\_str }
}

(End definition for \_\_regex\_compile\_::)

\_\_regex\_compile\_end\_cs:
\_\_regex\_compile\_cs\_aux:NNn\_\_\_\_\_\_\_\_\_aux:NNn
\_\_\_\_\_\_\_\_aux:NNn

Non-escaped right braces are only special if they appear when compiling the regular ex-
pression for a csname, but not within a class: \c\{{}) matches the control sequences
\{ and \}. So, end compiling the inner regex (this closes any dangling class or group).
Then insert the corresponding test in the outer regex. As an optimization, if the con-
trol sequence test simply consists of several explicit possibilities (branches) then use
\_\_regex\_item\_exact\_cs:n with an argument consisting of all possibilities separated
by \\scan_stop:\.

\flag_new:n { \_\_regex\_cs }
\cs_new_protected:cpn { \_\_regex\_compile\_ \c\_right\_brace\_str : }
{ \_\_regex\_if\_in\_cs:TF
    { \_\_regex\_compile\_end\_cs: }
    { \exp_after:wN \_\_regex\_compile\_raw:N \c\_right\_brace\_str }
}
\cs_new_protected:Npn \_\_regex\_compile\_end\_cs:
{ \_\_regex\_compile\_end:
    \flag\_clear:n { \_\_regex\_cs }
    \_\_kernel\_tl\_set:Nx \l\_\_regex\_internal\_a\_tl
    { \exp_after:wN \_\_regex\_compile\_cs\_aux:NNn \l\_\_regex\_internal\_regex
        \q__regex\_nil \q__regex\_nil \q__regex\_recursion\_stop
    }
    \exp\_args:Nx \_\_regex\_compile\_one:n
}

486
(End definition for \_regex_compile_): and others.)
45.3.12 Raw token lists with \u

\_regex_compile/_u:
The \u escape is invalid in classes and directly following a catcode test. Otherwise test
for a following r (for \ur), and call an auxiliary responsible for finding the variable name.

\cs_new_protected:cpn { \_regex_compile/_u: } #1#2
\{ \_regex_if_in_class_or_catcode:TF
\{ \_regex_compile_raw_error:N u #1 #2 \}
\{ \_regex_two_if_eq:NNNNTF #1 #2 \_regex_compile_raw:N r
\{ \_regex_compile_u_brace:NNN \_regex_compile_ur_end: \}
\{ \_regex_compile_u_brace:NNN \_regex_compile_u_end: #1 #2 \}
\}
\}

(End definition for \_regex_compile/_u:)

\_regex_compile_u_brace:NNN
This enforces the presence of a left brace, then starts a loop to find the variable name.

\cs_new:Npn \_regex_compile_u_brace:NNN #1#2#3
\{ \_regex_two_if_eq:NNNNTF #2 #3 \_regex_compile_special:N \c_left_brace_str
\{ \tl_set:Nn \l__regex_internal_b_tl {#1}
\_kernel_tl_set:Nx \l__regex_internal_a_tl { \if_false: } \fi:
\_regex_compile_u_loop:NN
\}
\{ \msg_error:nn { regex } { u-missing-lbrace }
\token_if_eq_meaning:NNTF #1 \_regex_compile_ur_end:
\{ \_regex_compile_raw:N u \_regex_compile_raw:N r \}
\{ \_regex_compile_raw:N u \}
\}
\}

(End definition for \_regex_compile_u_brace:NNN)

\_regex_compile_u_loop:NN
We collect the characters for the argument of \u within an x-expanding assignment. In
principle we could just wait to encounter a right brace, but this is unsafe: if the right
brace was missing, then we would reach the end-markers of the regex, and continue,
leading to obscure fatal errors. Instead, we only allow raw and special characters, and
stop when encountering a special right brace, any escaped character, or the end-marker.

\cs_new:Npn \_regex_compile_u_loop:NN #1#2
\{ \token_if_eq_meaning:NNTF #1 \_regex_compile_raw:N
\{ #2 \_regex_compile_u_loop:NN \}
\{ \token_if_eq_meaning:NNTF #1 \_regex_compile_special:N
\{ \exp_after:wN \token_if_eq_charcode:NNTF \c_right_brace_str #2
\{ \if_false: { \fi: } \l__regex_internal_b_tl \}
\{ \if_charcode:w \c_left_brace_str #2
\msg_expandable_error:nnn { regex } { cu-brace } { u }
\}

488
\__regex_compile_u_loop:NN

\__regex_compile_u_end:
\__regex_compile_u_payload:

Once we have extracted the variable's name, we check for quantifiers, in which case we set up a non-capturing group with a single branch. Inside this branch (we omit it and the group if there is no quantifier), \__regex_compile_u_payload: puts the right tests
corresponding to the contents of the variable, which we store in \l__regex_internal_a_tl. The behaviour of \u then depends on whether we are within a \c{...} escape (in this case, the variable is turned to a string), or not.

```
\cs_new_protected:Npn \__regex_compile_u_end:
\{\
  \__regex_compile_if_quantifier:TFw
  \{\
    \tl_build_put_right:Nn \l__regex_build_tl
    \{\
      \__regex_group_no_capture:nnnN { \if_false: } \fi:
      \__regex_branch:n { \if_false: } \fi:
    \}
    \__regex_compile_u_payload:
    \tl_build_put_right:Nn \l__regex_build_tl \{ \if_false: \{ \fi: \}
    \__regex_compile_quantifier:w
  \}
  \}
\}
\cs_new_protected:Npn \__regex_compile_u_payload:
\{\
  \tl_set:Nv \l__regex_internal_a_tl { \l__regex_internal_a_tl }
  \if_int_compare:w \l__regex_mode_int = \c__regex_outer_mode_int
    \__regex_compile_u_not_cs:
  \else:
    \__regex_compile_u_in_cs:
  \fi:
\}
```

(End definition for \__regex_compile_u_end: and \__regex_compile_u_payload.)

\__regex_compile_u_in_cs:

When \u appears within a control sequence, we convert the variable to a string with escaped spaces. Then for each character insert a class matching exactly that character, once.

```
\cs_new_protected:Npn \__regex_compile_u_in_cs:
\{\
  \__kernel_tl_gset:Nx \g__regex_internal_tl
  \{\exp_args:No \__kernel_str_to_other_fast:n
    \{ \l__regex_internal_a_tl \}
  \}
  \tl_build_put_right:Nx \l__regex_build_tl
  \{\tl_map_function:NN \g__regex_internal_tl
    \__regex_compile_u_in_cs_aux:n
  \}
\}
\cs_new:Npn \__regex_compile_u_in_cs_aux:n #1
\{\
  \__regex_class:NnnnN \c_true_bool
  \{ \__regex_item_caseful_equal:n \{ \int_value:w \#1 \}
    \} \{ 0 \} \c_false_bool
\}
```

(End definition for \__regex_compile_u_in_cs.)
In mode 0, the \u escape adds one state to the NFA for each token in \_regex_internal_a_tl. If a given (token) is a control sequence, then insert a string comparison test, otherwise, \_regex_item_exact:nn which compares catcode and character code.

\cs_new_protected:Npn \_regex_compile_u_not_cs: {
\tl_analysis_map_inline:Nn \_regex_internal_a_tl
\{ \tl_build_put_right:Nx \_regex_build_tl
\{ \_regex_class:NnnnN \c_true_bool
\{ \if_int_compare:w "##3 = \c_zero_int
  \_regex_item_exact_cs:n { \exp_after:wN \cs_to_str:N ##1 }
\else:
  \_regex_item_exact:nn { \int_value:w "##3 } { ##2 }
\fi:
\}
\{ 1 \} \{ 0 \} \c_false_bool
\}
\}
(End definition for \_regex_compile_u_not_cs:.)

45313 Other
\_regex_compile_/K:
The \K control sequence is currently the only “command”, which performs some action, rather than matching something. It is allowed in the same contexts as \b. At the compilation stage, we leave it as a single control sequence, defined later.
\cs_new_protected:cpn { \_regex_compile_/K: } {
\int_compare:nNnTF \l__regex_mode_int = \c__regex_outer_mode_int
\{ \tl_build_put_right:Nn \_regex_build_tl \{ \_regex_command_K: \} \}
\{ \_regex_compile_raw_error:N K \}
\}
(End definition for \_regex_compile_/K:.)

45314 Showing regexes
Before showing a regex we check that it is “clean” in the sense that it has the correct internal structure. We do this (in the implementation of \regex_show:N and \regex_log:N) by comparing it with a cleaned-up version of the same regex. Along the way we also need similar functions for other types: all \_regex_clean_(type):n functions produce valid (type) tokens (bool, explicit integer, etc.) from arbitrary input, and the output coincides with the input if that was valid.
\cs_new:Npn \_regex_clean_bool:n #1 {
\tl_if_single:nTF {#1}
\{ \bool_if:NTF #1 \c_true_bool \c_false_bool \}
\}
(End definition for \_regex_clean_bool:n.)

491
\cs_new:Npn \__regex_clean_int:n #1
{
  \tl_if_head_eq_meaning:nNTF {#1} -
  { - \exp_args:No \__regex_clean_int:n { \use_none:n #1 } }
  { \int_eval:n { 0 \str_map_function:nN {#1} \__regex_clean_int_aux:N } }
}
\cs_new:Npn \__regex_clean_int_aux:N #1
{
  \if_int_compare:w 1 < 1 #1 ~
  #1
  \else:
  \exp_after:wN \str_map_break:
  \fi:
}
\cs_new:Npn \__regex_clean_regex:n #1
{
  \__regex_clean_regex_loop:w #1 \__regex_branch:n { \q_recursion_tail } \q_recursion_stop
}
\cs_new:Npn \__regex_clean_regex_loop:w #1 \__regex_branch:n #2
{
  \quark_if_recursion_tail_stop:n {#2}
  \__regex_branch:n { \__regex_clean_branch:n {#2} }
  \__regex_clean_regex_loop:w
}
\cs_new:Npn \__regex_clean_branch:n #1
{
  \__regex_clean_branch_loop:n #1
  \prg_break_point:
}
\cs_new:Npn \__regex_clean_branch_loop:n #1
{
  \tl_if_single:nF {#1} { \prg_break: }
  \token_case_meaning:NnF #1
  \__regex_command_K: { #1 \__regex_clean_branch_loop:n }
  \__regex_assertion:Nn { #1 \__regex_clean_assertion:Nn }
  \__regex_class:NnnnN { #1 \__regex_clean_class:NnnnN }
  \__regex_group:nnnN { #1 \__regex_clean_group:nnnN }
  \__regex_group_no_capture:nnnN { #1 \__regex_clean_group:nnnN }
  \__regex_group_resetting:nnnN { #1 \__regex_clean_group:nnnN }
  \prg_break: }
}
\cs_new:Npn \__regex_clean_assertion:Nn #1#2
{
  \__regex_clean_bool:n {#1}
  \tl_if_single:nF {#2} { { \__regex_A_test: } \prg_break: }
  \token_case_meaning:NnTF #2
  \__regex_clean_bool:n {#1}
  \tl_if_single:nF {#2} { { \__regex_A_test: } \prg_break: }
  \token_case_meaning:NnTF #2
  \__regex_A_test: { }
  \__regex_G_test: { }
  \__regex_Z_test: { }
  \__regex_b_test: { }

492
\cs_new:Npn __regex_clean_branch_loop:n
{\__regex_clean_branch_loop:n}
\cs_new:Npn __regex_clean_class:NnnnN #1#2#3#4#5
{\__regex_clean_branch_loop:n}
\cs_new:Npn __regex_clean_group:nnnN #1#2#3#4
{\__regex_clean_branch_loop:n}
\cs_new:Npn __regex_clean_class:n #1
{\__regex_clean_class_loop:nnn #1 ????? \prg_break_point: }
\cs_new:Npn __regex_clean_class_loop:nnn #1#2#3
{\tl_if_single:nF {#1} { \prg_break: }
\token_case_meaning:NnTF #1
{ \__regex_item_caseful_range:nn { } { } }
{ \__regex_item_caseless_range:nn { } { } }
{ \__regex_item_exact:nn { } { } }
}{ #1 { \__regex_clean_int:n {#2} } { \__regex_clean_int:n {#3} } }
\token_case_meaning:NnTF #1
{ \__regex_item_catcode:nT { } { } }
{ \__regex_item_catcode_reverse:nT { } { } }
{ #1 { \__regex_clean_int:n {#2} } { \__regex_clean_int:n {#3} } }
{ \__regex_clean_class_loop:n}
\__regex_clean_class_loop:nnn
\cs_new:Npn \__regex_clean_exact_cs:n #1
  \exp_last_unbraced:Nf \use_none:n
  \__regex_clean_exact_cs:w #1
  \scan_stop: \q_recursion_tail \scan_stop:
\q_recursion_stop
\cs_new:Npn \__regex_clean_exact_cs:w #1 \scan_stop:
  \quark_if_recursion_tail_stop:n {#1}
  \scan_stop: \tl_to_str:n {#1}
  \__regex_clean_exact_cs:w
\__regex_show:N
\__regex_clean_exact_cs:n

Within a group and within \tl_build_begin:N ... \tl_build_end:N we redefine all
the function that can appear in a compiled regex, then run the regex. The result stored
in \l__regex_internal_a_tl is then meant to be shown.
\cs_new_protected:Npn \__regex_show:N #1
  \group_begin:
  \tl_build_begin:N \l__regex_build_tl
  \cs_set_protected:Npn \__regex_branch:n
  { \seq_pop_right:NN \l__regex_show_prefix_seq
    \l__regex_internal_a_tl
    \__regex_show_one:n { +branch }
    \seq_put_right:No \l__regex_show_prefix_seq
    \l__regex_internal_a_tl
    \use:n
  }
  \cs_set_protected:Npn \__regex_group:nnnN
    { \__regex_show_group_aux:nnnnN { } }
  \cs_set_protected:Npn \__regex_group_no_capture:nnnN
    { \__regex_show_group_aux:nnnnN { ~(no~capture) } }
  \cs_set_protected:Npn \__regex_group_resetting:nnnN
    { \__regex_show_group_aux:nnnnN { ~(resetting) } }
  \cs_set_eq:NN \__regex_class:NnnnN \__regex_show_class:NnnnN
  \cs_set_protected:Npn \__regex_command_K: { \__regex_show_one:n { reset~match~start~(\iow_char:N\K) } }
  \cs_set_protected:Npn \__regex_assertion:Nn ##1##2
    { \__regex_show_one:n { \bool_if:NF ##1 { negative~ } assertion:~##2 } }
  \cs_set:Npn \__regex_b_test: { word~boundary }
  \cs_set:Npn \__regex_Z_test: { anchor~at~end~(\iow_char:N\Z) }
Show a single character, together with its ascii representation if available. This could be extended to beyond ascii. It is not ideal for parentheses themselves.

Every part of the final message go through this function, which adds one line to the output, with the appropriate prefix.
\__regex_show_one:n

Enter and exit levels of nesting. The scope function prints its first argument as an “introduction”, then performs its second argument in a deeper level of nesting.

\cs_new_protected:Npn \__regex_show_push:n \#1
\{ \seq_put_right:Nx \l__regex_show_prefix_seq \{ \#1 ~ \} \}

\cs_new_protected:Npn \__regex_show_pop:
\{
\seq_pop_right:NN \l__regex_show_prefix_seq \l__regex_internal_a_tl \}

\cs_new_protected:Npn \__regex_show_scope:nn \#1\#2
\{
\__regex_show_one:n \{\#1\}
\__regex_show_push:n \{ ~ \}
\use_ii:nn \#2
\__regex_show_pop:
\}

(End definition for \__regex_show_one:n.)

\__regex_show_group_aux:nnnnN

We display all groups in the same way, simply adding a message, (no capture) or (resetting), to special groups. The odd \use_ii:nn avoids printing a spurious +-branch for the first branch.

\cs_new_protected:Npn \__regex_show_group_aux:nnnnN \#1\#2\#3\#4\#5
\{
\__regex_show_one:n \{,-group~begin \#1 \}
\__regex_show_push:n \{ \} \use_ii:nn \#2
\__regex_show_pop:
\__regex_show_one:n
\{
\'~\text{group-end} \__regex_msg_repeated:nnn \{\#3 \} \{\#4 \} \#5 \} \}

(End definition for \__regex_show_group_aux:nnnnN.)

\__regex_show_class:NnnnN

I’m entirely unhappy about this function: I couldn’t find a way to test if a class is a single test. Instead, collect the representation of the tests in the class. If that had more than one line, write Match or Don’t match on its own line, with the repeating information if any. Then the various tests on lines of their own, and finally a line. Otherwise, we need to evaluate the representation of the tests again (since the prefix is incorrect). That’s clunky, but not too expensive, since it’s only one test.

\cs_set:Npn \__regex_show_class:NnnnN \#1\#2\#3\#4\#5
\{
group_begin:
\tl_build_begin:N \l__regex_build_tl
\int_zero:N \l__regex_show_lines_int
\__regex_show_push:n \{\}
#2
\int_compare:nTF { \l__regex_show_lines_int = 0 } { \group_end: \__regex_show_one:n { \bool_if:nTF { \l__regex_show_lines_int = 0 } { Fail } { Pass } } } { \bool_if:nTF { \l__regex_show_lines_int = 1 } { \tl_build_put_right:Nn \l__regex_build_tl \{ \_regex_msg_repeated:nN {#3} {#4} #5 \} } { \tl_build_end:N \l__regex_build_tl \exp_args:NNNo \group_end: \tl_set:Nn \l__regex_internal_a_tl \l__regex_build_tl \__regex_show_one:n { \bool_if:NTF { \l__regex_show_lines_int = 0 } { Match } { Don't~match } \__regex_msg_repeated:nN {#3} {#4} #5 } \tl_build_put_right:Nx \l__regex_build_tl \exp_not:o \l__regex_internal_a_tl } } (End definition for \_regex_show_class:NnnnN.)

\_regex_show_item_catcode:NnT Produce a sequence of categories which the catcode bitmap #2 contains, and show it, indenting the tests on which this catcode constraint applies.
\cs_new_protected:Npn \_regex_show_item_catcode:NnT #1#2 { \seq_set_split:Nnn \l__regex_internal_seq { } { CBEMTPUDSLOA } \seq_set_filter:NNn \l__regex_internal_seq \l__regex_internal_seq \{ \int_if_odd_p:n { #2 / \int_use:c { c__regex_catcode_##1_int } } \} \_regex_show_scope:nn { \_regex_show_item_catcode:NnT \cs_new_protected:Npn \_regex_show_item_exact_cs:n #1#2 } { \bool_if:NTF { \l__regex_show_lines_int = 0 } { negative~ } class } } (End definition for \_regex_show_item_catcode:NnT.)

\_regex_show_item_exact_cs:n \cs_new_protected:Npn \_regex_show_item_exact_cs:n #1 #2 { \seq_set_split:Nnn \l__regex_internal_seq { } { CBEMTPUDSLOA } \seq_set_filter:NNn \l__regex_internal_seq \l__regex_internal_seq \{ \int_if_odd_p:n { #2 / \int_use:c { c__regex_catcode_##1_int } } \} \_regex_show_scope:nn { \_regex_show_item_exact_cs:n \cs_new_protected:Npn \_regex_show_item_exact_cs:n #1#2 } { \bool_if:NTF { \l__regex_show_lines_int = 0 } { negative~ } class } }
45.4 Building

45.4.1 Variables used while building

The last state that was allocated is \l__regex_max_state_int − 1, so that \l__regex_max_state_int always points to a free state. The \texttt{min_state} variable is 1 to begin with, but gets shifted in nested calls to the matching code, namely in \texttt{\c{...}} constructions.

Alternatives are implemented by branching from a \texttt{left} state into the various choices, then merging those into a \texttt{right} state. We store information about those states in two sequences. Those states are also used to implement group quantifiers. Most often, the left and right pointers only differ by 1.

\l__regex_capturing_group_int is the next ID number to be assigned to a capturing group. This starts at 0 for the group enclosing the full regular expression, and groups are counted in the order of their left parenthesis, except when encountering \texttt{resetting} groups.

This phase is about going from a compiled regex to an NFA. Each state of the NFA is stored in a \texttt{\toks}. The operations which can appear in the \texttt{\toks} are:

- \texttt{\__regex_action_start_wildcard:N \langle boolean \rangle} inserted at the start of the regular expression, where a \texttt{true} \langle boolean \rangle makes it unanchored.
- \texttt{\__regex_action_success:} marks the exit state of the NFA.
- \texttt{\__regex_action_cost:n \{<shift>\}} is a transition from the current \langle state \rangle to \langle state \rangle + \langle shift \rangle, which consumes the current character: the target state is saved and will be considered again when matching at the next position.
• \_\_regex_action_free:n \{\langle shift\rangle\}, and \_\_regex_action_free_group:n \{\langle shift\rangle\} are free transitions, which immediately perform the actions for the state \langle state \rangle + \langle shift \rangle of the NFA. They differ in how they detect and avoid infinite loops. For now, we just need to know that the group variant must be used for transitions back to the start of a group.

• \_\_regex_action_submatch:nN \{\langle group \rangle\} \{\langle key \rangle\} where the \langle key \rangle is < or > for the beginning or end of group numbered \langle group \rangle. This causes the current position in the query to be stored as the \langle key \rangle submatch boundary.

• One of these actions, within a conditional.

We strive to preserve the following properties while building.

• The current capturing group is capturing_group − 1, and if a group opened now it would be labelled capturing_group.

• The last allocated state is max_state − 1, so max_state is a free state.

• The left_state points to a state to the left of the current group or of the last class.

• The right_state points to a newly created, empty state, with some transitions leading to it.

• The left/right sequences hold a list of the corresponding end-points of nested groups.

The n-type function first compiles its argument. Reset some variables. Allocate two states, and put a wildcard in state 0 (transitions to state 1 and 0 state). Then build the regex within a (capturing) group numbered 0 (current value of capturing_group). Finally, if the match reaches the last state, it is successful. A false boolean for argument #1 for the auxiliaries will suppress the wildcard and make the match anchored: used for \peek_regex:nTF and similar.
The matching code relies on some global intarray variables, but only uses a range of their entries. Specifically,

- `\g__regex_state_active_intarray from \l__regex_min_state_int to \l__regex_max_state_int - 1`;

Here, in this nested call to the matching code, we need the new versions of this range to involve completely new entries of the intarray variables, so we begin by setting (the new) \l__regex_min_state_int to (the old) \l__regex_max_state_int to use higher entries.

When using a regex to match a cs, we don’t insert a wildcard, we anchor at the end, and since we ignore submatches, there is no need to surround the expression with a group. However, for branches to work properly at the outer level, we need to put the appropriate left and right states in their sequence.

500
Add a transition from #2 to #3 using the function #1. The left function is used for higher priority transitions, and the right function for lower priority transitions (which should be performed later). The signatures differ to reflect the differing usage later on. Both functions could be optimized.

\begin{verbatim}
\cs_new_protected:Npn \__regex_build_transition_left:NNN #1#2#3
{ \__regex_toks_put_left:Nx #2 { #1 { \int_eval:n { #3 - #2 } } } }
\cs_new_protected:Npn \__regex_build_transition_right:nNn #1#2#3
{ \__regex_toks_put_right:Nx #2 { #1 { \int_eval:n { #3 - #2 } } } }
\end{verbatim}

(End definition for \__regex_build_transition_left:NNN and \__regex_build_transition_right:nNn.)

\subsection*{\__regex_build_new_state:}

Add a new empty state to the NFA. Then update the left, right, and max states, so that the right state is the new empty state, and the left state points to the previously “current” state.

\begin{verbatim}
\cs_new_protected:Npn \__regex_build_new_state:
{ \__regex_toks_clear:N \l__regex_max_state_int
 \int_set_eq:NN \l__regex_left_state_int \l__regex_right_state_int
 \int_set_eq:NN \l__regex_right_state_int \l__regex_max_state_int
 \int_incr:N \l__regex_max_state_int
}
\end{verbatim}

(End definition for \__regex_build_new_state:)

\subsection*{\__regex_build_transitions_lazyness:NNNNN}

This function creates a new state, and puts two transitions starting from the old current state. The order of the transitions is controlled by #1, true for lazy quantifiers, and false for greedy quantifiers.

\begin{verbatim}
\cs_new_protected:Npn \__regex_build_transitions_lazyness:NNNNN #1#2#3#4#5
{ \__regex_build_new_state:
 \__regex_toks_put_right:Nx \l__regex_left_state_int
{ \if_meaning:w \c_true_bool #1
 #2 { \int_eval:n { #3 - \l__regex_left_state_int } } 
 #4 { \int_eval:n { #5 - \l__regex_left_state_int } } 
 \else:
 #4 { \int_eval:n { #5 - \l__regex_left_state_int } } 
 #2 { \int_eval:n { #3 - \l__regex_left_state_int } } 
 \fi:
}
\end{verbatim}

(End definition for \__regex_build_transitions_lazyness:NNNNN.)

\section*{45.4.4 Building classes}

The arguments are: \{boolean\} \{\{tests\}\} \{\{min\}\} \{\{more\}\} \{lazyness\}. First store the tests with a trailing \__regex_action_cost:n, in the true branch of \__regex_break_point:TF for positive classes, or the false branch for negative classes. The integer \{more\} is 0 for fixed repetitions, −1 for unbounded repetitions, and \{max\} − \{min\} for a range of repetitions.

\begin{verbatim}
\cs_new_protected:Npn \__regex_class:NnnnN \__regex_tests_action_cost:n #1#2#3#4#5
{ }
\end{verbatim}
\_\_regex_class_repeat:n This is used for a fixed number of repetitions. Build one state for each repetition, with a transition controlled by the tests that we have collected. That works just fine for \#1 = 0 repetitions: nothing is built.

\cs_new_protected:Npm \_\_regex_class_repeat:n #1
{ \prg_replicate:nn {#1}
{ \_\_regex_build_new_state:
 \_\_regex_build_transition_right:nNn \_\_regex_tests_action_cost:n
 \l__regex_left_state_int \l__regex_right_state_int
}
}

\_\_regex_class_repeat:nN This implements unbounded repetitions of a single class (e.g. the * and + quantifiers). If the minimum number \#1 of repetitions is 0, then build a transition from the current state to itself governed by the tests, and a free transition to a new state (hence skipping the tests). Otherwise, call \_\_regex_class_repeat:n for the code to match \#1 repetitions, and add free transitions from the last state to the previous one, and to a new one. In both cases, the order of transitions is controlled by the lazyness boolean \#2.

\cs_new_protected:Npm \_\_regex_class_repeat:nN #1#2
{ \if_int_compare:w #1 = \c_zero_int
 \_\_regex_build_transitions_lazyness:NNNNNN \_\_regex_tests_action_cost:n \l__regex_right_state_int \l__regex_left_state_int
 \else:
 \_\_regex_class_repeat:n \#1
 \int_set_eq:NN \l__regex_internal_a_int \l__regex_left_state_int
 \_\_regex_build_transitions_lazyness:NNNNNN \_\_regex_tests_action_cost:n \l__regex_right_state_int
 \_\_regex_action_free:n \l__regex_right_state_int
 \_\_regex_action_free:n \l__regex_internal_a_int
 \fi:
}

(End definition for \_\_regex_class_repeat:nN.)
We want to build the code to match from \#1 to \#1 + \#2 repetitions. Match \#1 repetitions (can be 0). Compute the final state of the next construction as a. Build \#2 > 0 states, each with a transition to the next state governed by the tests, and a transition to the final state a. The computation of a is safe because states are allocated in order, starting from \texttt{max_state}.

\begin{verbatim}
\cs_new_protected:Npn \__regex_class_repeat:nnN #1#2#3
  { \__regex_class_repeat:n {#1} \int_set:Nn \l__regex_internal_a_int
    { \l__regex_max_state_int + #2 - 1 } \prg_replicate:nn { #2 }
      { \__regex_build_transitions_lazyness:NNNNN #3 \__regex_action_free:n \l__regex_internal_a_int \__regex_tests_action_cost:n \l__regex_right_state_int } }
\end{verbatim}

(End definition for \texttt{\__regex_class_repeat:nnN}.)

45.4.5 Building groups

Arguments: \{(label)\} \{(contents)\} \{(min)\} \{(more)\} \{(lazyness)\}. If \{(min)\} is 0, we need to add a state before building the group, so that the thread which skips the group does not also set the start-point of the submatch. After adding one more state, the \texttt{left_state} is the left end of the group, from which all branches stem, and the \texttt{right_state} is the right end of the group, and all branches end their course in that state. We store those two integers to be queried for each branch, we build the \texttt{NFA} states for the contents \#2 of the group, and we forget about the two integers. Once this is done, perform the repetition: either exactly \#3 times, or \#3 or more times, or between \#3 and \#3 + \#4 times, with lazyness \#5. The \{(label)\} \#1 is used for submatch tracking. Each of the three auxiliaries expects \texttt{left_state} and \texttt{right_state} to be set properly.

\begin{verbatim}
\cs_new_protected:Npn \__regex_group_aux:nnnnN #1#2#3#4#5
  { \if_int_compare:w #3 = \c_zero_int \__regex_build_new_state: \__regex_build_transition_right:nNn \__regex_action_free_group:n \l__regex_internal_a_int \l__regex_right_state_int \fi:
    \__regex_build_new_state: \__regex_push_lr_states: \#2 \__regex_pop_lr_states: \if_case:w - #4 \exp_stop_f: \__regex_group_repeat:nn {#1} {#3} \or: \__regex_group_repeat:nnn {#1} {#3} {#4} \else: \__regex_group_repeat:nnnnN {#1} {#3} {#4} #5 \fi:
  }
\end{verbatim}

(End definition for \texttt{\__regex_group_aux:nnnnN}.)
\_regex_group:nnnN
\_regex_group_no_capture:nnnN
Hand to \_regex_group_aux:nnnnN the label of that group (expanded), and the group itself, with some extra commands to perform.

\cs_new_protected:Npn \_regex_group:nnnN #1
\exp_args:No \_regex_group_aux:nnnnN
{ \int_use:N \l__regex_capturing_group_int }
\int_incr:N \l__regex_capturing_group_int
#1
\)
\cs_new_protected:Npn \_regex_group_no_capture:nnnN
{ \_regex_group_aux:nnnnN { -1 } }
(End definition for \_regex_group:nnnN and \_regex_group_no_capture:nnnN.)

\_regex_group_resetting:nnnN
\_regex_group_resetting_loop:nnNn
Again, hand the label −1 to \_regex_group_aux:nnnnN, but this time we work a little bit harder to keep track of the maximum group label at the end of any branch, and to reset the group number at each branch. This relies on the fact that a compiled regex always is a sequence of items of the form \_regex_branch:n {⟨branch⟩}.

\cs_new_protected:Npn \_regex_group_resetting:nnnN #1
\exp_args:Noo \_regex_group_resetting_loop:nnNn
{ \int_use:N \l__regex_capturing_group_int }
\int_use:N \l__regex_capturing_group_int
#1
{ ?? \prg_break:n } { }
\prg_break_point:
}
\cs_new_protected:Npn \_regex_group_resetting_loop:nnNn #1#2#3#4
{ \use_none:nn #3 { \int_set:Nn \l__regex_capturing_group_int {#1} } \int_set:Nn \l__regex_capturing_group_int {#2} \int_max:nn {#1} { \l__regex_capturing_group_int } }
\exp_args:Nf \_regex_group_resetting_loop:nnNn
(End definition for \_regex_group_resetting:nnnN and \_regex_group_resetting_loop:nnNn.)

\_regex_branch:n
Add a free transition from the left state of the current group to a brand new state, starting point of this branch. Once the branch is built, add a transition from its last state to the right state of the group. The left and right states of the group are extracted from the relevant sequences.

\cs_new_protected:Npn \_regex_branch:n #1
{ \_regex_build_new_state:
\seq_get:NN \l__regex_left_state_seq \l__regex_internal_a_tl
\int_set:Nn \l__regex_left_state_int \l__regex.EventSystems_state
\_regex_build_transition_right:nNn \_regex_action_free:n
}
This function is called to repeat a group a fixed number of times \#2; if this is 0 we remove the group altogether (but don’t reset the capturing_group label). Otherwise, the auxiliary `\__regex_group_repeat_aux:n` copies \#2 times the \toks for the group, and leaves internal_a pointing to the left end of the last repetition. We only record the submatch information at the last repetition. Finally, add a state at the end (the transition to it has been taken care of by the replicating auxiliary).

\begin{verbatim}
\cs_new_protected:Npn \__regex_group_repeat:nn #1#2
{\if_int_compare:w #2 = \c_zero_int
 \int_set:Nn \l__regex_max_state_int { \l__regex_left_state_int - 1 }
 \__regex_build_new_state:
\else:
 \__regex_group_repeat_aux:n {#2}
 \__regex_group_submatches:nNN {#1}
 \l__regex_internal_a_int \l__regex_right_state_int
 \__regex_build_new_state:
\fi:}
\end{verbatim}

(End definition for `\__regex_group_repeat:nn`.)

\__regex_group_submatches:nNN
This inserts in states \#2 and \#3 the code for tracking submatches of the group \#1, unless inhibited by a label of \texttt{-1}.

\begin{verbatim}
\cs_new_protected:Npn \__regex_group_submatches:nNN #1#2#3
{\if_int_compare:w #1 > - \c_one_int
 \__regex_toks_put_left:Nx #2 { \__regex_action_submatch:nN {#1} < }
 \__regex_toks_put_left:Nx #3 { \__regex_action_submatch:nN {#1} > }
\fi:}
\end{verbatim}

(End definition for `\__regex_group_submatches:nNN`.)

\__regex_group_repeat_aux:n
Here we repeat \toks ranging from \texttt{left_state} to \texttt{max_state}, \#1 > 0 times. First add a transition so that the copies “chain” properly. Compute the shift \( c \) between the original copy and the last copy we want. Shift the \texttt{right_state} and \texttt{max_state} to their final values. We then want to perform \( c \) copy operations. At the end, \texttt{b} is equal to the \texttt{max_state}, and \texttt{a} points to the left of the last copy of the group.

\begin{verbatim}
\cs_new_protected:Npn \__regex_group_repeat_aux:n #1
{ \__regex_build_transition_right:nNn \__regex_action_free:n
 \l__regex_right_state_int \l__regex_max_state_int
 \int_set_eq:NN \l__regex_internal_a_int \l__regex_left_state_int
 \int_set_eq:NN \l__regex_internal_b_int \l__regex_max_state_int
\end{verbatim}
This function is called to repeat a group at least \( n \) times; the case \( n = 0 \) is very different from \( n > 0 \). Assume first that \( n = 0 \). Insert submatch tracking information at the start and end of the group, add a free transition from the right end to the “true” left state \( a \) (remember: in this case we had added an extra state before the left state). This forms the loop, which we break away from by adding a free transition from \( a \) to a new state.

Now consider the case \( n > 0 \). Repeat the group \( n \) times, chaining various copies with a free transition. Add submatch tracking only to the last copy, then add a free transition from the right end back to the left end of the last copy, either before or after the transition to move on towards the rest of the NFA. This transition can end up before submatch tracking, but that is irrelevant since it only does so when going again through the group, recording new matches. Finally, add a state; we already have a transition pointing to it from \( \text{	exttt{__regex_group_repeat_aux:n}} \).
We wish to repeat the group between \#2 and \#2 + \#3 times, with a laziness controlled by \#4. We insert submatch tracking up front: in principle, we could avoid recording submatches for the first \#2 copies of the group, but that forces us to treat specially the case \#2 = 0. Repeat that group with submatch tracking \#2 + \#3 times (the maximum number of repetitions). Then our goal is to add \#3 transitions from the end of the \#2-th group, and each subsequent groups, to the end. For a lazy quantifier, we add those transitions to the left states, before submatch tracking. For the greedy case, we add the transitions to the right states, after submatch tracking and the transitions which go on with more repetitions. In the greedy case with \#2 = 0, the transition which skips over all copies of the group must be added separately, because its starting state does not follow the normal pattern: we had to add it “by hand” earlier.

\begin{verbatim}
\cs_new_protected:Npn \__regex_group_repeat:nnnN #1#2#3#4
{\__regex_group_submatches:nNN {#1}
 \l__regex_left_state_int \l__regex_right_state_int
 \__regex_group_repeat_aux:n { #2 + #3 }
 \if_meaning:w \c_true_bool #4
 \int_set_eq:NN \l__regex_left_state_int \l__regex_max_state_int
 \prg_replicate:nn { #3 }
 \{ \int_sub:Nn \l__regex_left_state_int
 \l__regex_internal_b_int - \l__regex_internal_a_int \}
 \__regex_build_transition_left:NNN \__regex_action_free:n
 \l__regex_left_state_int \l__regex_max_state_int
 }
\else:
 \prg_replicate:nn { #3 - 1 }
 \{ \int_sub:Nn \l__regex_right_state_int
 \l__regex_internal_b_int - \l__regex_internal_a_int \}
 \__regex_build_transition_right:mNn \__regex_action_free:n
 \l__regex_right_state_int \l__regex_max_state_int
 }
 \if_int_compare:w #2 = \c_zero_int
 \int_set:Nn \l__regex_left_state_int
 \l__regex_left_state_int - 1 \}
\else:
 \int_sub:Nn \l__regex_right_state_int
 \l__regex_internal_b_int - \l__regex_internal_a_int \}
 \fi:
 \__regex_build_transition_right:mNn \__regex_action_free:n
 \l__regex_right_state_int \l__regex_max_state_int
 \fi:
\end{verbatim}
45.4.6 Others

Usage: \_\_regex\_assertion:Nn (boolean) \{ (test) \}, where the \{test\} is either of the two other functions. Add a free transition to a new state, conditionally to the assertion test. The \_\_regex\_b\_test: test is used by the \b and \B escape: check if the last character was a word character or not, and do the same to the current character. The boundary-markers of the string are non-word characters for this purpose.
\exp_after:wN \__regex_break_true:w
\fi:
}

(End definition for \__regex_assertion:Nn and others.)

\__regex_command_K:  
Change the starting point of the 0-th submatch (full match), and transition to a new state, pretending that this is a fresh thread.
\cs_new_protected:Npn \__regex_command_K:  
{  
  \__regex_build_new_state:  
  \__regex_toks_put_right:Nx \l__regex_left_state_int  
  \__regex_action_submatch:nN { 0 } <  
  \bool_set_true:N \l__regex_fresh_thread_bool  
  \__regex_action_free:n  
  {  
    \int_eval:n  
    { \l__regex_right_state_int - \l__regex_left_state_int }  
  }  
  \bool_set_false:N \l__regex_fresh_thread_bool  
}

(End definition for \__regex_command_K:.)

45.5 Matching

We search for matches by running all the execution threads through the NFA in parallel, reading one token of the query at each step. The NFA contains “free” transitions to other states, and transitions which “consume” the current token. For free transitions, the instruction at the new state of the NFA is performed immediately. When a transition consumes a character, the new state is appended to a list of “active states”, stored in \g__regex_thread_info_intarray (together with submatch information): this thread is made active again when the next token is read from the query. At every step (for each token in the query), we unpack that list of active states and the corresponding submatch props, and empty those.

If two paths through the NFA “collide” in the sense that they reach the same state after reading a given token, then they only differ in how they previously matched, and any future execution would be identical for both. (Note that this would be wrong in the presence of back-references.) Hence, we only need to keep one of the two threads: the thread with the highest priority. Our NFA is built in such a way that higher priority actions always come before lower priority actions, which makes things work.

The explanation in the previous paragraph may make us think that we simply need to keep track of which states were visited at a given step: after all, the loop generated when matching (a?)* against a is broken, isn’t it? No. The group first matches a, as it should, then repeats; it attempts to match a again but fails; it skips a, and finds out that this state has already been seen at this position in the query: the match stops. The capturing group is (wrongly) a. What went wrong is that a thread collided with itself, and the later version, which has gone through the group one more times with an empty match, should have a higher priority than not going through the group.
We solve this by distinguishing “normal” free transitions \_\texttt{regex_action_free:n} from transitions \_\texttt{regex_action_free_group:n} which go back to the start of the group. The former keeps threads unless they have been visited by a “completed” thread, while the latter kind of transition also prevents going back to a state visited by the current thread.

45.5.1 Variables used when matching

The tokens in the query are indexed from \texttt{min_pos} for the first to \texttt{max_pos} − 1 for the last, and their information is stored in several arrays and \texttt{toks} registers with those numbers. We match without backtracking, keeping all threads in lockstep at the \texttt{curr_pos} in the query. The starting point of the current match attempt is \texttt{start_pos}, and \texttt{success_pos}, updated whenever a thread succeeds, is used as the next starting position.

The character and category codes of the token at the current position and a token list expanding to that token; the character code of the token at the previous position; the character code of the token just before a successful match; and the character code of the result of changing the case of the current token (A-Z ↔ a-z). This last integer is only computed when necessary, and is otherwise \texttt{c_max_int}. The \texttt{curr_char} variable is also used in various other phases to hold a character code.

For every character in the token list, each of the active states is considered in turn. The variable \texttt{\_\_regex_curr_state_int} holds the state of the \texttt{NFA} which is currently considered: transitions are then given as shifts relative to the current state.

The submatches for the thread which is currently active are stored in the \texttt{curr_submatches} list, which is almost a comma list, but ends with a comma. This list is stored by \texttt{\_\_regex_store_state:n} into an intarray variable, to be retrieved when matching at the next position. When a thread succeeds, this list is copied to \texttt{\_\_regex_success_submatches_tl}: only the last successful thread remains there.
This integer, always even, is increased every time a character in the query is read, and not reset when doing multiple matches. We store in \texttt{\_\_regex_state_active_intarray} the last step in which each \texttt{\langle state\rangle} in the NFA was encountered. This lets us break infinite loops by not visiting the same state twice in the same step. In fact, the step we store is equal to \texttt{step} when we have started performing the operations of \texttt{\langle toks(state)\rangle}, but not finished yet. However, once we finish, we store \texttt{step} + 1 in \texttt{\_\_regex_state_active_intarray}. This is needed to track submatches properly (see building phase). The \texttt{step} is also used to attach each set of submatch information to a given iteration (and automatically discard it when it corresponds to a past step).

\begin{verbatim}
\int_new:N \l__regex_step_int

(End definition for \l__regex_step_int.)
\end{verbatim}

All the currently active threads are kept in order of precedence in \texttt{\_\_regex_thread_info_intarray} together with the corresponding submatch information. Data in this intarray is organized as blocks from \texttt{min_thread} (included) to \texttt{max_thread} (excluded). At the start of every step, the whole array is unpacked, so that the space can immediately be reused, and \texttt{max_thread} is reset to \texttt{min_thread}, effectively clearing the array.

\begin{verbatim}
\int_new:N \l__regex_min_thread_int
\int_new:N \l__regex_max_thread_int

(End definition for \l__regex_min_thread_int and \l__regex_max_thread_int.)
\end{verbatim}

\texttt{\_\_regex_state_active_intarray} stores the last \texttt{\langle step\rangle} in which each \texttt{\langle state\rangle} was active. \texttt{\_\_regex_thread_info_intarray} stores threads to be considered in the next step, more precisely the states in which these threads are.

\begin{verbatim}
\intarray_new:Nn \g__regex_state_active_intarray \{ 65536 \}
\intarray_new:Nn \g__regex_thread_info_intarray \{ 65536 \}

(End definition for \g__regex_state_active_intarray and \g__regex_thread_info_intarray.)
\end{verbatim}

The list \texttt{\_\_regex_matched_analysis_tl} consists of a brace group containing three brace groups corresponding to the current token, with the same syntax as \texttt{\texttt{\_\_analysis_map_inline:nn}. The list \texttt{\_\_regex_matched_analysis_tl} (constructed under the \texttt{\_\_build} machinery) has one item for each token that has already been treated so far in a given match attempt: each item consists of three brace groups with the same syntax as \texttt{\_\_analysis_map_inline:nn}.

\begin{verbatim}
\tl_new:N \l__regex_matched_analysis_tl
\tl_new:N \l__regex_curr_analysis_tl

(End definition for \l__regex_matched_analysis_tl and \l__regex_curr_analysis_tl.)
\end{verbatim}

Every time a match is found, this token list is used. For single matching, the token list is empty. For multiple matching, the token list is set to repeat the matching, after performing some operation which depends on the user function. See \texttt{\_\_regex_single_match:} and \texttt{\_\_regex_multi_match:n}.

\begin{verbatim}
\tl_new:N \l__regex_every_match_tl

(End definition for \l__regex_every_match_tl.)
\end{verbatim}
When doing multiple matches, we need to avoid infinite loops where each iteration matches the same empty token list. When an empty token list is matched, the next successful match of the same empty token list is suppressed. We detect empty matches by setting \l__regex_fresh_thread_bool to true for threads which directly come from the start of the regex or from the \K command, and testing that boolean whenever a thread succeeds. The function \__regex_if_two_empty_matches:F is redefined at every match attempt, depending on whether the previous match was empty or not: if it was, then the function must cancel a purported success if it is empty and at the same spot as the previous match; otherwise, we definitely don’t have two identical empty matches, so the function is \use:n.

\bool_new:N \l__regex_fresh_thread_bool
\bool_new:N \l__regex_empty_success_bool
\cs_new_eq:NN \__regex_if_two_empty_matches:F \use:n

(End definition for \l__regex_fresh_thread_bool, \l__regex_empty_success_bool, and \__regex_if_two_empty_matches:F.)

The boolean \l__regex_match_success_bool is true if the current match attempt was successful, and \g__regex_success bool is true if there was at least one successful match. This is the only global variable in this whole module, but we would need it to be local when matching a control sequence with \c{...}. This is done by saving the global variable into \l__regex_saved_success_bool, which is local, hence not affected by the changes due to inner regex functions.

\bool_new:N \g__regex_success_bool
\bool_new:N \l__regex_saved_success_bool
\bool_new:N \l__regex_match_success_bool

(End definition for \g__regex_success_bool, \l__regex_saved_success_bool, and \l__regex_match_success_bool.)

### 45.5.2 Matching: framework

Initialize the variables that should be set once for each user function (even for multiple matches). Namely, the overall matching is not yet successful; none of the states should be marked as visited (\g__regex_state_active_intarray), and we start at step 0; we pretend that there was a previous match ending at the start of the query, which was not empty (to avoid smothering an empty match at the start). Once all this is set up, we are ready for the ride. Find the first match.

\cs_new_protected:Npn \__regex_match:n #1
  \__regex_match_init:
  \__regex_match_once_init:
  \tl_analysis_map_inline:nn {#1}
  \{ \__regex_match_one_token:nnN {##1} {##2} ##3 \}
  \__regex_match_one_token:nnN { } { -2 } F
  \prg_break_point:Nn \__regex_maplike_break: { }
}
\str_map_inline:nn {#1}
{
\tl_if_blank:nTF {#1}
   { \__regex_match_one_token:nnN {#1} {'##1} A }
   { \__regex_match_one_token:nnN {#1} {'##1} C }
}
\__regex_match_one_token:nnN { } {-2} F
\prg_break_point:Nn \__regex_maplike_break: { }
}
\cs_new_protected:Npn \__regex_match_init:
{
\bool_gset_false:N \g__regex_success_bool
\int_step_inline:nnn \l__regex_min_state_int { \l__regex_max_state_int - 1 }
\__kernel_intarray_gset:Nnn \g__regex_state_active_intarray {##1} { 1 }
\int_zero:N \l__regex_step_int
\int_set:Nn \l__regex_min_pos_int { 2 }
\int_set_eq:NN \l__regex_success_pos_int \l__regex_min_pos_int
\int_set:Nn \l__regex_last_char_success_int { -2 }
\tl_build_begin:N \l__regex_matched_analysis_tl
\tl_clear:N \l__regex_curr_analysis_tl
\int_set:Nn \l__regex_min_submatch_int { 1 }
\int_set_eq:NN \l__regex_submatch_int \l__regex_min_submatch_int
\bool_set_false:N \l__regex_empty_success_bool
}

(End definition for \__regex_match:n, \__regex_match_cs:n, and \__regex_match_init:.)

\__regex_match_once_init: This function resets various variables used when finding one match. It is called before the loop through characters, and every time we find a match, before searching for another match (this is controlled by the every_match token list).

First initialize some variables: set the conditional which detects identical empty matches; this match attempt starts at the previous success_pos, is not yet successful, and has no submatches yet; clear the array of active threads, and put the starting state 0 in it. We are then almost ready to read our first token in the query, but we actually start one position earlier than the start because \__regex_match_one_token:nnN increments \l__regex_curr_pos_int and saves \l__regex_curr_char_int as the last_char so that word boundaries can be correctly identified.

\cs_new_protected:Npn \__regex_match_once_init:
{
\if_meaning:w \c_true_bool \l__regex_empty_success_bool
\cs_set:Npn \__regex_if_two_empty_matches:F
{\int_compare:nNnF \l__regex_start_pos_int = \l__regex_curr_pos_int
\int_set_eq:NN \l__regex_start_pos_int \l__regex_curr_pos_int
}
\else:
\cs_set_eq:NN \__regex_if_two_empty_matches:F \use:n
\fi:
\int_set_eq:NN \l__regex_start_pos_int \l__regex_success_pos_int
\bool_set_false:N \l__regex_match_success_bool

513
\_\_regex_single_match: \_\_regex_multi_match:n

For a single match, the overall success is determined by whether the only match attempt is a success. When doing multiple matches, the overall matching is successful as soon as any match succeeds. Perform the action \#1, then find the next match.

\_\_regex_single_match: \_\_regex_multi_match:n

At each new position, set some variables and get the new character and category from the query. Then unpack the array of active threads, and clear it by resetting its length (max_thread). This results in a sequence of \_\_regex_use_state_and_submatches:w ⟨state⟩, ⟨submatch-clist⟩; and we consider those states one by one in order. As soon
as a thread succeeds, exit the step, and, if there are threads to consider at the next position, and we have not reached the end of the string, repeat the loop. Otherwise, the last thread that succeeded is the match. We explain the \texttt{fresh\_thread} business when describing \_\_regex\_action\_wildcard:

\begin{verbatim}
\cs_new_protected:Npn \__regex_match_one_token:nnN #1#2#3
  \int_add:Nn \l__regex_step_int { 2 }
  \int_incr:N \l__regex_curr_pos_int
  \int_set_eq:NN \l__regex_last_char_int \l__regex_curr_char_int
  \l_tl_set:Nn \l__regex_crr_token_tl {#1}
  \tl_set:Nn \l__regex_curr_token_tl {#1}
  \int_set:Nn \l__regex_curr_char_int {#2}
  \int_set:Nn \l__regex_curr_catcode_int { #3 }
  \tl_build_put_right:Nx \l__regex_matched_analysis_tl
    { \exp_not:o \l__regex_curr_analysis_tl }
  \tl_set:Nn \l__regex_curr_analysis_tl { { {#1} {#2} #3 } }
  \use:x
    \begin{verbatim}
    \int_set_eq:NN \l__regex_max_thread_int \l__regex_min_thread_int
    \int_step_function:nnN { \l__regex_min_thread_int }
    { \l__regex_max_thread_int - 1 }
    \__regex_match_one_active:n
    \end{verbatim}
  \prg_break_point:
  \bool_set_false:N \l__regex_fresh_thread_bool
  \if_int_compare:w \l__regex_max_thread_int > \l__regex_min_thread_int
    \if_int_compare:w -2 < \l__regex_curr_char_int
      \exp_after:wN \exp_after:wN \exp_after:wN \use_none:n
    \fi:
  \fi:
  \l__regex_every_match_tl
\end{verbatim}
\end{verbatim}

(End definition for \_\_regex\_match\_one\_token:nnN and \_\_regex\_match\_one\_active:n.)

\subsection{Using states of the nfa}

\_\_regex\_use\_state: Use the current NFA instruction. The state is initially marked as belonging to the current step: this allows normal free transition to repeat, but group-repeating transitions won’t. Once we are done exploring all the branches it spawned, the state is marked as step + 1: any thread hitting it at that point will be terminated.

\begin{verbatim}
\cs_new_protected:Npn \__regex_use_state:
  \begin{verbatim}
\end{verbatim}
\end{verbatim}
\_\_regex_use_state_and_submatches:w

This function is called as one item in the array of active threads after that array has been unpacked for a new step. Update the curr_state and curr_submatches and use the state if it has not yet been encountered at this step.

\cs_new_protected:Npn \_\_regex_use_state_and_submatches:w #1 , #2 ;
\{\int_set:Nn \l__regex_curr_state_int {#1}\ifs_int_compare:w\__kernel_intarray_item:Nn \g__regex_state_active_intarray{\l__regex_curr_state_int}<\l__regex_step_int\tl_set:Nn \l__regex_curr_submatches_tl {#2, }\exp_after:wN \_\_regex_use_state:\fi:\scan_stop:\}\}

(End definition for \_\_regex_use_state_and_submatches:w.)

45.5.4 Actions when matching

\_\_regex_action_start wildcard:N

For an unanchored match, state 0 has a free transition to the next and a costly one to itself, to repeat at the next position. To catch repeated identical empty matches, we need to know if a successful thread corresponds to an empty match. The instruction resetting \l__regex_fresh_thread_bool may be skipped by a successful thread, hence we had to add it to \_\_regex_match_one_token:nnN too.

\cs_new_protected:Npn \_\_regex_action_start wildcard:N #1 
\{\bool_set_true:N \l__regex_fresh_thread_bool 
\_\_regex_action_free:n {1}\bool_set_false:N \l__regex_fresh_thread Bool \bool_if:NT #1 { \_\_regex_action_cost:n {0} } \}

(End definition for \_\_regex_action_start wildcard:N.)

\_\_regex_action_free:n
\_\_regex_action_free_group:n
\_\_regex_action_free_aux:n

These functions copy a thread after checking that the NFA state has not already been used at this position. If not, store submatches in the new state, and insert the instructions for that state in the input stream. Then restore the old value of \l__regex_curr_state_int and of the current submatches. The two types of free transitions differ by how they test that the state has not been encountered yet: the group version is stricter, and will not use a state if it was used earlier in the current thread, hence forcefully breaking the loop, while the “normal” version will revisit a state even within the thread itself.

\cs_new_protected:Npn \_\_regex_action_free:n

516
\_\_regex_action_free_aux:nn { > \l\_\_regex_step_int \else: } }
\cs_new_protected:Npn \_\_regex_action_free_group:n
{ \_\_regex_action_free_aux:nn { < \l\_\_regex_step_int } }
\cs_new_protected:Npn \_\_regex_action_free_aux:nn #1\#2
{
\use:x
{
\int_add:Nn \l\_\_regex_curr_state_int {#2}
\exp_not:n
{
\if_int_compare:w
\_\_kernel_intarray_item:Nn \g\_\_regex_state_active_intarray
{ \l\_\_regex_curr_state_int }
#1
\exp_after:wN \_\_regex_use_state:
\fi:
}
\int_set:Nn \l\_\_regex_curr_state_int
{ \int_use:N \l\_\_regex_curr_state_int }
\tl_set:Nn \exp_not:N \l\_\_regex_curr_submatches_tl
{ \exp_not:o \l\_\_regex_curr_submatches_tl }
}
}

(End definition for \_\_regex_action_free:n, \_\_regex_action_free_group:n, and \_\_regex_action_free_aux:nn.)

\_\_regex_action_cost:n
A transition which consumes the current character and shifts the state by \#1. The resulting state is stored in the appropriate array for use at the next position, and we also store the current submatches.
\cs_new_protected:Npn \_\_regex_action_cost:n #1
{
\exp_args:Nx \_\_regex_store_state:n
{ \int_eval:n { \l\_\_regex_curr_state_int + #1 } }
}

(End definition for \_\_regex_action_cost:n.)

\_\_regex_store_state:n
\_\_regex_store_submatches:
Put the given state and current submatch information in \g\_\_regex_thread_info_intarray, and increment the length of the array.
\cs_new_protected:Npn \_\_regex_store_state:n #1
{
\exp_args:No \_\_regex_store_submatches:nn
{ \l\_\_regex_curr_submatches_tl (#1) }
\int_incr:N \l\_\_regex_max_thread_int
}
\cs_new_protected:Npn \_\_regex_store_submatches:nn #1\#2
{
\_\_kernel_intarray_gset_range_from_clist:Nnn
\g\_\_regex_thread_info_intarray
{ \_\_regex_int_eval:w
1 + \l\_\_regex_max_thread_int * (\l\_\_regex_capturing_group_int * 2 + 1) }

517
\_\_regex_disable_submatches:

Some user functions don’t require tracking submatches. We get a performance improvement by simply defining the relevant functions to remove their argument and do nothing with it.

\begin{verbatim}
\cs_new_protected:Npn \__regex_disable_submatches:
{ \cs_set_protected:Npn \__regex_store_submatches:n ##1 { } }
\end{verbatim}

(End definition for \__regex_disable_submatches:)

\_\_regex_action_submatch:nN

Update the current submatches with the information from the current position. Maybe a bottleneck.

\begin{verbatim}
\cs_new_protected:Npn \__regex_action_submatch:nN #1#2
{ \exp_after:wN \__regex_action_submatch_aux:w \l__regex_curr_submatches_tl ; {#1} #2 }
\cs_new_protected:Npn \__regex_action_submatch_aux:w #1 ; #2#3
{ \tl_set:Nx \l__regex_curr_submatches_tl
{ \prg_replicate:nn { #2 \if_meaning:w > #3 + \l__regex_capturing_group_int \fi: } { \__regex_action_submatch_auxii:w } \__regex_action_submatch_auxiii:w #1 } }
\cs_new:Npn \__regex_action_submatch_auxii:w #1 \__regex_action_submatch_auxiii:w #2 , { #2 , #1 \__regex_action_submatch_auxiii:w }
\cs_new:Npn \__regex_action_submatch_auxiii:w #1 , { \int_use:N \l__regex_curr_pos_int , }
\end{verbatim}

(End definition for \__regex_action_submatch:nN and others.)

\_\_regex_action_success:

There is a successful match when an execution path reaches the last state in the NFA, unless this marks a second identical empty match. Then mark that there was a successful match; it is empty if it is “fresh”; and we store the current position and submatches. The current step is then interrupted with \prg_break:, and only paths with higher precedence are pursued further. The values stored here may be overwritten by a later success of a path with higher precedence.

\begin{verbatim}
\cs_new_protected:Npn \__regex_action_success:
{ \__regex_if_two_empty_matches:F }
\end{verbatim}

(End definition for \__regex_action_success: and others.)

518
45.6 Replacement

45.6.1 Variables and helpers used in replacement

\l__regex_replacement_csnames_int

The behaviour of closing braces inside a replacement text depends on whether a sequences \{ or \} has been encountered. The number of “open” such sequences that should be closed by \} is stored in \l__regex_replacement_csnames_int, and decreased by 1 by each \}.

\int_new:N \l__regex_replacement_csnames_int

(End definition for \l__regex_replacement_csnames_int.)

\l__regex_replacement_category_tl \l__regex_replacement_category_seq

This sequence of letters is used to correctly restore categories in nested constructions such as \c{L(abc\cD(_)d)}.

\tl_new:N \l__regex_replacement_category_tl
\seq_new:N \l__regex_replacement_category_seq

(End definition for \l__regex_replacement_category_tl and \l__regex_replacement_category_seq.)

\l__regex_balance_tl

This token list holds the replacement text for \__regex_replacement_balance_one_match:n while it is being built incrementally.

\tl_new:N \l__regex_balance_tl

(End definition for \l__regex_balance_tl.)

\__regex_replacement_balance_one_match:n

This expects as an argument the first index of a set of entries in \g__regex_submatch_begin_intarray (and related arrays) which hold the submatch information for a given match. It can be used within an integer expression to obtain the brace balance incurred by performing the replacement on that match. This combines the braces lost by removing the match, braces added by all the submatches appearing in the replacement, and braces appearing explicitly in the replacement. Even though it is always redefined before use, we initialize it as for an empty replacement. An important property is that concatenating several calls to that function must result in a valid integer expression (hence a leading + in the actual definition).

\cs_new:Npn \__regex_replacement_balance_one_match:n #1
\{ - \__regex_submatch_balance:n #1 \} 

(End definition for \__regex_replacement_balance_one_match:n.)
The input is the same as \texttt{\__regex_replacement_balance_one_match:n}. This function is redefined to expand to the part of the token list from the end of the previous match to a given match, followed by the replacement text. Hence concatenating the result of this function with all possible arguments (one call for each match), as well as the range from the end of the last match to the end of the string, produces the fully replaced token list. The initialization does not matter, but (as an example) we set it as for an empty replacement.

\begin{verbatim}
\cs_new:Npn \__regex_replacement_do_one_match:n #1
{\__regex_query_range:nn
 { \__kernel_intarray_item:Nn \g__regex_submatch_prev_intarray {#1} }
 { \__kernel_intarray_item:Nn \g__regex_submatch_begin_intarray {#1} }}
\end{verbatim}

(End definition for \texttt{\__regex_replacement_do_one_match:n})

\texttt{\__regex_replacement_exp_not:N} This function lets us navigate around the fact that the primitive \texttt{\exp_not:n} requires a braced argument. As far as I can tell, it is only needed if the user tries to include in the replacement text a control sequence set equal to a macro parameter character, such as \texttt{\c_parameter_token}. Indeed, within an x-expanding assignment, \texttt{\exp_not:N \#} behaves as a single \#, whereas \texttt{\exp_not:n \{\#\}} behaves as a doubled ##.

\begin{verbatim}
\cs_new:Npn \__regex_replacement_exp_not:N #1 { \exp_not:n {#1} }
\end{verbatim}

(End definition for \texttt{\__regex_replacement_exp_not:N})

\texttt{\__regex_replacement_exp_not:V} This is used for the implementation of \texttt{\u}, and it gets redefined for \texttt{\peek_regex_replace_once:nnTF}.

\begin{verbatim}
\cs_new_eq:NN \__regex_replacement_exp_not:V \exp_not:V
\end{verbatim}

(End definition for \texttt{\__regex_replacement_exp_not:V})

### 45.6.2 Query and brace balance

When it is time to extract submatches from the token list, the various tokens are stored in \texttt{\toks} registers numbered from \texttt{\__regex_min_pos_int} inclusive to \texttt{\__regex_max_pos_int} exclusive. The function \texttt{\__regex_query_range:nn \{\langle min\rangle\} \{\langle max\rangle\}} unpacks registers from the position \langle min\rangle to the position \langle max\rangle − 1 included. Once this is expanded, a second x-expansion results in the actual tokens from the query. That second expansion is only done by user functions at the very end of their operation, after checking (and correcting) the brace balance first.

\begin{verbatim}
\cs_new:Npn \__regex_query_range:nn #1#2
{\exp_after:wN \__regex_query_range_loop:ww
 \int_value:w \__regex_int_eval:w #1 \exp_after:wN ;
 \int_value:w \__regex_int_eval:w #2 ;
 \prg_break_point:
}
\cs_new:Npn \__regex_query_range_loop:ww #1 ; #2 ;
{\if_int_compare:w #1 < #2 \exp_stop_f:
 \else:
 \exp_after:wN \prg_break:
 \fi:
\end{verbatim}

520
\_\_regex_submatch:n

Find the start and end positions for a given submatch (of a given match).

\cs_new:Npn \_\_regex_query_submatch:n #1
\{ \_\_regex_query_range:nn { \_kernel_intarray_item:Nn \g__regex_submatch_begin_intarray {#1} } { \_kernel_intarray_item:Nn \g__regex_submatch_end_intarray {#1} } \}

(End definition for \_\_regex_query_submatch:n.)

\_\_regex_submatch_balance:n

Every user function must result in a balanced token list (unbalanced token lists cannot be stored by \TeX). When we unpacked the query, we kept track of the brace balance, hence the contribution from a given range is the difference between the brace balances at the \langle max pos \rangle and \langle min pos \rangle. These two positions are found in the corresponding “submatch” arrays.

\cs_new_protected:Npn \_\_regex_submatch_balance:n #1
\{ \int_eval:n { \_\_regex_intarray_item:NnF \g__regex_balance_intarray { \_kernel_intarray_item:Nn \g__regex_submatch_begin_intarray {#1} } { 0 } - \_\_regex_intarray_item:NnF \g__regex_balance_intarray { \_kernel_intarray_item:Nn \g__regex_submatch_end_intarray {#1} } { 0 } } \}

(End definition for \_\_regex_submatch_balance:n.)

45.6.3 Framework

The replacement text is built incrementally. We keep track in \_\_regex_balance_int of the balance of explicit begin- and end-group tokens and we store in \_\_regex_balance_int some code to compute the brace balance from submatches (see its description). Detect unescaped right braces, and escaped characters, with \texttt{prg\_do\_nothing}; because some of the later function look-ahead. Once the whole replacement text has been parsed, make sure that there is no open csname. Finally, define the balance\_one\_match and do\_one\_match functions.
\cs_new_protected:Npn \__regex_replacement:n #1
{
\group_begin:
\tl_build_begin:N \l__regex_build_tl
\int_zero:N \l__regex_balance_int
\tl_clear:N \l__regex_balance_tl
\__regex_escape_use:nnnn
{
\if_charcode:w \c_right_brace_str ##1
\__regex_replacement_rbrace:N
\else:
\if_charcode:w \c_left_brace_str ##1
\__regex_replacement_lbrace:N
\else:
\__regex_replacement_normal:n
\fi:
\fi:
##1
}
\__regex_replacement_escaped:N ##1
\__regex_replacement_normal:n ##1
\seq_if_empty:NF \l__regex_replacement_category_seq
{ \msg_error:nnx { regex } { replacement-missing-rparen } \seq_count:N \l__regex_replacement_category_seq \seq_clear:N \l__regex_replacement_category_seq }
\cs_gset:Npx \__regex_replacement_balance_one_match:n ##1
{ + \int_use:N \l__regex_balance_int
\l__regex_balance_tl
- \__regex_submatch_balance:n {##1}
}
\tl_build_end:N \l__regex_build_tl
\exp_args:NNo
\group_end:
\__regex_replacement_aux:n \l__regex_build_tl
}\cs_new_protected:Npn \__regex_replacement_aux:n #1
{
\cs_set:Npn \__regex_replacement_do_one_match:n ##1
{ \__regex_query_range:nn
{ \__kernel_intarray_item:Nn \g__regex_submatch_prev_intarray {##1}
\exp_args:NNo
\group_end:
\__regex_replacement_aux:n \l__regex_build_tl
}\cs_new_protected:Npn \__regex_replacement_aux:n #1
{ \cs_set:Npn \__regex_replacement_do_one_match:n ##1
{ \__regex_query_range:nn
{ \__kernel_intarray_item:Nn\g__regex_submatch_prev_intarray {##1}
\exp_args:NNo
\group_end:
\__regex_replacement_aux:n \l__regex_build_tl
}
6960} \__kernel_intarray_item:Nn
6961} \g__regex_submatch_begin_intarray {##1}
6962} #1
6963}
6964}
6965}
6966}
6967} (End definition for \__regex_replacement:n and \__regex_replacement_aux:n.)
6968} (End definition for \__regex_replacement_put:n.)
6969} Most characters are simply sent to the output by \tl_build_put_right:Nn, unless a particular category code has been requested: then \__regex_replacement_c_A:w or a similar auxiliary is called. One exception is right parentheses, which restore the category code in place before the group started. Note that the sequence is non-empty there: it contains an empty entry corresponding to the initial value of \l__regex_replacement_category_tl. The argument #1 is a single character (including the case of a catcode-other space). In case no specific catcode is requested, we take into account the current catcode regime (at the time the replacement is performed) as much as reasonable, with all impossible catcodes (escape, newline, etc.) being mapped to “other”.
6970} { \int_compare:nNnTF { \l__regex_replacement_csnames_int } > 0
6971} { \exp_args:No \__regex_replacement_put:n { \token_to_str:N #1 } }
6972} { \tl_if_empty:NTF \l__regex_replacement_category_tl
6973} { \__regex_replacement_normal_aux:N #1 }
6974} % ( \token_if_eq_charcode:NNTF #1 )
6975} { \seq_pop:NN \l__regex_replacement_category_seq
6976} \l__regex_replacement_category_tl
6977} \use:c { \__regex_replacement_c_ \l__regex_replacement_category_tl :w } ? #1
6978}
6979}
6980} { \token_if_eq_charcode:NNTF #1 \c_space_token
6981} \c_space_token
6982} \__regex_replacement_c_S:w }
6983} \exp_after:wN \exp_after:wN
6984} \if_case:w \tex_catcode:D ‘#1 \exp_stop_f:
6985} \__regex_replacement_c_0:w
6986

523
As in parsing a regular expression, we use an auxiliary built from \#1 if defined. Otherwise, check for escaped digits (standing from submatches from 0 to 9): anything else is a raw character.

\cs_new_protected:Npn \__regex_replacement_escaped:N \#1
\cs_if_exist_use:cF { \__regex_replacement_#1:w }
\if_int_compare:w 1 < \#1 \exp_stop_f:
    \__regex_replacement_put_submatch:n \{#1\}
\else:
    \__regex_replacement_normal:n \{#1\}
\fi:
\end.definition for \__regex_replacement_escaped:N.

45.6.4 Submatches

Insert a submatch in the replacement text. This is dropped if the submatch number is larger than the number of capturing groups. Unless the submatch appears inside a \c{...} or \u{...} construction, it must be taken into account in the brace balance. Later on, \#1 will be replaced by a pointer to the 0-th submatch for a given match. There is an \exp_not:N here as at the point-of-use of \l__regex_balance_tl there is an \x-type expansion which is needed to get \#1 in correctly.

\cs_new_protected:Npn \__regex_replacement_put_submatch:n \#1
\cs_if_int_compare:w \#1 < \l__regex_capturing_group_int \__regex_replacement_put_submatch:n \{#1\}
\else:
    \msg_expandable_error:nnff { regex } { submatch-too-big }
    \{#1\} \{ \int_eval:n \{ \l__regex_capturing_group_int - 1 \} \}
\fi:
\cs_new_protected:Npn \_regex_replacement_put_submatch_aux:n \#1
\{
  \tl_build_put_right:Nn \l__regex_build_tl
  { \_regex_query_submatch:n \{ \int_eval:n { \#1 + \##1 } \} }
\if_int_compare:w \l__regex_replacement_csnames_int = \c_zero_int
  \tl_put_right:Nn \l__regex_balance_tl
  { \exp_not:N \int_eval:n { \#1 + \##1 } }
\fi:
\}
\exp_stop_f:  
\}(End definition for \_regex_replacement_put_submatch:n and \_regex_replacement_put_submatch_minus_aux:n.)

\_regex_replacement_g:w \_regex_replacement_g_digits:NN  
Grab digits for the \textbackslash g escape sequence in a primitive assignment to the integer \_\_regex_internal_a_int. At the end of the run of digits, check that it ends with a right brace.
\cs_new_protected:Npn \_regex_replacement_g:w #1#2
\{
  \token_if_eq_meaning:NNTF #1 \_regex_replacement_lbrace:N
    { \l__regex_internal_a_int = \_regex_replacement_g_digits:NN }
  { \_regex_replacement_error:NNN g #1 #2 }
\}
\cs_new:Npn \_regex_replacement_g_digits:NN #1#2
\{
  \token_if_eq_meaning:NNTF #1 \_regex_replacement_normal:n
  { \if_int_compare:w 1 < 1#2 \exp_stop_f:
     \#2
\exp_after:wN \use_i:nnn
\exp_after:wN \_regex_replacement_g_digits:NN
\else:
  \exp_stop_f:
\exp_after:wN \_regex_replacement_error:NNN g
\exp_after:wN \_regex_replacement_normal:n
\fi:}
\}
\{
  \exp_stop_f:
\if_meaning:w \_regex_replacement_rbrace:N #1
  \exp_args:No \_regex_replacement_put_submatch:n
    { \int_use:N \l__regex_internal_a_int }
  \exp_after:wN \use_none:nn
\else:
  \exp_after:wN \_regex_replacement_error:NNN g
\fi:
\}
\#1 \#2
\}
\}(End definition for \_regex_replacement_g:w and \_regex_replacement_g_digits:NN.)
45.6.5 Csnames in replacement

_c_regex_replacement_c:w c may only be followed by an unescaped character. If followed by a left brace, start a
control sequence by calling an auxiliary common with \u. Otherwise test whether the
category is known; if it is not, complain.
\cs_new_protected:Npn \_regex_replacement_c:w \#1#2
\{
\token_if_eq_meaning:NNTF \#1 \_regex_replacement_normal:n
\{
\cs_if_exist:cTF { \_regex_replacement_c_#2:w }
\{ \_regex_replacement_cat:NNN \#2 \}
\{ \_regex_replacement_error:NNN c \#1#2 \}
\}
\{
\token_if_eq_meaning:NNTF \#1 \_regex_replacement_lbrace:N
\{ \_regex_replacement_cu_aux:Nw \_regex_replacement_exp_not:N \}
\{ \_regex_replacement_error:NNN c \#1#2 \}
\}
\}
(End definition for \_regex_replacement_c:w.)

\_regex_replacement_cu_aux:Nw Start a control sequence with \cs:w, protected from expansion by #1 (either \_regex_replacement_exp_not:N or \exp_not:V), or turned to a string by \tl_to_str:V if inside
another csnames construction \c or \u. We use \tl_to_str:V rather than \tl_to_str:N
to deal with integers and other registers.
\cs_new_protected:Npn \_regex_replacement_cu_aux:Nw \#1
\{
\if_case:w \l__regex_replacement_csnames_int
\tl_build_put_right:Nn \l__regex_build_tl \{ \exp_not:n { \exp_after:wN \cs:w } \}
\else:
\tl_build_put_right:Nn \l__regex_build_tl \{ \exp_not:n { \exp_after:wN \tl_to_str:V \cs:w } \}
\fi:
\int_incr:N \l__regex_replacement_csnames_int
\}
(End definition for \_regex_replacement_cu_aux:Nw.)

\_regex_replacement_u:w Check that \u is followed by a left brace. If so, start a control sequence with \cs:w, which is then unpacked either with \exp_not:V or \tl_to_str:V depending on the
current context.
\cs_new_protected:Npn \_regex_replacement_u:w \#1#2
\{
\token_if_eq_meaning:NNTF \#1 \_regex_replacement_lbrace:N
\{ \_regex_replacement_cu_aux:Nw \_regex_replacement_exp_not:V \}
\{ \_regex_replacement_error:NNN u \#1#2 \}
\}
(End definition for \_regex_replacement_u:w.)
Within a \{...\} or \u{...} construction, end the control sequence, and decrease the brace count. Otherwise, this is a raw right brace.

\cs_new_protected:Npn \__regex_replacement_rbrace:N #1
\begin{verbatim}
\if_int_compare:w \l__regex_replacement_csnames_int > \c_zero_int
 \tl_build_put_right:Nn \l__regex_build_tl { \cs_end: }
 \int_decr:N \l__regex_replacement_csnames_int
 \else:
 \__regex_replacement_normal:n {#1}
 \fi:
\end{verbatim}

(End definition for \__regex_replacement_rbrace:N.)

Within a \{...\} or \u{...} construction, this is forbidden. Otherwise, this is a raw left brace.

\cs_new_protected:Npn \__regex_replacement_lbrace:N #1
\begin{verbatim}
\if_int_compare:w \l__regex_replacement_csnames_int > \c_zero_int
 \msg_error:nnn { regex } { cu-lbrace } { u }
 \else:
 \__regex_replacement_normal:n {#1}
 \fi:
\end{verbatim}

(End definition for \__regex_replacement_lbrace:N.)

45.6.6 Characters in replacement

Here, #1 is a letter among BEMTPUDSLOA and #2#3 denote the next character. Complain if we reach the end of the replacement or if the construction appears inside \{...\} or \u{...}, and detect the case of a parenthesis. In that case, store the current category in a sequence and switch to a new one.

\cs_new_protected:Npn \__regex_replacement_cat:NNN #1#2#3
\begin{verbatim}
\token_if_eq_meaning:NNTF #3
 { \msg_error:nn { regex } { replacement-catcode-end } }
 \int_compare:nNnTF { \l__regex_replacement_csnames_int } > 0
  { \msg_error:nnnn { regex } { replacement-catcode-in-cs } {#1} {#3} #2 #3 }
 \__regex_two_if_eq:NNNNTF #2 #3 \__regex_replacement_normal:n (\seq_push:NV \l__regex_replacement_category_seq
 \l__regex_replacement_category_tl
 \tl_set:NN \l__regex_replacement_category_tl {#1})
 \token_if_eq_meaning:NNT \__regex_replacement_escaped:N
\end{verbatim}
\newcommand{\__regex_char_if_alphanumeric}{NTF \#3}
\newcommand{\msg_error}{nnnn
{ regex } { replacement-catcode-escaped }
{#1} {#3}
{ }
\use:c { \__regex_replacement_c_{#1}:w } #2 #3}

(End definition for \_\_regex_replacement_cat:NNN.)

We now need to change the category code of the null character many times, hence
work in a group. The catcode-specific macros below are defined in alphabetical order;
if you are trying to understand the code, start from the end of the alphabet as those
categories are simpler than active or begin-group.

\group_begin:
\__regex_replacement_char:nNN
The only way to produce an arbitrary character–catcode pair is to use the \lowercase
or \uppercase primitives. This is a wrapper for our purposes. The first argument is the
null character with various catcodes. The second and third arguments are grabbed from
the input stream: \#3 is the character whose character code to reproduce. We could use
\char_generate:nn but only for some catcodes (active characters and spaces are not
supported).
\cs_new_protected:Npn \__regex_replacement_char:nNN #1#2#3
{ \tex_lccode:D 0 = '#3 \scan_stop:
\tex_lowercase:D { \__regex_replacement_put:n {#1} }
}

(End definition for \_\_regex_replacement_char:nNN.)

\_\_regex_replacement_c_A:w
For an active character, expansion must be avoided, twice because we later do two x-
expansions, to unpack \toks for the query, and to expand their contents to tokens of the
query.
\char_set_catcode_active:N \^^@ \cs_new_protected:Npn \_\_regex_replacement_c_A:w
{ \_\_regex_replacement_char:nNN \exp_not:n { \exp_not:N \^^@ } }

(End definition for \_\_regex_replacement_c_A:w.)

\_\_regex_replacement_c_B:w
An explicit begin-group token increases the balance, unless within a \c{...} or \u{...}
construction. Add the desired begin-group character, using the standard \if_false:
trick. We eventually x-expand twice. The first time must yield a balanced token list,
and the second one gives the bare begin-group token. The \exp_after:wN is not strictly
needed, but is more consistent with \l3tl-analysis.
\char_set_catcode_group_begin:N \^^@ \cs_new_protected:Npn \_\_regex_replacement_c_B:w
{ }
\__regex_replacement_c_B:w \_ \lowercase \_l_replacement csnames_int = \c_zero_int \int_incr:N \_ \lowercase \_l_replacement balance_int \fi: \_ \lowercase \_l_replacement_char:nN \# \if_int_compare:w \l__regex_replacement_csnames_int = \c_zero_int \int_decr:N \l__regex_balance_int \fi: \{ \exp_not:n { \exp_after:wN \^^@ \if_false: } \fi: \} \} \_ \lowercase \_l_replacement_char:nN \{ \exp_not:n { \exp_after:wN \^^@ \if_false: } \fi: \} \} \} \} \} \} \} \end{definition for \_ \lowercase \_l_replacement c_B:w.\}

\_ \lowercase \_l_replacement c_C:w This is not quite catcode-related: when the user requests a character with category “control sequence”, the one-character control symbol is returned. As for the active character, we prepare for two \_ \lowercase \_l_replacement x-expansions.
\cs_new_protected:Npn \_ \lowercase \_l_replacement c_C:w \# \if_int_compare:w \l__regex_replacement_csnames_int = \c_zero_int \int_decr:N \l__regex_balance_int \fi: \_ \lowercase \_l_replacement_char:nN \{ \exp_not:n { \exp_after:wN \^^@ \if_false: } \fi: \} \} \end{definition for \_ \lowercase \_l_replacement c_C:w.\}

\_ \lowercase \_l_replacement c_D:w Subscripts fit the mould: \lowercase the null byte with the correct category.
\char_set_catcode_math_subscript:N \^^@ \cs_new_protected:Npn \_ \lowercase \_l_replacement c_D:w \{ \_ \lowercase \_l_replacement_char:nN \{ \^^@ \} \} \end{definition for \_ \lowercase \_l_replacement c_D:w.\}

\_ \lowercase \_l_replacement c_E:w Similar to the begin-group case, the second \_ \lowercase \_l_replacement x-expansion produces the bare end-group token.
\char_set_catcode_group_end:N \^^@ \cs_new_protected:Npn \_ \lowercase \_l_replacement c_E:w \{ \_ \lowercase \_l_replacement_char:nN \{ \^^@ \} \} \end{definition for \_ \lowercase \_l_replacement c_E:w.\}

\_ \lowercase \_l_replacement c_L:w Simply \lowercase a letter null byte to produce an arbitrary letter.
\char_set_catcode_letter:N \^^@ \cs_new_protected:Npn \_ \lowercase \_l_replacement c_L:w \{ \_ \lowercase \_l_replacement_char:nN \{ \^^@ \} \} \end{definition for \_ \lowercase \_l_replacement c_L:w.\}

\_ \lowercase \_l_replacement c_M:w Similar to the begin-group case, the second \_ \lowercase \_l_replacement x-expansion produces the bare end-group token.
\char_set_catcode_math_toggle:N \^^@ \cs_new_protected:Npn \_ \lowercase \_l_replacement c_M:w \{ \_ \lowercase \_l_replacement_char:nN \{ \^^@ \} \} \end{definition for \_ \lowercase \_l_replacement c_M:w.\}
\_\_regex_replacement_c\_0:w  Lowercase an other null byte.

\_\_regex_replacement_c\_P:w  For macro parameters, expansion is a tricky issue. We need to prepare for two x-expansions and passing through various macro definitions. Note that we cannot replace one \exp_not:n by doubling the macro parameter characters because this would misbehave if a mischievous user asks for \c{\cP\#}, since that macro parameter character would be doubled.

\_\_regex_replacement_c\_S:w  Spaces are normalized on input by \TeX{} to have character code \texttt{32}. It is in fact impossible to get a token with character code \texttt{0} and category code \texttt{10}. Hence we use \texttt{32} instead of \texttt{0} as our base character.

\_\_regex_replacement_c\_T:w  No surprise for alignment tabs here. Those are surrounded by the appropriate braces whenever necessary, hence they don’t cause trouble in alignment settings.

\_\_regex_replacement_c\_U:w  Simple call to \_\_regex_replacement_char:nNN which lowercases the math superscript ^^@.

Restore the catcode of the null byte.
45.6.7 An error

Simple error reporting by calling one of the messages replacement-c, replacement-g, or replacement-u.

\__regex_replacement_error:NNN
\cs_new_protected:Npn \__regex_replacement_error:NNN #1#2#3
{ \msg_error:nnnx { regex } { replacement-#1 } {#3} #2 #3 }

(End definition for \__regex_replacement_error:NNN.)

45.7 User functions

\regex_new:N Before being assigned a sensible value, a regex variable matches nothing.

\cs_new_protected:Npn \regex_new:N #1
{ \cs_new_eq:NN #1 \c__regex_no_match_regex }

(End definition for \regex_new:N. This function is documented on page 54.)

\l_tmpa_regex \l_tmpb_regex \g_tmpa_regex \g_tmpb_regex

The usual scratch space.

\regex_new:N \l_tmpa_regex \regex_new:N \l_tmpb_regex \regex_new:N \g_tmpa_regex \regex_new:N \g_tmpb_regex

(End definition for \l_tmpa_regex and others. These variables are documented on page 57.)

\regex_set:Nn \regex_gset:Nn \regex_const:Nn

Compile, then store the result in the user variable with the appropriate assignment function.

\cs_new_protected:Npn \regex_set:Nn #1#2
{ \__regex_compile:n {#2} \tl_set_eq:NN #1 \l__regex_internal_regex }
\cs_new_protected:Npn \regex_gset:Nn #1#2
{ \__regex_compile:n {#2} \tl_gset_eq:NN #1 \l__regex_internal_regex }
\cs_new_protected:Npn \regex_const:Nn #1#2
{ \__regex_compile:n {#2} \tl_const:Nx #1 { \exp_not:o \l__regex_internal_regex } }

(End definition for \regex_set:Nn, \regex_gset:Nn, and \regex_const:Nn. These functions are documented on page 54.)

\regex_show:n \regex_log:n \__regex_show:N \regex_show:N \regex_gset:Nn \regex_log:N \__regex_show:NN

User functions: the n variant requires compilation first. Then show the variable with some appropriate text. The auxiliary \__regex_show:N is defined in a different section.

\cs_new_protected:Npn \regex_show:n { \__regex_show:N \msg_show:nnxxxx }
\cs_new_protected:Npn \regex_log:n { \__regex_show:N \msg_log:nnxxxx }
\cs_new_protected:Npn \__regex_show:NN { \__regex_show:NN \msg_show:nnxxxx }
\cs_new_protected:Npn \regex_show:N \regex_gset:Nn \regex_log:N \__regex_show:NN

531
We define here 40 user functions, following a common pattern in terms of :nnN auxiliaries, defined in the coming subsections. The auxiliary is handed \_\_regex_build:n or \_\_regex_build:N with the appropriate regex argument, then all other necessary arguments (replacement text, token list, etc. The conditionals call \_\_regex_return: to return either true or false once matching has been performed.
45.7.1 Variables and helpers for user functions

The number of matches found so far is stored in \l__regex_match_count_int. This is only used in the \regex_count:nnN functions.

Those flags are raised to indicate begin-group or end-group tokens that had to be added when extracting submatches.

The end-points of each submatch are stored in two arrays whose index (submatch) ranges from \l__regex_min_submatch_int (inclusive) to \l__regex_submatch_int (exclusive). Each successful match comes with a 0-th submatch (the full match), and one match for each capturing group: submatches corresponding to the last successful match are labelled starting at zeroth_submatch. The entry \l__regex_zeroth_submatch_int in \g__regex_submatch_prev_intarray holds the position at which that match attempt started: this is used for splitting and replacements.

Hold the place where the match attempt begun and the end-points of each submatch.
The first thing we do when matching is to store the balance of begin-group/end-group characters into \texttt{\g__regex_balance_intarray}.

\begin{verbatim}
\intarray_new:Nn \g__regex_balance_intarray { 65536 }
\end{verbatim}

Keep track of the number of left/right braces to add when performing a regex operation such as a replacement.

\begin{verbatim}
\int_new:N \l__regex_added_begin_int
\int_new:N \l__regex_added_end_int
\end{verbatim}

This function triggers either \texttt{\prg_return_false:} or \texttt{\prg_return_true:} as appropriate to whether a match was found or not. It is used by all user conditionals.

\begin{verbatim}
\cs_new_protected:Npn \__regex_return: {
  \if_meaning:w \c_true_bool \g__regex_success_bool
  \prg_return_true:
  \else:
  \prg_return_false:
  \fi:
}
\end{verbatim}

To easily extract subsets of the input once we found the positions at which to cut, store the input tokens one by one into successive \texttt{\toks} registers. Also store the brace balance (used to check for overall brace balance) in an array.

\begin{verbatim}
\cs_new_protected:Npn \__regex_query_set:n #1 {
  \int_zero:N \l__regex_balance_int
  \int_zero:N \l__regex_curr_pos_int
  \__regex_query_set_aux:nN { } F
  \tl_analysis_map_inline:nn {#1} { \__regex_query_set_aux:nN {##1} ##3 }
  \__regex_query_set_aux:nN { } F
  \int_set_eq:NN \l__regex_max_pos_int \l__regex_curr_pos_int
}
\end{verbatim}
45.7.2 Matching

\__regex_if_match:nn

We don't track submatches, and stop after a single match. Build the NFA with \#1, and perform the match on the query \#2.

\begin{verbatim}
\cs_new_protected:Npn \__regex_if_match:nn #1#2
\group_begin:
\__regex_disable_submatches:
\__regex_single_match:
 #1
\__regex_match:n \{#2\}
\group_end:
\end{verbatim}

(End definition for \__regex_if_match:nn.)

\__regex_count:nnN

Again, we don't care about submatches. Instead of aborting after the first “longest match” is found, we search for multiple matches, incrementing \l__regex_match_count_int every time to record the number of matches. Build the NFA and match. At the end, store the result in the user's variable.

\begin{verbatim}
\cs_new_protected:Npn \__regex_count:nnN #1#2#3
\group_begin:
\__regex_disable_submatches:
\int_zero:N \l__regex_match_count_int
\__regex_multi_match:n \{ \int_incr:N \l__regex_match_count_int \}
 #1
\__regex_match:n \{#2\}
\exp_args:NNNo
\group_end:
\int_set:Nn #3 { \int_use:N \l__regex_match_count_int }
\end{verbatim}

(End definition for \__regex_count:nnN.)

45.7.3 Extracting submatches

\__regex_extract_once:nnN \__regex_extract_all:nnN

Match once or multiple times. After each match (or after the only match), extract the submatches using \__regex_extract:. At the end, store the sequence containing all the submatches into the user variable \#3 after closing the group.

\begin{verbatim}
\cs_new_protected:Npn \__regex_extract_once:nnN #1#2#3
\group_begin:
\__regex_single_match:
 #1
\__regex_match:n \{#2\}
\__regex_extract:
\__regex_query_set:n \{#2\}
\__regex_group_end_extract_seq:N #3
\group_end:
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__regex_extract_all:nnN #1#2#3
\group_begin:
\__regex_multi_match:n \{ \__regex_extract: \}
\end{verbatim}

\end{verbatim}


\_\_regex\_split:nnN

Splitting at submatches is a bit more tricky. For each match, extract all submatches, and replace the zeroth submatch by the part of the query between the start of the match attempt and the start of the zeroth submatch. This is inhibited if the delimiter matched an empty token list at the start of this match attempt. After the last match, store the last part of the token list, which ranges from the start of the match attempt to the end of the query. This step is inhibited if the last match was empty and at the very end: decrement \l__regex\_submatch\_int, which controls which matches will be used.

\cs\_new\_protected:Npn \_\_regex\_split:nnN \#1\#2\#3
{\group\_begin:
 \_\_regex\_multi\_match:n
 {\if\_int\_compare:w
 \l__regex\_start\_pos\_int < \l__regex\_success\_pos\_int
 \_\_regex\_extract:
 \_\_kernel\_intarray\_gset:Nnn \g__regex\_submatch\_prev\_intarray
 { \l__regex\_zeroth\_submatch\_int } { 0 } 
 \_\_kernel\_intarray\_gset:Nnn \g__regex\_submatch\_end\_intarray
 { \l__regex\_zeroth\_submatch\_int } 
 { \_\_kernel\_intarray\_item:Nn \g__regex\_submatch\_begin\_intarray
 { \l__regex\_zeroth\_submatch\_int } 
 \_\_kernel\_intarray\_gset:Nnn \g__regex\_submatch\_begin\_intarray
 { \l__regex\_zeroth\_submatch\_int } 
 { \l__regex\_start\_pos\_int } 
 \fi:}
 \_\_regex\_match:n \#2
 \_\_regex\_query\_set:n \#2
 \_\_kernel\_intarray\_gset:Nnn \g__regex\_submatch\_prev\_intarray
 { \l__regex\_submatch\_int } { 0 } 
 \_\_kernel\_intarray\_gset:Nnn \g__regex\_submatch\_end\_intarray
 { \l__regex\_submatch\_int } 
 { \l__regex\_max\_pos\_int } 
 \_\_kernel\_intarray\_gset:Nnn \g__regex\_submatch\_begin\_intarray
 { \l__regex\_submatch\_int } 
 { \l__regex\_start\_pos\_int } 
 \int\_incr:N \l__regex\_submatch\_int
 \if\_meaning:w \c\_true\_bool \l__regex\_empty\_success\_bool
 \if\_int\_compare:w \l__regex\_start\_pos\_int = \l__regex\_max\_pos\_int
 \int\_decr:N \l__regex\_submatch\_int
 \fi:
 \fi:
 \_\_regex\_group\_end\_extract\_seq:N \#3

The end-points of submatches are stored as entries of two arrays from \texttt{\_\_regex_min_submatch_int} to \texttt{\_\_regex_submatch_int} (exclusive). Extract the relevant ranges into \texttt{\_\_regex_internal_tl}, separated by \texttt{\_\_regex_tmp:w}. We keep track in the two flags \texttt{\_\_regex_begin} and \texttt{\_\_regex_end} of the number of begin-group or end-group tokens added to make each of these items overall balanced. At this step, \texttt{\_\_regex_submatch_int} is counted as being balanced (same number of begin-group and end-group tokens). This problem is caught by \texttt{\_\_regex_extract_check:w}, explained later. After complaining about any begin-group or end-group tokens we had to add, we are ready to construct the user's sequence outside the group.

```latex
\cs_new_protected:Npn \_\_regex_group_end_extract_seq:N #1
  { \flag_clear:n { \_\_regex_begin } \flag_clear:n { \_\_regex_end } \cs_set_eq:NN \__regex_tmp:w \scan_stop: \__kernel_tl_gset:Nx \g__regex_internal_tl { \int_step_function:nnN { \l__regex_min_submatch_int } { \l__regex_submatch_int - 1 } \_\_regex_extract_seq_aux:n \__regex_tmp:w } \int_set:Nn \l__regex_added_begin_int { \flag_height:n { \_\_regex_begin } } \int_set:Nn \l__regex_added_end_int { \flag_height:n { \_\_regex_end } } \tex_afterassignment:D \_\_regex_extract_check:w \__kernel_tl_gset:Nx \g__regex_internal_tl \if_false: { \fi: } \int_compare:nNnT { \l__regex_added_begin_int + \l__regex_added_end_int } > 0 { \msg_error:nnnn { regex } { result-unbalanced } { splitting-or-extracting-submatches } { \int_use:N \l__regex_added_begin_int } { \int_use:N \l__regex_added_end_int } } \group_end: \cs_set_eq:NN \__regex_tmp:w \_\_regex_extract_map_loop:w \seq_set_from_function:NnN \_\_regex_extract_map:N \exp_not:n }
```

The \texttt{\_\_regex_extract_seq_aux:n} auxiliary builds one item of the sequence of submatches. First compute the brace balance of the submatch, then extract the submatch from the query, adding the appropriate braces and raising a flag if the submatch is not balanced.

```latex
\cs_new:Npn \_\_regex_extract_seq_aux:n #1
  { \_\_regex_tmp:w { } \exp_after:wN \_\_regex_extract_seq_aux:ww }
```

(End definition for \texttt{\_\_regex_split:nnN}).

(End definition for \texttt{\_\_regex_group_end_extract_seq:N}).
\begin{verbatim}
\cs_new:Npn \__regex_extract_seq_aux:ww #1; #2; {
\if_int_compare:w #1 < \c_zero_int
\prg_replicate:nn {-#1}
\flag_raise:n { \__regex_begin }
\exp_not:n { { \if_false: } \fi: }
\fi:
\__regex_query_submatch:n {#2}
\if_int_compare:w #1 > \c_zero_int
\prg_replicate:nn {#1}
\flag_raise:n { \__regex_end }
\exp_not:n { \if_false: { \fi: } }
\fi:
\fi:
(End definition for \__regex_extract_seq_aux:n and \__regex_extract_seq_aux:ww.)
\end{verbatim}

In \__regex_group_end_extract_seq:N we had to expand \g__regex_internal_tl to turn \if_false: constructions into actual begin-group and end-group tokens. This is done with a \__kernel_tl_gset:Nx assignment, and \__regex_extract_check:w is run immediately after this assignment ends, thanks to the \afterassignment primitive. If all of the items were properly balanced (enough begin-group tokens before end-group tokens, so \} is not) then \__regex_extract_check:w is called just before the closing brace of the \__kernel_tl_gset:Nx (thanks to our sneaky \if_false: { \fi: } construction), and finds that there is nothing left to expand. If any of the items is unbalanced, the assignment gets ended early by an extra end-group token, and our check finds more tokens needing to be expanded in a new \__kernel_tl_gset:Nx assignment. We need to add a begin-group and an end-group tokens to the unbalanced item, namely to the last item found so far, which we reach through a loop.
\cs_new:Npn \__regex_extract_check_loop:w #1 \__regex_tmp:w #2
{ #2 \exp_not:o {#1} \__regex_extract_check_loop:w \prg_do_nothing: }

Arguments of \__regex_extract_check_end:w are: #1 is the part of the item before
the extra end-group token; #2 is junk; #3 is \prg_do_nothing: followed by the not-yet-
expanded part of the item after the extra end-group token. In the replacement text,
the first brace and the \if_false: { \fi: } construction are the added begin-group and
end-group tokens (the latter being not-yet expanded, just like #3), while the closing brace
after \exp_not:o {#1} replaces the extra end-group token that had ended the assignment
early. In particular this means that the character code of that end-group token is lost.

This receives a seq internal function and maps it over all items in \g__regex_internal_tl.
This token list takes the form \__regex_tmp:w \{ item_1 \} \__regex_tmp:w \{ item_2 \} ...
\__regex_tmp:w, and the calling code has set \__regex_tmp:w equal to
\__regex_extract_map_loop:w. The loop is otherwise pretty standard, with \prg_do_nothing: to avoid losing braces.

Our task here is to store the list of end-points of submatches, and store them in appro-
priate array entries, from \l__regex_zeroth_submatch_int upwards. First, we store in
\g__regex_submatch_prev_intarray the position at which the match attempt started.
We extract the rest from the comma list \l__regex_success_submatches_tl, which
starts with entries to be stored in \g__regex_submatch_begin_intarray and continues
with entries for \g__regex_submatch_end_intarray.

\cs_new_protected:Npn \__regex_extract: ({
  \if_meaning:w \c_true_bool \g__regex_success_bool
  \int_set_eq:NN \l__regex_zeroth_submatch_int \l__regex_submatch_int
  \prg_replicate:nn \l__regex_capturing_group_int
    { \__kernel_intarray_gset:Nnn \g__regex_submatch_prev_intarray
        { \l__regex_submatch_int } { 0 } 
    \int_incr:N \l__regex_submatch_int }
  \__kernel_intarray_gset:Nnn \g__regex_submatch_prev_intarray
    { \l__regex_zeroth_submatch_int } { \l__regex_start_pos_int }
  \int_zero:N \l__regex_internal_a_int
  \exp_after:wN \__regex_extract_aux:w \l__regex_success_submatches_tl
  \prg_break_point: \__regex_use_none_delimit_by_q_recursion_stop:w ,
  \q__regex_recursion_stop
}fi:
\cs_new_protected:Npn \__regex_extract_aux:w #1 ,
  \prg_break: #1 \prg_break_point:
  \if_int_compare:w \l__regex_internal_a_int < \l__regex_capturing_group_int
    \__kernel_intarray_gset:Nnn \g__regex_submatch_begin_intarray
      { \__regex_int_eval:w \l__regex_zeroth_submatch_int + \l__regex_internal_a_int } {#1}
  \else:
    \__kernel_intarray_gset:Nnn \g__regex_submatch_end_intarray
      { \__regex_int_eval:w \l__regex_zeroth_submatch_int + \l__regex_internal_a_int - \l__regex_capturing_group_int } {#1}
  \fi:
  \int_incr:N \l__regex_internal_a_int
  \__regex_extract_aux:w
}

(End definition for \__regex_extract: and \__regex_extract_aux:w.)

\__regex_replace_once:nnN Build the NFA and the replacement functions, then find a single match. If the match failed,
simply exit the group. Otherwise, we do the replacement. Extract submatches. Compute
the brace balance corresponding to replacing this match by the replacement (this depends
on submatches). Prepare the replaced token list: the replacement function produces the
tokens from the start of the query to the start of the match and the replacement text for
this match; we need to add the tokens from the end of the match to the end of the query.
Finally, store the result in the user’s variable after closing the group: this step involves
an additional \texttt{x}-expansion, and checks that braces are balanced in the final result.

\cs_new_protected:Npn \__regex_replace_once:nnN #1#2#3
  \group_begin:
  \__regex_single_match: #1
  \exp_args:No \__regex_match:n {#3}
(End definition for __regex_replace_once:nnN.)

__regex_replace_all:nnN

Match multiple times, and for every match, extract submatches and additionally store the position at which the match attempt started. The entries from \l__regex_min_submatch_int to \l__regex_submatch_int hold information about submatches of every match in order; each match corresponds to \l__regex_capturing_group_int consecutive entries. Compute the brace balance corresponding to doing all the replacements: this is the sum of brace balances for replacing each match. Join together the replacement texts for each match (including the part of the query before the match), and the end of the query.
\int_step_function:nnnN
\l__regex_min_submatch_int \l__regex_capturing_group_int \l__regex_submatch_int - 1 \l__regex_start_pos_int \l__regex_max_pos_int
\__regex_group_end_replace:N \__regex_replace_all:nnN.

At this stage \l__regex_internal_a_tl (x-expands to the desired result). Guess from \l__regex_balance_int the number of braces to add before or after the result then try expanding. The simplest case is when \l__regex_internal_a_tl together with the braces we insert via \prg_replicate:nn give a balanced result, and the assignment ends at the \if_false: { \fi: } construction: then \__regex_group_end_replace_check:w sees that there is no material left and we successfully found the result. The harder case is that expanding \l__regex_internal_a_tl may produce extra closing braces and end the assignment early. Then we grab the remaining code using; importantly, what follows has not yet been expanded so that \__regex_group_end_replace_check:n grabs everything until the last brace in \__regex_group_end_replace_try:, letting us try again with an extra surrounding pair of braces.

45.7.5 Peeking ahead

\l__regex_peek_true_tl
\l__regex_peek_false_tl

True/false code arguments of \peek_regex:nTF or similar.

\l__regex_replacement_tl

When peeking in \peek_regex_replace_once:nnTF we need to store the replacement text.

\l__regex_input_tl
\__regex_input_item:n

Stores each token found as \__regex_input_item:n \{\langle tokens\rangle\}, where the \langle tokens\rangle o-expand to the token found, as for \tl_analysis_map_inline:nn.

\peek_regex:nTF
\peek_regex:NTF
\peek_regex_remove_once:nTF
\peek_regex_remove_once:NTF

The T and F functions just call the corresponding TF function. The four TF functions differ along two axes: whether to remove the token or not, distinguished by using \__regex_-
\peek_end: or \__regex_peek_remove_end:n (the latter case needs an argument, as we will see), and whether the regex has to be compiled or is already in an N-type variable, distinguished by calling \__regex_build_aux:Nn or \__regex_build_aux:NN. The first argument of these functions is \c_false_bool to indicate that there should be no implicit insertion of a wildcard at the start of the pattern: otherwise the code would keep looking further into the input stream until matching the regex.
\cs_new_protected:Npn \peek_regex:NTF #1
\__regex_peek:nnTF
\__regex_peek_aux:nnTF
\__regex_disable_submatches:
\__regex_single_match:
\cs_new_protected:Npn \peek_regex:NT #1#2
\cs_new_protected:Npn \peek_regex:NF #1 { \peek_regex:NTF {#1} { } }
\cs_new_protected:Npn \peek_regex_remove_once:NTF #1
\__regex_peek:nnTF
\__regex_peek_aux:nnTF
Store the user’s true/false codes (plus \group_end:) into two token lists. Then build the automaton with #1, without submatch tracking, and aiming for a single match. Then start matching by setting up a few variables like for any regex matching like \regex_match:nnTF, with the addition of \l\_regex_input_tl that keeps track of the tokens seen, to reinsert them at the end. Instead of \tl_analysis_map_inline:nn on the input, we call \peek_analysis_map_inline:n to go through tokens in the input stream. Since \__regex_match_one_token:nnN calls \__regex_maplike_break: we need to catch that and break the \peek_analysis_map_inline:n loop instead.
\cs_new_protected:Npn \__regex_peak:nnTF #1
\__regex_peek_aux:nnTF
\__regex_disable_submatches:
\__regex_single_match:
\cs_new_protected:Npn \__regex_peak:nnTF #1#2
\cs_new_protected:Npn \__regex_peak:nnTF #1
\__regex_disable_submatches:
\__regex_single_match:
\cs_new_protected:Npn \__regex_peak:nnTF #1#2#3#4
\group_begin:
\tl_set:Nn \l\_regex_peak_true_tl { \group_end: #3 }
\tl_set:Nn \l\_regex_peak_false_tl { \group_end: #4 }
\__regex_single_match:
Once the regex matches (or permanently fails to match) we call \_\_regex_end: or \_\_regex_remove_end:n with argument the last token seen. For \peek:nTF we reinsert tokens seen by calling \_\_regex_reinsert:n regardless of the result of the match. For \peek_remove_once:nTF we reinsert the tokens seen only if the match failed; otherwise we just reinsert the tokens #1, with one expansion. To be more precise, #1 consists of tokens that o-expand and x-expand to the last token seen, for example it is \exp_not:N ⟨cs⟩ for a control sequence. This means that just doing \exp_after:wN \l__regex_true_tl #1 would be unsafe because the expansion of ⟨cs⟩ would be suppressed.

\cs_new_protected:Npn \_\_regex_end: {\bool_if:NTF \g__regex_success_bool { \_\_regex_reinsert:n \l__regex_true_tl } }{ \_\_regex_reinsert:n \l__regex_false_tl }\cs_new_protected:Npn \_\_regex_remove_end:n #1 {\bool_if:NTF \g__regex_success_bool { \exp_args:NNo \use:nn \l__regex_true_tl {#1} } }{ \_\_regex_reinsert:n \l__regex_false_tl }\cs_new_protected:Npn \_\_regex_reinsert:n #1 {\tl_build_end:N \l__regex_input_tl \cs_set_eq:NN \_\_regex_reinsert_item:n \_\_regex_item:n {#1} \exp_after:wN \l__regex_input_tl \exp_end:} (End definition for \_\_regex_end: and \_\_regex_remove_end:n.)\_\_regex_reinsert:n Insert the true/false code #1, followed by the tokens found, which were stored in \l__regex_input_tl. For this, loop through that token list using \_\_regex_reinsert_item:n, which expands #1 once to get a single token, and jumps over it to expand what follows, with suitable \exp:w and \exp_end:. We cannot just use \use:e on the whole token list because the result may be unbalanced, which would stop the primitive prematurely, or let it continue beyond where we would like.
Similar to \peek_regex:nTF above.

\peek_regex_replace_once:nnTF
\peek_regex_replace_once:NnTF
\peek_regex_replace_once:NnT
\peek_regex_replace_once:NnF
\peek_regex_replace_once:Nn

(End definition for \peek_regex_replace_once:nnTF and \peek_regex_replace_once:NnTF. These functions are documented on page 195.)

\__regex_peek_replace:nnTF
Same as \__regex_peek:nTF (used for \peek_regex:nTF above), but without disabling submatches, and with a different end. The replacement text #2 is stored, to be analyzed later.

(End definition for \__regex_peek_replace:nnTF.)

If the match failed \__regex_peek_reinsert:N reinserts the tokens found. Otherwise, finish storing the submatch information using \__regex_extract:, and store the input into \toks. Redefine a few auxiliaries to change slightly their expansion behaviour as explained below. Analyse the replacement text with \__regex_replacement:n, which as usual defines \__regex_replacement_do_one_match:n to insert the tokens from the start of the match attempt to the beginning of the match, followed by the replacement text. The \use:x expands for instance the trailing \__regex_query_range:nn down to a sequence of \__regex_reinsert_item:n \{⟨tokens⟩⟩ where ⟨tokens⟩ o-expand to a single token that we want to insert. After x-expansion, \use:x does \use:n, so we have \exp_after:wN \exp_after:wN \__regex_peek_reinsert_item:n \exp:w ... \exp_end:. This is set up such as to
obtain \texttt{\_\_regex\_peek\_true\_tl} followed by the replaced tokens (possibly unbalanced) in the input stream.

```
\cs_new_protected:Npn \_\_regex\_peek\_replace\_end:n {
\__regex\_extract:
\cs_set_eq:NN \_\_regex\_replacement\_put:n \_\_regex\_peek\_replacement\_put:n
\cs_set_eq:NN \_\_regex\_replacement\_put\_submatch\_aux:n \_\_regex\_peek\_replacement\_put\_submatch\_aux:n
\cs_set_eq:NN \_\_regex\_input\_item:n \_\_regex\_reinsert\_item:n
\cs_set_eq:NN \_\_regex\_replacement\_exp\_not:N \_\_regex\_peek\_replacement\_token:n
\cs_set_eq:NN \_\_regex\_replacement\_exp\_not:V \_\_regex\_peek\_replacement\_var:N
\exp_args:N \_\_regex\_replacement:n { \l__regex\_replacement\_tl }
\use:x
\exp_not:n { \exp_after:wN \l__regex\_peek\_true\_tl \exp:w }
\_\_regex\_replacement\_do\_one\_match:n
\__regex\_zeroth\_submatch\_int
\__regex\_query\_range:nn
\__kernel\_intarray\_item:Nn \g__regex\_submatch\_end\_intarray
\__regex\_zeroth\_submatch\_int
\__regex\_max\_pos\_int
\exp_end:
\exp_after:wN \l__regex\_peek\_false\_tl }
(End definition for \_\_regex\_peek\_replace\_end:n.)
```

\texttt{\_\_regex\_set\_from\_input\_tl:} The input was stored into \texttt{\_\_regex\_input\_tl} as successive items \texttt{\_\_regex\_input\_item:n \{\{tokens\}\}}. Store that in successive \texttt{\toks}. It’s not clear whether the empty entries before and after are both useful.

```
\cs_new_protected:Npn \_\_regex\_set\_from\_input\_tl: {
\tl_build_end:N \l__regex\_input\_tl
\int_zero:N \l__regex\_curr\_pos\_int
\cs_eq:NN \_\_regex\_set\_item:n \_\_regex\_query\_set\_item:n
\_\_regex\_query\_set\_item:n { }
\_\_regex\_input\_tl
\__kernel\_intarray\_item:Nn \g__regex\_submatch\_end\_intarray
\__regex\_zeroth\_submatch\_int
\_\_regex\_max\_pos\_int
\\}
\cs_new_protected:Npn \_\_regex\_set\_item:n #1 {
\int_incr:N \l__regex\_curr\_pos\_int
\__regex\_toks\_set:Nn \l__regex\_curr\_pos\_int { \_\_regex\_input\_item:n \{#1\} }
\}
(End definition for \_\_regex\_set\_from\_input\_tl: and \_\_regex\_set\_item:n.)
```
While building the replacement function \_\_regex_replacement_do_one_match:n, we often want to put simple material, given as \#1, whose x-expansion o-expands to a single token. Normally we can just add the token to \_\_regex_build_tl, but for \peek-_\_regex_replace_once:nnTF we eventually want to do some strange expansion that is basically using \exp_after:wN to jump through numerous tokens (we cannot use x-expansion like for \regex_replace_once:nnNTF because it is ok for the result to be unbalanced since we insert it in the input stream rather than storing it. When within a csname we don’t do any such shenanigan because \cs:w ... \cs_end: does all the expansion we need.

\cs_new_protected:Npn \_\_regex_peek_replacement_put:n \#1
\{ \if_case:w \l__regex_replacement_csnames_int 
\tl_build_put_right:Nn \l__regex_build_tl 
\{ \exp_not:N \_\_regex_reinsert_item:n \{\#1\} \} \else: 
\tl_build_put_right:Nn \l__regex_build_tl \{\#1\} \fi: \}

(End definition for \_\_regex_peek_replacement_put:n.)

\cs_new_protected:Npn \_\_regex_peek_replacement_token:n \#1
\{ \exp_after:wN \exp_end: \exp_after:wN \exp:w \#1 \exp:w \}

(End definition for \_\_regex_peek_replacement_token:n.)

\_\_regex_peek_replacement_put_submatch_aux:n
While analyzing the replacement we also have to insert submatches found in the query. Since query items \_\_regex_input_item:n \{(tokens)\} expand correctly only when surrounded by \exp:w ... \exp_end:, and since these expansion controls are not there within csnames (because \cs:w ... \cs_end: make them unnecessary in most cases), we have to put \exp:w and \exp_end: by hand here.

\cs_new_protected:Npn \_\_regex_peek_replacement_put_submatch_aux:n \#1
\{ \if_case:w \l__regex_replacement_csnames_int 
\tl_build_put_right:Nn \l__regex_build_tl 
\{ \_\_regex_query_submatch:n \{ \int_eval:n \{ \#1 + ##1 \} \} \} \else: 
\tl_build_put_right:Nn \l__regex_build_tl 
\{ \exp:w \_\_regex_query_submatch:n \{ \int_eval:n \{ \#1 + ##1 \} \} \exp_end: \} \fi: \}

(End definition for \_\_regex_peek_replacement_put_submatch_aux:n.)

\_\_regex_peek_replacement_var:N
This is used for \u outside csnames. It makes sure to continue expansion with \exp:w before expanding the variable \#1 and stopping the \exp:w that precedes.

\cs_new_protected:Npn \_\_regex_peek_replacement_var:N \#1
\{ \exp_after:wN \exp_last_unbraced:NV \exp_after:wN \exp_end: \exp_after:wN \#1 \}

548
Messages for the preparser phase.

Invalid quantifier.

Messages for missing or extra closing brackets and parentheses, with some fancy singular/plural handling for the case of parentheses.
LaTeX came across a closing parenthesis when no submatch group was open. The parenthesis will be ignored.

Some escaped alphanumerics are not allowed everywhere.

\msg_new:nnnn { regex } { bad-escape }
{ Invalid-escape-`\iow_char:N\#1`- \_regex_if_in_cs:TF { within-a-control-sequence. }
{ \_regex_if_in_class:TF
{ in-a-character-class. }
{ following-a-category-test. }
}
}

\_regex_if_in_cs:TF
{ The-escape-sequence-`\iow_char:N\#1`-may-not-appear-
\_regex_if_in_cs:TF
{ within-a-control-sequence-test-introduced-by-
`\iow_char:N\c\iow_char:N\{`. }
{ \_regex_if_in_class:TF
{ within-a-character-class- }
{ following-a-category-test-such-as-`\iow_char:N\cL`- }
because-it-does-not-match-exactly-one-character. }
}

Range errors.
\msg_new:nnnn { regex } { range-missing-end }
{ Invalid-end-point-for-range-`\#1-\#2`-in-character-class. }
{ The-end-point-`\#2`-of-the-range-`\#1-\#2`-may-not-serve-as-an-
end-point-for-a-range:-alphanumeric-characters-should-not-be-
escaped,-and-non-alphanumeric-characters-should-be-escaped. }
\msg_new:nnnn { regex } { range-backwards }
{ Range-`\#1-\#2`-out-of-order-in-character-class. }
{ In-ranges-of-characters-`[x-y]`-appearing-in-character-classes,-
the-first-character-code-must-not-be-larger-than-the-second.-
Here,-`\#1`-has-character-code-\int_eval:n {\'#1},-while-
`\#2`-has-character-code-\int_eval:n {\'#2}. }

Errors related to |c and \u.
\msg_new:nnnn { regex } { c-bad-mode }
{ Invalid-nested-`\iow_char:N\c`-escape-in-regular-expression. }
{ The-`\iow_char:N\c`-escape-cannot-be-used-within-
a-control-sequence-test-`\iow_char:N\c(...)`-
nor-another-category-test.-
To-combine-several-category-tests,-use-`\iow_char:N\c[...]`.  

550
The `\iow_char:N\cC` construction restricts the next item to be a control sequence or the next group to be made of control sequences. It only makes sense to follow it by `.`, `(`, or by a group.

Left-braces must be escaped in `\iow_char:N\#1{...}`.

Constructions such as `\iow_char:N\#1{...\iow_char:N\{...}` are not allowed and should be replaced by `\iow_char:N\#1{...\token_to_str:N\{...}`.

LaTeX was given a regular expression where a `\iow_char:N\c\iow_char:N\{...` construction was not ended with a closing brace `\iow_char:N\}`.

A construction `\iow_char:N\c[...` appears in a regular expression, but the closing `]` is not present.

In regular expressions, the `\iow_char:N\c` escape sequence may only be followed by a left brace, a left bracket, or a capital letter representing a character category, namely one of `\'ABCDEFGHJKLMNPQRSTU`. 

A regular expression ends with `\iow_char:N\c` followed by a letter. It will be ignored.

The `\iow_char:N\u` escape sequence must be followed by a brace group with the name of the variable to use.
\msg_new:nnnn { regex } { u-missing-rbrace }
{ Missing-right-brace-inserted-for-'\iow_char:N\u'--escape. }
{ LaTeX-
{ reached-the-end-of-the-string- }
{ encountered-an-escaped-alphanumeric-character '\iow_char:N\#2'- }
when-parsing-the-argument-of-an-
'\iow_char:N\\iow_char:N\{(...)\}'--escape.
}

Errors when encountering the POSIX syntax [:...:].
\msg_new:nnnn { regex } { posix-unsupported }
{ POSIX-collating-element-'[#1 - #1]'--not-supported. }
{ The-'.foo.'-and-'[=bar=]'-syntaxes-have-a-special-meaning-
in-POSIX-regular-expressions.-This-is-not-supported-by-LaTeX.-
Maybe-you-forgot-to-escape-a-left-bracket-in-a-character-class?
}
\msg_new:nnnn { regex } { posix-unknown }
{ POSIX-class-['[:#1:]']--unknown. }
{ '[:#1:]'--is-not-among-the-known-POSIX-classes-
'[:alnum:]','-[:alpha:]','-[:ascii:]','-[:blank:]','-[:cntrl:]','-[:digit:]','-[:graph:]','-[:lower:]','-[:print:]','-[:punct:]','-[:space:]','-[:upper:]','-[:word:]','-and-'[:xdigit:].
}
\msg_new:nnnn { regex } { posix-missing-close }
{ Missing-closing-':]'--for-POSIX-class. }
{ The-POSIX-syntax-'#1'-must-be-followed-by-':]',-not-'#2'. }

In various cases, the result of a \l3regex operation can leave us with an unbalanced
token list, which we must re-balance by adding begin-group or end-group character-
tokens.
\msg_new:nnnn { regex } { result-unbalanced }
{ Missing-brace-inserted-when-#2. }
{ LaTeX-was-asked-to-do-some-regular-expression-operation,-
and-the-resulting-token-list-would-not-have-the-same-number-
of-begin-group-and-end-group-tokens.-Braces-were-inserted:-
#2-left,-#3-right.
}

Error message for unknown options.
\msg_new:nnnn { regex } { unknown-option }
{ Unknown-option-'#1'-for-regular-expressions. }
{ The-only-available-option-is-'case-insensitive',-toggled-by-
(?i)'--and-!'(?-i)'.
}
\msg_new:nnnn { regex } { special-group-unknown }
{ Unknown-special-group-'#1-...'-in-a-regular-expression. }
{ The-only-valid-constructions-starting-with-!'(?'-are-
Errors in the replacement text.

\msg_new:nnnn \{ regex \} \{ replacement-c \}
\{
  Misused-'\iow_char:N\c'\-command\-in\-a\-replacement\-text. \}
\{ In\-a\-replacement\-text,\-the\-'\iow_char:N\c'\-escape\-sequence-
  can\-be\-followed\-by\-one\-of\-the\-letters\-'ABCDELMOPSTU'-
or\-a\-brace\-group,\-not\-by\-'#1'. \}
\}
\msg_new:nnnn \{ regex \} \{ replacement-u \}
\{
  Misused-'\iow_char:N\u'\-command\-in\-a\-replacement\-text. \}
\{ In\-a\-replacement\-text,\-the\-'\iow_char:N\u'\-escape\-sequence-
  must\-be\-followed\-by\-a\-brace\-group\-holding\-the\-name\-of\-the-
  variable\-to\-use. \}
\}
\msg_new:nnnn \{ regex \} \{ replacement-g \}
\{
  Missing\-brace\-for\-the\-'\iow_char:N\g'\-construction-
in\-a\-replacement\-text. \}
\}
\msg_new:nnnn \{ regex \} \{ replacement-catcode-end \}
\{
  Missing\-character\-for\-the\-'\iow_char:N\c<category><character>'\-construction\-in\-a\-replacement\-text. \}
\}
\msg_new:nnnn \{ regex \} \{ replacement-catcode-escaped \}
\{
  Escaped\-letter\-or\-digit\-after\-category\-code\-in\-replacement\-text. \}
\}
\msg_new:nnnn \{ regex \} \{ replacement-catcode-in-cs \}
\{
  Category\-code\-'\iow_char:N\c#1#3'\-ignored\-inside-
  \-'\iow_char:N\c{...\}';\-in\-a\-replacement\-text. \}
\}
In a replacement text, the category codes of the argument of `\io\char:N \c{...\} are ignored when building the control sequence name.

\msg_new:nnnn \{ regex \} \{ replacement-null-space \}
{ TeX cannot build a space token with character code 0. }
{ You asked for a character token with category space, and character code 0, for instance through `\io\char:N \cs\io\char:N \x00'. }
This specific case is impossible and will be replaced by a normal space.

\msg_new:nnnn \{ regex \} \{ replacement-missing-rbrace \}
{ Missing right brace inserted in replacement text. }
{ There \int_compare:nTF \{ #1 \} \{ was \} \{ were \} \{ \int\-missing-right\-\int_compare:nTF \{ #1 \} \{ brace \} \{ braces \} . }

\msg_new:nnnn \{ regex \} \{ replacement-missing-rparen \}
{ Missing right parenthesis inserted in replacement text. }
{ There \int_compare:nTF \{ #1 \} \{ was \} \{ were \} \{ \int\-missing-right\-\int_compare:nTF \{ #1 \} \{ parenthesis \} \{ parentheses \} . }

\msg_new:nnnn \{ regex \} \{ submatch-too-big \}
{ Submatch \#1 \{ used\-but\-regex\-only\-has\-\#2\-group(s) \} }
Some escaped alphanumerics are not allowed everywhere.

\msg_new:nnnn \{ regex \} \{ backwards-quantifier \}
{ Quantifier "\{#1,#2\}" is backwards. }
{ The values given in a quantifier must be in order. }
Used when showing a regex.

\msg_new:nnnn \{ regex \} \{ show \}
{ \\{ \text \} \{ #3 \} }

\prop_gput:Nnn \g_msg_module_name_prop \{ regex \} \{ LaTeX3 \}
\prop_gput:Nnn \g_msg_module_type_prop \{ regex \} \{ \}

\__regex_msg_repeated:nnN
This is not technically a message, but seems related enough to go there. The arguments are: \#1 is the minimum number of repetitions; \#2 is the number of allowed extra repetitions (−1 for infinite number), and \#3 tells us about lazyness.
\cs_new:Npn \__regex_msg_repeated:nnN \#1\#2\#3
\{ \str_if_eq:eeF \{ \#1 \} \{ \#2 \} \{ 1 0 \}
\{ , \text \} \{ repeated \}
\int_case:nnF \{ \#2 \}
\{ \{ \#1\} \{ \#3 \} \{ lazy \} \{ greedy \} \}
45.9 Code for tracing

There is a more extensive implementation of tracing in the l3trial package \texttt{l3trace}. Function names are a bit different but could be merged.

\begin{verbatim}
\__regex_trace_push:nnN \__regex_trace_pop:nnN \__regex_trace:nnx

Here \#1 is the module name (\texttt{regex}) and \#2 is typically 1. If the module's current tracing level is less than \#2 show nothing, otherwise write \#3 to the terminal.

\cs_new_protected:Npn \__regex_trace_push:nnN #1#2#3 { \__regex_trace:nnx {#1} {#2} { entering \token_to_str:N #3 } }
\cs_new_protected:Npn \__regex_trace_pop:nnN #1#2#3 { \__regex_trace:nnx {#1} {#2} { leaving \token_to_str:N #3 } }
\cs_new_protected:Npn \__regex_trace:nnx #1#2#3 {
    \int_compare:nNnF { \int_use:c { g__regex_trace_#1_int } } < {#2} {
        \iow_term:x { Trace:~#3 } }
}
\end{verbatim}

(End definition for \__regex_trace_push:nnN, \__regex_trace_pop:nnN, and \__regex_trace:nnx.)

\begin{verbatim}
\__regex_trace_states:n

This function lists the contents of all states of the NFA, stored in \texttt{\toks} from 0 to \texttt{\l__-regex_max_state_int} (excluded).

\cs_new_protected:Npn \__regex_trace_states:n #1 {
    \int_step_inline:nnn \l__regex_min_state_int { \l__regex_max_state_int - 1 } {
        \__regex_trace:nnx { regex } {#1} {
            \low_char:N \toks ##1 = { \__regex_toks_use:w ##1 } }
    }
}
\end{verbatim}

(End definition for \__regex_trace_states:n.)

\begin{verbatim}
⟨/package⟩
\end{verbatim}
Chapter 46

\l3prg implementation

The following test files are used for this code: \texttt{m3prg001.lvt,m3prg002.lvt,m3prg003.lvt}.

46.1 Primitive conditionals

Those two primitive \TeX conditionals are synonyms. \texttt{\if_bool:N} is defined in \l3basics, as it’s needed earlier to define quark test functions.

\begin{verbatim}
\cs_new_eq:NN \if_predicate:w \tex_ifodd:D
\end{verbatim}

(End definition for \texttt{\if_bool:N} and \texttt{\if_predicate:w}. These functions are documented on page 68.)

46.2 Defining a set of conditional functions

These are all defined in \l3basics, as they are needed “early”. This is just a reminder!

(End definition for \texttt{\prg_set_conditional:Npnn} and others. These functions are documented on page 60.)

46.3 The boolean data type

Boolean variables have to be initiated when they are created. Other than that there is not much to say here.

\begin{verbatim}
\cs_new_protected:Npn \bool_new:N #1 { \cs_new_eq:NN #1 \c_false_bool }
\cs_generate_variant:Nn \bool_new:N { c }
\end{verbatim}

(End definition for \texttt{\bool_new:N}. This function is documented on page 63.)

A merger between \texttt{\tl_const:Nn} and \texttt{\bool_set:Nn}.

\begin{verbatim}
\cs_new_protected:Npn \bool_const:Nn \bool_const:cn #1 #2
\{ \_kernel_chk_if_free_cs:NN #1 \tex_global:D \tex_chardef:D #1 = \bool_if_p:n {#2} \}
\cs_generate_variant:Nn \bool_const:Nn { c }
\end{verbatim}

556
Setting is already pretty easy. When `check-declarations` is active, the definitions are
patched to make sure the boolean exists. This is needed because booleans are not based
on token lists nor on TeX registers.

```latex
\bool_set_true:N \bool_set_true:c
\bool_gset_true:N \bool_gset_true:c
\bool_set_false:N \bool_set_false:c
\bool_gset_false:N \bool_gset_false:c
```

The usual copy code. While it would be cleaner semantically to copy the `\cs_set_eq:NN`
family of functions, we copy `\tl_set_eq:NN` because that has the correct checking code.

```latex
\bool_set_eq:NN \bool_set_eq:cN
\bool_set_eq:Nc \bool_set_eq:cc
\bool_gset_eq:NN \bool_gset_eq:cN
\bool_gset_eq:Nc \bool_gset_eq:cc
```

This function evaluates a boolean expression and assigns the first argument the meaning
`\c_true_bool` or `\c_false_bool`. Again, we include some checking code. It is important
to evaluate the expression before applying the `\chardef` primitive, because that primitive
sets the left-hand side to `\scan_stop:` before looking for the right-hand side.

```latex
\bool_set:Nn \bool_set:cn
\bool_gset:Nn \bool_gset:cn
```

(End definition for \bool_const:Nn. This function is documented on page 63.)

(End definition for \bool_set_true:N and others. These functions are documented on page 63.)

(End definition for \bool_set_eq:NN and \bool_gset_eq:NN. These functions are documented on page 63.)
46.4 Internal auxiliaries

Internal recursion quarks.

Functions to gobble up to a quark.

Functions to query recursion quarks.

Straight forward here. We could optimize here if we wanted to as the boolean can just be input directly.

Expands to true or false with category code letter.

Show the truth value of the boolean.

Show the truth value of the boolean, as true or false.
46.5 Boolean expressions

Evaluating the truth value of a list of predicates is done using an input syntax somewhat similar to the one found in other programming languages with ( and ) for grouping, ! for logical “Not”, && for logical “And” and || for logical “Or”. However, they perform eager evaluation. We shall use the terms Not, And, Or, Open and Close for these operations.

Any expression is terminated by a Close operation. Evaluation happens from left to right in the following manner using a GetNext function:

- If an Open is seen, start evaluating a new expression using the Eval function and call GetNext again.
- If a Not is seen, remove the ! and call a GetNext function with the logic reversed.
- If none of the above, reinsert the token found (this is supposed to be a predicate function) in front of an Eval function, which evaluates it to the boolean value (true) or (false).

The Eval function then contains a post-processing operation which grabs the instruction following the predicate. This is either And, Or or Close. In each case the truth value is used to determine where to go next. The following situations can arise:
\textbf{true} And Current truth value is true, logical And seen, continue with GetNext to examine truth value of next boolean (sub-)expression.

\textbf{false} And Current truth value is false, logical And seen, stop using the values of predicates within this sub-expression until the next Close. Then return (false).

\textbf{true} Or Current truth value is true, logical Or seen, stop using the values of predicates within this sub-expression until the nearest Close. Then return (true).

\textbf{false} Or Current truth value is false, logical Or seen, continue with GetNext to examine truth value of next boolean (sub-)expression.

\textbf{true} Close Current truth value is true, Close seen, return (true).

\textbf{false} Close Current truth value is false, Close seen, return (false).

\begin{verbatim}
\prg_new_conditional:Npn \bool_if:n { T , F , TF } \{ \if_predicate:w \bool_if_p:n {#1} \prg_return_true: \else: \prg_return_false: \fi: \}
\end{verbatim}

To speed up the case of a single predicate, \texttt{f-expand} and check whether the result is one token (possibly surrounded by spaces), which must be \texttt{c_true_bool} or \texttt{c_false_bool}. We use a version of \texttt{tl_if_single:nTF} optimized for speed since we know that an empty \texttt{#1} is an error. The auxiliary \texttt{__bool_if_p_aux:w} removes the trailing parenthesis and gets rid of any space. For the general case, first issue a \texttt{\group_align_safe_begin:} as we are using && as syntax shorthand for the And operation and we need to hide it for \TeX. This group is closed after \texttt{\__bool_get_next:NN} returns \texttt{c_true_bool} or \texttt{c_false_bool}. That function requires the trailing parenthesis to know where the expression ends.

\begin{verbatim}
\cs_new:Npn \bool_if_p:n { \exp_args:Nf \__bool_if_p:n } \cs_new:Npn \__bool_if_p:n #1 \{ \tl_if_empty:oT { \use_none:nn #1 . } { \__bool_if_p_aux:w } \group_align_safe_begin: \exp_after:wN \group_align_safe_end: \exp:w \exp_end_continue_f:w \{ \__bool_get_next:NN \use_i:nnnn #1 \}
\end{verbatim}

The GetNext operation. Its first argument is \texttt{\use_i:nnnn, \use_ii:nnnn, \use_iii:nnnn, or \use_iv:nnnn} (we call these “states”). In the first state, this function eventually expand to the truth value \texttt{c_true_bool} or \texttt{c_false_bool} of the expression.
which follows until the next unmatched closing parenthesis. For instance "\_\_bool_-get_next:NN \use_i:nnnn \c_true_bool && \c_true_bool )" (including the closing parenthesis) expands to \c_true_bool. In the second state (after a !) the logic is reversed. We call these two states “normal” and the next two “skipping”. In the third state (after \c_true_bool||) it always returns \c_true_bool. In the fourth state (after \c_false_bool&&) it always returns \c_false_bool and also stops when encountering ||, not only parentheses. This code itself is a switch: if what follows is neither ! nor (, we assume it is a predicate.

\cs_new:Npn \__bool_get_next:NN #1#2
\use:c
\__bool_
\if_meaning:w !#2 ! \else: \if_meaning:w (#2 ( \else: p \fi: \fi:
\:Nw
#1 #2}
(End definition for \__bool_get_next:NN.)

\__bool_:Nw\cs_new:cpn { \__bool_!:Nw } #1#2
\exp_after:wN \__bool_get_next:NN \use_i:nnnn \use_i:nnnn \use_iii:nnnn \use_iv:nnnn
(End definition for \__bool_!:Nw.)

\__bool_:Nw\cs_new:cpn { \__bool_(:Nw } #1#2
\exp_after:wN \__bool_choose:NNN \exp_after:wN #1
\int_value:w \__bool_get_next:NN \use_i:nnnn
(End definition for \__bool_(:Nw.)

\__bool_p:Nw\cs_new:cpn { \__bool_p:Nw } #1
\exp_after:wN \__bool_choose:NNN \exp_after:wN #1 \int_value:w
(End definition for \__bool_p:Nw.)
The arguments are #1: a function such as \use_i:nnnn, #2: 0 or 1 encoding the current truth value, #3: the next operation, And, Or or Close. We distinguish three cases according to a combination of #1 and #2. Case 2 is when #1 is \use_ii:nnnn (state 3), namely after \c_true_bool ||. Case 1 is when #1 is \use_i:nnnn and #2 is true or when #1 is \use_i:nnnn and #2 is false, for instance for !\c_false_bool. Case 0 includes the same with true/false interchanged and the case where #1 is \use_ii:nnnn namely after \c_false_bool &&.

When seeing ) the current subexpression is done, leave the appropriate boolean.

In case 1, namely when the argument is true and we are in a normal state continue in the normal state 1. In case 2, namely when skipping alternatives in an Or, continue in the same state. When seeing | in case 0, continue in a normal state; in particular stop skipping for \c_false_bool && because that binds more tightly than ||. In the other two cases start skipping for \c_true_bool ||.

(End definition for \__bool_choose:NNN and others.)

Go through the list of expressions, stopping whenever an expression is false. If the end is reached without finding any false expression, then the result is true.
\bool_lazy_and_p:nn \bool_lazy_and:nnTF

Only evaluate the second expression if the first is true. Note that \#2 must be removed as an argument, not just by skipping to the \else: branch of the conditional since \#2 may contain unbalanced \TeX conditionals.

\begin{verbatim}
\prg_new_conditional:Npnn \bool_lazy_and:nn #1#2 { p , T , F , TF } 
\{ 
  \if_predicate:w 
  \bool_if:nTF {#1} { \bool_if_p:n {#2} } { \c_false_bool }
  \prg_return_true:
  \else:
  \prg_return_false:
  \fi:
\}
\end{verbatim}

(End definition for \bool_lazy_and:nnTF and \_bool_lazy_and:n. This function is documented on page 66.)

\bool_lazy_any_p:n \bool_lazy_any:nTF \__bool_lazy_any:n

Go through the list of expressions, stopping whenever an expression is true. If the end is reached without finding any true expression, then the result is false.

\begin{verbatim}
\cs_new:Npn \bool_lazy_any_p:n #1 
\{ \__bool_lazy_any:n #1 \q__bool_recursion_tail \q__bool_recursion_stop \}
\prg_new_conditional:Npnn \bool_lazy_any:n #1 { T , F , TF }
\{ 
  \if_predicate:w \bool_lazy_any_p:n {#1}
  \prg_return_true:
  \else:
  \prg_return_false:
  \fi:
\}
\cs_new:Npn \__bool_lazy_any:n #1 
\{ \__bool_if_recursion_tail_stop_do:nn {#1} { \c_false_bool }
\bool_if:nT {#1}
\{ \__bool_use_i_delimit_by_q_recursion_stop:nw { \c_true_bool } \}
\__bool_lazy_any:n
\}
\end{verbatim}

(End definition for \bool_lazy_any:nTF and \_bool_lazy_any:n. This function is documented on page 66.)

\bool_lazy_or_p:nn \bool_lazy_or:nnTF

Only evaluate the second expression if the first is false.

\begin{verbatim}
\prg_new_conditional:Npnn \bool_lazy_or:nn #1#2 { p , T , F , TF } 
\{ 
  \if_predicate:w 
  \bool_if:nTF {#1} { \c_true_bool } { \bool_if_p:n {#2} }
  \prg_return_true:
  \else:
  \prg_return_false:
  \fi:
\}
\end{verbatim}

(End definition for \bool_lazy_or:nnTF. This function is documented on page 66.)
\bool_not_p:n The Not variant just reverses the outcome of \bool_if_p:n. Can be optimized but this is nice and simple and according to the implementation plan. Not even particularly useful to have it when the infix notation is easier to use.

\bool_xor_p:nn Exclusive or. If the boolean expressions have same truth value, return false, otherwise return true.

46.6 Logical loops

A while loop where the boolean is tested before executing the statement. The “while” version executes the code as long as the boolean is true; the “until” version executes the code as long as the boolean is false.

A do-while loop where the body is performed at least once and the boolean is tested after executing the body. Otherwise identical to the above functions.

Loop functions with the test either before or after the first body expansion.

564
8401 \bool_while_do:nn {#1} {#2}
8402 \}
8403 \cs_new:Npn \bool_do_while:nn #1#2
8404 {
8405 \bool_if:nT {#1} { \bool_do_while:nn {#1} {#2} }
8406 \}
8407 \cs_new:Npn \bool_until_do:nn #1#2
8408 {
8409 \bool_if:nF {#1}
8410 { #2
8411 \bool_until_do:nn {#1} {#2}
8412 }
8413 \}
8414 \cs_new:Npn \bool_do_until:nn #1#2
8415 {
8416 #2
8417 \bool_if:nF {#1} { \bool_do_until:nn {#1} {#2} }
8418 }
8419 (End definition for \bool_while_do:nn and others. These functions are documented on page 67.)

46.7 Producing multiple copies

This function uses a cascading csname technique by David Kastrup (who else :-) The idea is to make the input 25 result in first adding five, and then 20 copies of the code to be replicated. The technique uses cascading csnames which means that we start building several csnames so we end up with a list of functions to be called in reverse order. This is important here (and other places) because it means that we can for instance make the function that inserts five copies of something to also hand down ten to the next function in line. This is exactly what happens here: in the example with 25 then the next function is the one that inserts two copies but it sees the ten copies handed down by the previous function. In order to avoid the last function to insert say, 100 copies of the original argument just to gobble them again we define separate functions to be inserted first. These functions also close the expansion of \exp:w, which ensures that \prg_replicate:nn only requires two steps of expansion.

This function has one flaw though: Since it constantly passes down ten copies of its previous argument it severely affects the main memory once you start demanding hundreds of thousands of copies. Now I don’t think this is a real limitation for any ordinary use, and if necessary, it is possible to write \prg_replicate:nn \{1000\} \{\prg_replicate:nn \{1000\} \{\langle code\rangle\}\}. An alternative approach is to create a string of m’s with \exp:w which can be done with just four macros but that method has its own problems since it can exhaust the string pool. Also, it is considerably slower than what we use here so the few extra csnames are well spent I would say.

\cs_new:Npn \prg_replicate:nn #1
\{ #2
\exp:w
Then comes all the functions that do the hard work of inserting all the copies. The first function takes \texttt{\:n} as a parameter.

Users shouldn’t ask for something to be replicated once or even not at all but...

*(End definition for \texttt{\prg_replicate:nn} and others. This function is documented on page 67.)*
46.8 Detecting \TeX{}'s mode

For testing vertical mode. Strikes me here on the bus with David, that as long as we are just talking about returning true and false states, we can just use the primitive conditionals for this and gobbling the \texttt{\exp_end} in the input stream. However this requires knowledge of the implementation so we keep things nice and clean and use the return statements.

\begin{verbatim}
467 \prg_new_conditional:Npnn \mode_if_vertical: { p , T , F , TF } \\
467 \{ \if_mode_vertical: \prg_return_true: \else: \prg_return_false: \fi: \\
\end{verbatim}

(End definition for \texttt{\mode_if_vertical:TF}. This function is documented on page 68.)

For testing horizontal mode.

\begin{verbatim}
467 \prg_new_conditional:Npnn \mode_if_horizontal: { p , T , F , TF } \\
467 \{ \if_mode_horizontal: \prg_return_true: \else: \prg_return_false: \fi: \\
\end{verbatim}

(End definition for \texttt{\mode_if_horizontal:TF}. This function is documented on page 68.)

For testing inner mode.

\begin{verbatim}
467 \prg_new_conditional:Npnn \mode_if_inner: { p , T , F , TF } \\
467 \{ \if_mode_inner: \prg_return_true: \else: \prg_return_false: \fi: \\
\end{verbatim}

(End definition for \texttt{\mode_if_inner:TF}. This function is documented on page 68.)

For testing math mode. At the beginning of an alignment cell, this should be used only inside a non-expandable function.

\begin{verbatim}
467 \prg_new_conditional:Npnn \mode_if_math: { p , T , F , TF } \\
467 \{ \if_mode_math: \prg_return_true: \else: \prg_return_false: \fi: \\
\end{verbatim}

(End definition for \texttt{\mode_if_math:TF}. This function is documented on page 68.)

46.9 Internal programming functions

\TeX{}'s alignment structures present many problems. As Knuth says himself in \textit{\TeX{}: The Program}: “It’s sort of a miracle whenever \texttt{\halign} or \texttt{\valign} work, […]” One problem relates to commands that internally issue a \texttt{\cr} but also peek ahead for the next character for use in, say, an optional argument. If the next token happens to be a \& with category code 4 we get some sort of weird error message because the underlying \texttt{\futurelet} stores the token at the end of the alignment template. This could be a \& giving a message like ! \texttt{Misplaced \cr.} or even worse: it could be the \texttt{\endtemplate} token causing even more trouble! To solve this we have to open a special group so that \TeX{} still thinks it’s on safe ground but at the same time we don’t want to introduce any brace group that may find its way to the output. The following functions help with this by using behaviour documented only in Appendix D of \textit{The \TeX{}book}… In short evaluating ‘\{’ and ‘\}’ as numbers will not change the counter \TeX{} uses to keep track of its state in an alignment, whereas gobbling a brace using \texttt{\if_false:} will affect \TeX{}’s state without producing any real group. We place the \texttt{\if_false: \{ \fi:} part at that place so that the successive expansions of \texttt{\group_align_safe_begin/end:} are always brace balanced.

\begin{verbatim}
467 \group_align_safe_begin: \\
467 \group_begin: \\
467 \tex_catcode:D '\^-0 = 2 \exp_stop_f: \\
467 \cs_new:Npn \group_align_safe_begin: \\
467 \{ \exp:w \if_false: \{ \fi: '^-0 \exp_stop_f: \} \\
\end{verbatim}
\group_end:
\cs_new:Npn \group_align_safe_end:
  { \if_int_compare:w '{ = \c_zero_int } \fi: }

("End definition for \group_align_safe_begin: and \group_align_safe_end:. These functions are documented on page 69.")

\g__kernel_prg_map_int 
\int_new:N \g__kernel_prg_map_int

("End definition for \g__kernel_prg_map_int.")

\prg_break_point:Nn \prg_map_break:Nn

These are defined in l3basics, as they are needed “early”. This is just a reminder that is the case!

("End definition for \prg_break_point:Nn and \prg_map_break:Nn. These functions are documented on page 68.")

\prg_break_point: \prg_break: \prg_break:n

Also done in l3basics.

("End definition for \prg_break_point:, \prg_break:, and \prg_break:n. These functions are documented on page 69.")

568
Chapter 47

\texttt{l3sys} implementation

\section*{Kernel code}

\subsection*{Detecting the engine}

\begin{verbatim}
\__sys_const:nn \texttt{__sys_const:nn} \texttt{\_sys_const:nn}
\end{verbatim}

Set the T, F, TF, p forms of \texttt{\_sys_const:nn} to be constants equal to the result of evaluating the boolean expression \texttt{\_sys_const:nn}.

\begin{verbatim}
cs_new_protected:Npn \_sys_const:nn #1#2
{
  \bool_if:nTF {#2}
  {\cs_new_eq:cN { #1 :T } \use:n \cs_new_eq:cN { #1 :F } \use_none:n \cs_new_eq:cN { #1 :TF } \use_i:nn \cs_new_eq:cN { #1 _p: } \c_true_bool }
  {
    \cs_new_eq:cN { #1 :T } \use_none:n \cs_new_eq:cN { #1 :F } \use:n \cs_new_eq:cN { #1 :TF } \use_ii:nn \cs_new_eq:cN { #1 _p: } \c_false_bool }
}
\end{verbatim}

\texttt{\_sys_const:nn}

(End definition for \texttt{\_sys_const:nn}.)

\begin{verbatim}
\sys_if_engine_luatex_p: \sys_if_engine_luatex:TF \sys_if_engine_pdftex_p: \sys_if_engine_pdftex:TF \sys_if_engine_ptex_p: \sys_if_engine_ptex:TF \sys_if_engine_uptex_p: \sys_if_engine_uptex:TF \sys_if_engine_xetex_p: \sys_if_engine_xetex:TF \c_sys_engine_str
\end{verbatim}

Set up the engine tests on the basis exactly one test should be true. Mainly a case of looking for the appropriate marker primitive.

\begin{verbatim}
\str_const:Nx \c_sys_engine_str \cs_if_exist:NT \tex_luatexversion:D { luatex } \cs_if_exist:NT \tex_pdftexversion:D { pdftex } \cs_if_exist:NT \tex_kanjiskip:D 
\end{verbatim}
\cs_if_exist:NTF \tex_enableCJKtoken:D
{ \ uptex }
{ \ ptx \ }
\cs_if_exist:NT \tex_XeTeXversion:D { \ xetex }
\}
\tl_map_inline:nn { { \luatex } { \pdfTeX } { \ ptx } { \ uptex } { \ xetex } }
{ \___sys_const:nn { \sys_if_engine_ \#1 }
\{ \str_if_eq_p:Vn \c_sys_engine_str \#1 \} }
(End definition for \sys_if_engine_luatex:TF and others. These functions are documented on page 71.)
\c\sys_engine_exec_str
\c\sys_engine_format_str
Take the functions defined above, and set up the engine and format names. \c\sys_engine_exec_str differs from \c\sys_engine_str as it is the actual engine name, not a “filtered” version. It differs for \ptex and \uptex, which have a leading e, and for \luatex, because \LaTeX uses the LuaHBTeX engine.
\c\sys_engine_format_str is quite similar to \c\sys_engine_str, except that it differentiates \pdfTeX from \latex (which is pdfTeX in DVI mode). This differentiation, however, is reliable only if the user doesn’t change \tex_pdfoutput:D before loading this code.
\group_begin:
\cs_set_eq:NN \lua_now:e \tex_directlua:D
\str_const:Nx \c_sys_engine_exec_str
{ \sys_if_engine_pdfTeX:T \{ pdf \ }
\sys_if_engine_xetex:T \{ xe \ }
\sys_if_engine_ptex:T \{ ep \ }
\sys_if_engine_uptex:T \{ eup \ }
\sys_if_engine_luatex:T
{ \lua \lua_now:e
{ \if (pcall(require, 'luaharfbuzz')) then |
\tex.print("hb") |
\end
\}
\latex
}
\group_end:
\str_const:Nx \c_sys_engine_format_str
{ \cs_if_exist:NTF \fmtname
{ \bool_lazy_or:nnTF
{ \str_if_eq_p:Vn \fmtname \plain }
{ \str_if_eq_p:Vn \fmtname \LaTeX2e }
{ \sys_if_engine_pdfTeX:T
{ \int_compare:nNnT \{ \tex_pdfoutput:D \} = \{ 1 \} \{ pdf \ }
\sys_if_engine_xetex:T \{ xe \ }
570
47.1.2 Randomness

This candidate function is placed there because `\sys_if_rand_exist:TF` is used in `l3fp-rand`.

Currently, randomness exists under pdfTEX, LuaTEX, pTEX and upTEX.

(End definition for `\sys_if_rand_exist:TF`. This function is documented on page 305.)

47.1.3 Platform

Setting these up requires the file module (file lookup), so is actually implemented there.

(End definition for `\sys_if_platform_unix:TF`, `\sys_if_platform_windows:TF`, and `\c_sys_platform_str`. These functions are documented on page 72.)

47.1.4 Configurations

Loading the backend code is pretty simply: check that the backend is valid, then load it up.

(End definition for `\c_sys_engine_exec_str` and `\c_sys_engine_format_str`. These variables are documented on page 71.)
\cs_new_protected:Npn \__sys_load_backend_check:N \#1
  {
    \sys_if_engine_xetex:TF
      {
        \str_case:VnF \#1
          {
            \{ dvisvgm \} \{ \}
            \{ xdvipdfmx \} \{ \tl_gset:Nn \#1 \{ xetex \} \}
            \{ xetex \} \{ \}
          }
        
        \msg_error:nxx \{ sys \} \{ wrong-backend \}
        \#1 \{ xetex \}
        \tl_gset:Nn \#1 \{ xetex \}
      }
    
    \sys_if_output_pdf:TF
      {
        \str_if_eq:VnTF \#1 \{ pdfmode \}
          {
            \sys_if_engine_luatex:TF
              {
                \tl_gset:Nn \#1 \{ luatex \}
                \tl_gset:Nn \#1 \{ pdftex \}
              }
            
            \bool_lazy_or:nnF
              {
                \str_if_eq_p:Vn \#1 \{ luatex \}
                \str_if_eq_p:Vn \#1 \{ pdftex \}
              }
              {
                \msg_error:nxx \{ sys \} \{ wrong-backend \}
                \#1 \{ \sys_if_engine_luatex:TF \{ luatex \} \{ pdfmode \} \}
                \sys_if_engine_luatex:TF
                  {
                    \tl_gset:Nn \#1 \{ luatex \}
                    \tl_gset:Nn \#1 \{ pdftex \}
                  }
            }
          }
        
        \str_case:VnF \#1
          {
            \{ dvipdfmx \} \{ \}
            \{ dvips \} \{ \}
            \{ dvisvgm \} \{ \}
          }
  }
\msg_error:nnxx { sys } { wrong-backend }
#1 { dvips }
\tl_gset:Nn #1 { dvips }
}
}
}
}
}

(End definition for \sys_load_backend:, \_sys_load_backend_check:N, and \c_sys_backend_str. These functions are documented on page 73.)

\g__sys_debug_bool \g__sys_deprecation_bool
\bool_new:N \g__sys_debug_bool
\bool_new:N \g__sys_deprecation_bool

(End definition for \g__sys_debug_bool and \g__sys_deprecation_bool.)

\sys_load_debug:
\sys_load_deprecation:
Simple.

\cs_new_protected:Npn \sys_load_debug:
\cs_new_protected:Npn \sys_load_deprecation:

(End definition for \sys_load_debug: and \sys_load_deprecation:. These functions are documented on page 74.)

47.1.5 Access to the shell

\l__sys_internal_tl
\tl_new:N \l__sys_internal_tl

(End definition for \l__sys_internal_tl.)

\c__sys_marker_tl
The same idea as the marker for rescanning token lists.
\tl_const:Nx \c__sys_marker_tl { \token_to_str:N : }

(End definition for \c__sys_marker_tl.)

\sys_get_shell:nnN\sys_get_shell:nnN \_sys_get:nnN \_sys_get_do:Nw
Setting using a shell is at this level just a slightly specialised file operation, with an additional check for quotes, as these are not supported.
\cs_new_protected:Npn \sys_get_shell:nnN #1#2#3
\cs_new_protected:Npn \sys_get_shell:nnN #1 { dvips }
\tl_gset:Nn #1 { dvips }

573
\prg_new_protected_conditional:Npn \sys_get_shell:nnN #1#2#3 { T , F , TF } 
{ \sys_if_shell:TF 
{ \exp_args:No \__sys_get:nnN \tl_to_str:n \{#1\} \{#2\} #3 } \prg_return_false: 
} \cs_new_protected:Npn \__sys_get:nnN #1#2#3 
{ \tl_if_in:nnTF \{#1\} \{ " \} \msg_error:nnx \{ kernel \} \{ quote-in-shell \} \{#1\} \prg_return_false: 
} \exp_args:Nno \use:nn { \cs_new_protected:Npn \__sys_get_do:Nw #1#2 } \c__sys_marker_tl 
{ \group_end: \if_false: \fi: \int_set_eq:NN \tex_tracingnesting:D \c_zero_int \exp_args:No \tex_everyeof:D \{ \c__sys_marker_tl \} \exp_after:wN \__sys_get_do:Nw \exp_after:wN \prg_do_nothing: \exp_after:wN \prg_return_true: 
} \sys_if_engine_luatex:F { \int_const:Nn \c__sys_shell_stream_int { 18 } } 
(End definition for \sys_get_shell:nnNTF and others. These functions are documented on page 72.) \c__sys_shell_stream_int 
This is not needed for Lua\TeX{}: shell escape there isn’t done using a \TeX{} interface.
\sys_if_engine_luatex:F \{ \int_const:Nn \c__sys_shell_stream_int \{ 18 \} \} 
(End definition for \c__sys_shell_stream_int.) \sys_shell_now:n 
\__sys_shell_now:e 
Execute commands through shell escape immediately.
For Lua\TeX{}, we use a pseudo-primitive to do the actual work.
⟨/tex⟩⟨∗lua⟩
do
local os_exec = os.execute
do
local function shellescape(cmd)
local status, msg = os.execute(cmd)
if status == nil then

8713      write_nl("log","runsystem(" .. cmd .. ")
")
8714      elseif status == 0 then
8715      write_nl("log","runsystem(" .. cmd .. ")...executed
")
8716      else
8717      write_nl("log","runsystem(" .. cmd .. ")...failed " .. (msg or ")") .. "
")
8718      end
8719      end
8720      luacmd("__sys_shell_now:e", function()
8721          shellescape(scan_string())
8722      end, "global", "protected")
8723      ⟨/lua⟩
8724      ⟨∗tex⟩
8725      \sys_if_engine_luatex:TF
8726      {
8727          \cs_new_protected:Npn \sys_shell_now:n #1
8728              { \__sys_shell_now:e { \exp_not:n {#1} } }
8729      }
8730      {
8731          \cs_new_protected:Npn \sys_shell_now:n #1
8732              { \iow_now:Nn \c__sys_shell_stream_int {#1} }
8733      }
8734      \cs_generate_variant:Nn \sys_shell_now:n { x }
8735      ⟨/tex⟩
8736      ⟨/lua⟩
8737      ⟨∗tex⟩
8738      \sys_if_engine_luatex:TF
8739      {
8740          \cs_new_protected:Npn \sys_shell_now:n #1
8741              { \__sys_shell_now:e { \exp_not:n {#1} } }
8742      }
8743      {
8744          \cs_new_protected:Npn \sys_shell_now:n #1
8745              { \iow_now:Nn \c__sys_shell_stream_int {#1} }
8746      }
8747      \cs_generate_variant:Nn \sys_shell_now:n { x }
8748      ⟨/tex⟩
8749      ⟨/lua⟩
8750      ⟨∗tex⟩
8751      \sys_if_engine_luatex:TF
8752      {
8753          \cs_new_protected:Npn \sys_shell_now:n #1
8754              { \__sys_shell_now:e { \exp_not:n {#1} } }
8755      }
8756      {
8757          \cs_new_protected:Npn \sys_shell_now:n #1
8758              { \iow_now:Nn \c__sys_shell_stream_int {#1} }
8759      }
8760      \cs_generate_variant:Nn \sys_shell_now:n { x }
8761      ⟨/tex⟩
8762      (End definition for \sys_shell_now:n and \__sys_shell_now:e. This function is documented on page 73.)

\sys_shell_shipout:n Execute commands through shell escape at shipout.
\__sys_shell_shipout:e For LuaTeX, we use the same helper as above but delayed using a late_lua whatsit.
Creating a late_lua whatsit works a bit different if we are running under ConTeXt.
8763      ⟨/lua⟩
8764      ⟨∗tex⟩
8765      \sys_if_engine_luatex:TF
8766      {
8767          local new_latelua = nodes and nodes.nuts and nodes.nuts.pool and nodes.nuts.pool.latelua
8768              local whatsit_id = node.id’whatst’
8769          local latelua_sub = node.subtype’late_lua’
8770          local node_new = node.direct.new
8771          local setfield = node.direct.setwhatsitfield or node.direct.setfield
8772          return function(f)
8773              local n = node_new(whatsit_id, latelua_sub)
8774              setfield(n, ‘data’, f)
8775          return n
8776      end
8777      end)
8778      local node_write = node.direct.write
8779      ⟨/lua⟩
8780      ⟨∗tex⟩
8781      \sys_if_engine_luatex:TF
8782      {
8783          local new_latelua = nodes and nodes.nuts and nodes.nuts.pool and nodes.nuts.pool.latelua
8784              local whatsit_id = node.id’whatst’
8785          local latelua_sub = node.subtype’late_lua’
8786          local node_new = node.direct.new
8787          local setfield = node.direct.setwhatsitfield or node.direct.setfield
8788          return function(f)
8789              local n = node_new(whatsit_id, latelua_sub)
8790              setfield(n, ‘data’, f)
8791          return n
8792      end
8793      end)
8794      local node_write = node.direct.write
8795      ⟨/lua⟩
8796      ⟨∗tex⟩
8797      \sys_if_engine_luatex:TF
8798      {
8799          local new_latelua = nodes and nodes.nuts and nodes.nuts.pool and nodes.nuts.pool.latelua
8800              local whatsit_id = node.id’whatst’
8801          local latelua_sub = node.subtype’late_lua’
8802          local node_new = node.direct.new
8803          local setfield = node.direct.setwhatsitfield or node.direct.setfield
8804          return function(f)
8805              local n = node_new(whatsit_id, latelua_sub)
8806              setfield(n, ‘data’, f)
8807          return n
8808      end
8809      end)
8810      ⟨/lua⟩
8811      ⟨∗tex⟩
8812      \sys_if_engine_luatex:TF
8813      {
8799          local new_latelua = nodes and nodes.nuts and nodes.nuts.pool and nodes.nuts.pool.latelua
8800              local whatsit_id = node.id’whatst’
8801          local latelua_sub = node.subtype’late_lua’
8802          local node_new = node.direct.new
8803          local setfield = node.direct.setwhatsitfield or node.direct.setfield
8804          return function(f)
8805              local n = node_new(whatsit_id, latelua_sub)
8806              setfield(n, ‘data’, f)
8807          return n
8808      end
8809      end)
47.2 Dynamic (every job) code

\cs_new_protected:Npn \sys_shell_shipout:n #1
  \__sys_shell_shipout:e { \exp_not:n {#1} }
\cs_new_protected:Npn \sys_shell_shipout:n #1
  \iow_shipout:Nn \c__sys_shell_stream_int {#1}
\cs_generate_variant:Nn \sys_shell_shipout:n { x }

(End definition for \sys_shell_shipout:n and \__sys_shell_shipout:e. This function is documented on page 73.)

47.2.1 The name of the job

\c_sys_jobname_str

Inherited from the \LaTeX3 name for the primitive. This has to be the primitive as it’s set in \everyjob. If the user does

\pdflatex \input some-file-name

then \everyjob is inserted before \jobname is changed form texput, and thus we would have the wrong result.

\__sys_everyjob:n
\cs_set:Npn \__sys_tmp:w #1
  \str_if_eq:eeTF { \cs_meaning:N #1 } { \token_to_str:N #1 }

(End definition for \__sys_everyjob:n. This variable is documented on page 70.)

47.2.2 Time and date

\c_sys_minute_int
\c_sys_hour_int
\c_sys_day_int
\c_sys_month_int
\c_sys_year_int

Copies of the information provided by \TeX. There is a lot of defensive code in package mode: someone may have moved the primitives, and they can only be recovered if we have \texttt{primitive} and it is working correctly. For Ini\TeX of course that is all redundant but does no harm.

\__sys_everyjob:n
\group_begin:
\cs_set:Npn \__sys_tmp:w #1
  \str_if_eq:eeTF { \cs_meaning:N #1 } { \token_to_str:N #1 }

576
\begin{verbatim}
\cs_if_exist:NTF \tex_primitive:D
\{
  \bool_lazy_and:nnTF
  \{ \sys_if_engine_xetex_p: \}
  \{ \int_compare_p:nNn \use_none:n \tex_XeTeXrevision:D \}
  < { 99999 }
  \{ \tex_primitive:D #1 \}
\}
\int_const:Nn \c_sys_minute_int { \int_mod:nn \__sys_tmp:w \time \60 }
\int_const:Nn \c_sys_hour_int { \int_div_truncate:nn \__sys_tmp:w \time \60 }
\int_const:Nn \c_sys_day_int { \__sys_tmp:w \day }
\int_const:Nn \c_sys_month_int { \__sys_tmp:w \month }
\int_const:Nn \c_sys_year_int { \__sys_tmp:w \year }
\group_end:
\end{verbatim}

(End definition for \c_sys_minute_int and others. These variables are documented on page 70.)

### 47.2.3 Random numbers

\texttt{\sys_rand_seed}: Unpack the primitive. When random numbers are not available, we return zero after an error (and incidentally make sure the number of expansions needed is the same as with random numbers available).

\begin{verbatim}
\__sys_everyjob:n
{ \sys_if_rand_exist:TF
  \{ \cs_new:Npn \sys_rand_seed: \{ \tex_the:D \tex_randomseed:D \} \}
  \{ \cs_new:Npn \sys_rand_seed:
    \int_value:w \msg_expandable_error:nnn { kernel } { fp-no-random }
    \{ \sys_rand_seed: \}
    \c_zero_int
  \}
\}
\end{verbatim}

(End definition for \sys_rand_seed:. This function is documented on page 72.)

\texttt{\sys_gset_rand_seed:n} The primitive always assigns the seed globally.

\begin{verbatim}
\__sys_everyjob:n
{ \cs_new:Npn \sys_gset_rand_seed:n
  \int_value:w \c_zero_int
\}
\end{verbatim}

577
\sys_if_rand_exist:TF
{
  \cs_new_protected:Npn \sys_gset_rand_seed:n #1
  { \tex_setrandomseed:D \int_eval:n {#1} \exp_stop_f: }
}
{
  \cs_new_protected:Npn \sys_gset_rand_seed:n #1
  {
    \msg_error:nnn { kernel } { fp-no-random } { \sys_gset_rand_seed:n {#1} }
  }
}

(End definition for \sys_gset_rand_seed:n. This function is documented on page 72.)

\sys_timer:
\_sys_elapsedtime:
\sys_if_timer_exist_p:TF
\sys_if_timer_exist:TF

\tex

\lua
local gettimeofday = os.gettimeofday
local epoch = gettimeofday() - os.clock()
local write = tex.write
local tointeger = math.tointeger
luacmd('__sys_elapsedtime:', function()
  write(tointeger((gettimeofday() - epoch)*65536 // 1))
end, 'global')
\tex
\lua
\sys_if_engine_luatex:TF
{
  \cs_new:Npn \sys_timer:
  { \_sys_elapsedtime: }
}
{
  \cs_if_exist:NTF \tex_elapsedtime:D
    { \cs_new:Npn \sys_timer:
        { \int_value:w \tex_elapsedtime:D }
    }
  }
  \msg_new:nnnn { kernel } { no-elapsed-time }
  { No-clock-detected-for-\#1. }
  { The-current-engine-provides-no-way-to-access-the-system-time. }
  \cs_new:Npn \sys_timer:
  {
    \int_value:w
    \msg_expandable_error:nnn { kernel } { no-elapsed-time }
    { \sys_timer: }
    \c_zero_int
  }
}
\_sys_const:nn { sys_if_timer_exist }
{ \cs_if_exist_p:N \tex_elapsedtime:D || \cs_if_exist_p:N \_sys_elapsedtime: }

578
47.2.4 Access to the shell

Expose the engine's shell escape status to the user.

\c_sys_shell_escape_int

\sys_everyjob:n
\int_const:Nn \c_sys_shell_escape_int
\sys_if_engine_luatex:TF
\tex_directlua:D { \tex_sprint(status.shell_escape~or~os.execute()) }
\tex_shellescape:D

\sys_if_shell:TF
\sys_if_shell_unrestricted:TF
\sys_if_shell_restricted:TF
\sys_everyjob:n
\int_compare_p:nNn \c_sys_shell_escape_int > 0
\int_compare_p:nNn \c_sys_shell_unrestricted
\int_compare_p:nNn \c_sys_shell_unrestricted = 1
\int_compare_p:nNn \c_sys_shell_restricted
\int_compare_p:nNn \c_sys_shell_restricted = 2

\sys_finalise:
\sys_everyjob:n
\cs_gset_eq:NN \g_file_curr_name_str \tex_jobname:D

47.3 Last-minute code

A simple hook to finalise the system-dependent layer. This is forced by the backend loader, which is forced by the main loader, so we do not need to include that here.
47.3.1 Detecting the output

This is a simple enough concept: the two views here are complementary.

```
\cs_new_protected:Npn \__sys_finalise:n #1
{ \tl_gput_right:Nn \g__sys_finalise_tl {#1} }
\tl_new:N \g__sys_finalise_tl

(End definition for \sys_finalise:, \__sys_finalise:n, and \g__sys_finalise_tl. This function is documented on page 74.)
```

```
\sys_if_output_dvi_p: \sys_if_output_dvi:TF
\sys_if_output_pdf_p: \sys_if_output_pdf:TF
\c_sys_output_str

This is a simple enough concept: the two views here are complementary.

```
\__sys_finalise:n
{ \__kernel_tl_gset:Nx \g__sys_backend_tl
  { \sys_if_engine_xetex:TF
    { xetex }
    \sys_if_output_pdf:TF
    { \sys_if_engine_pdfx:TF
      { pdftex }
      { luatex }
    }
    { dvips }
  }
}

(End definition for \sys_if_output_dvi:TF, \sys_if_output_pdf:TF, and \c_sys_output_str. These functions are documented on page 71.)
```

47.3.2 Configurations

```
\g__sys_backend_tl
As the backend has to be checked and possibly adjusted, the approach here is to create
a variable and use that in a one-shot to set a constant.
```

```
\tl_new:N \g__sys_backend_tl
\__sys_finalise:n
{ \__kernel_tl_gset:Nx \g__sys_backend_tl
  \sys_if_engine_xetex:TF
  { xetex }
  \sys_if_output_pdf:TF
  \sys_if_engine_pdfx:TF
  { pdftex }
  { luatex }
  \sys_if_engine_dvips:TF
  { dvips }
}
```

(End definition for \g__sys_backend_tl. These functions are documented on page 71.)

580
If there is a class option set, and recognised, we pick it up: these will over-ride anything set automatically but will themselves be over-written if there is a package option.

\__sys_finalise:n
\{\cs_if_exist:NT \@classoptionslist
\{\cs_if_eq:NNF \@classoptionslist \scan_stop:
\{\clist_map_inline:Nn \@classoptionslist
\{\str_case:nnT {#1}
\{\tl_gset:Nn \g__sys_backend_tl { dvipdfmx }
\{\tl_gset:Nn \g__sys_backend_tl { dvips }
\{\tl_gset:Nn \g__sys_backend_tl { dvisvgm }
\{\tl_gset:Nn \g__sys_backend_tl { pdftex }
\{\tl_gset:Nn \g__sys_backend_tl { xetex }
\{\clist_remove_all:Nn \@unusedoptionlist {#1} \}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\)

(End definition for \g__sys_backend_tl.)

\{/tex\}
\{/package\}
Chapter 48

l3msg implementation

\l__msg_internal_tl A general scratch for the module.
\tl_new:N \l__msg_internal_tl
(End definition for \l__msg_internal_tl.)

\l__msg_name_str Used to save module info when creating messages.
\l__msg_text_str
\str_new:N \l__msg_name_str
\str_new:N \l__msg_text_str
(End definition for \l__msg_name_str and \l__msg_text_str.)

48.1 Internal auxiliaries

\s__msg_mark Internal scan marks.
\s__msg_stop
\scan_new:N \s__msg_mark
\scan_new:N \s__msg_stop
(End definition for \s__msg_mark and \s__msg_stop.)

\msg_use_none_delimit_by_s_stop:w Functions to gobble up to a scan mark.
\cs_new:Npn \msg_use_none_delimit_by_s_stop:w #1 \s__msg_stop { }
(End definition for \msg_use_none_delimit_by_s_stop:w.)

48.2 Creating messages

Messages are created and used separately, so there two parts to the code here. First, a
mechanism for creating message text. This is pretty simple, as there is not actually a lot
to do.

\c__msg_text_prefix_tl \c__msg_more_text_prefix_tl Locations for the text of messages.
\tl_const:Nn \c__msg_text_prefix_tl { msg~text~>~ }
\tl_const:Nn \c__msg_more_text_prefix_tl { msg~text~>~ }

582
\msg_if_exist_p:nn\msg_if_exist:nn\prg_new_conditional:Nnn \msg_if_exist:nn \#1\#2 \{ p , T , F , TF \}
\{ \cs_if_exist:cTF \{ \c__msg_text_prefix_tl \#1 / \#2 \}
\{ \prg_return_true: \} \{ \prg_return_false: \} \}

(End definition for \msg_if_exist:nnTF. This function is documented on page 76.)

\__msg_chk_if_free:nn\msg_new:nnn\msg_gset:nnn\msg_set:nnn\msg_new:nnnn
Testing whether the control sequence containing the message text exists or not.
\prg_new_conditional:Nnn \msg_if_exist:nn \#1\#2 \{ p , T , F , TF \}
\{ \cs_if_exist:cTF \{ \c__msg_text_prefix_tl \#1 / \#2 \}
\{ \prg_return_true: \} \{ \prg_return_false: \} \}

(End definition for \msg_if_exist:nnTF. This function is documented on page 76.)

This auxiliary is similar to \__kernel_chk_if_free_cs:N, and is used when defining
messages with \msg_new:nnnn.
\cs_new_protected:Npn \__msg_chk_if_free:nn \#1\#2
\msg_if_exist:nnT \{\#1\} \{\#2\}
\msg_error:nxx \{ \msg \} \{ already-defined \}
\{\#1\} \{\#2\}

(End definition for \__msg_chk_if_free:nn.)

Setting a message simply means saving the appropriate text into two functions. A sanity
check first.
\cs_new_protected:Npn \msg_new:nnnn \#1\#2\#3\#4
\__msg_chk_if_free:nn \#1\#2\#3\#4
\msg_gset:nnn \#1\#2\#3\#4
\cs_set:cpn \{ \c__msg_text_prefix_tl \#1 / \#2 \}
\cs_set:cpn \{ \c__msg_more_text_prefix_tl \#1 / \#2 \}

(End definition for \msg_new:nnnn and others. These functions are documented on page 76.)

583
48.3 Messages: support functions and text

Simple pieces of text for messages.
\begin{verbatim}
\tl_const:Nn \c__msg_coding_error_text_tl
\tl_const:Nn \c__msg_continue_text_tl
\tl_const:Nn \c__msg_critical_text_tl
\tl_const:Nn \c__msg_fatal_text_tl
\tl_const:Nn \c__msg_help_text_tl
\tl_const:Nn \c__msg_no_info_text_tl
\tl_const:Nn \c__msg_on_line_text_tl
\tl_const:Nn \c__msg_return_text_tl
\tl_const:Nn \c__msg_trouble_text_tl

\tl_const:Nn \c__msg_coding_error_text_tl
{ This-is-a-coding-error.
  \\ \\ }
\tl_const:Nn \c__msg_continue_text_tl
{ Type\textless return\textgreater -to-continue }
\tl_const:Nn \c__msg_critical_text_tl
{ Reading-the-current-file-'\g_file_curr_name_str'-will-stop. }
\tl_const:Nn \c__msg_fatal_text_tl
{ This-is-a-fatal-error:-LaTeX-will-abort. }
\tl_const:Nn \c__msg_help_text_tl
{ For-immediate-help-type-H\textless return\textgreater }
\tl_const:Nn \c__msg_no_info_text_tl
{ \LaTeX-does-not-know-anything-more-about-this-error,-sorry. }
\tl_const:Nn \c__msg_return_text_tl
\tl_const:Nn \c__msg_on_line_text_tl\{ on-line \}
\tl_const:Nn \c__msg_return_text_tl
{ \\ \\ Try-typing\textless return\textgreater-to-proceed. \\
  \\ \\ If-that-doesn't-work,-type-X\textless return\textgreater-to-quit. }
\tl_const:Nn \c__msg_trouble_text_tl
{ \\ \\ More-errors-will-almost-certainly-follow: \\ \\
  the-LaTeX-run-should-be-aborted. }
\end{verbatim}

(End definition for \c__msg_coding_error_text_tl and others.)

\begin{verbatim}
\msg_line_number: For writing the line number nicely. \msg_line_context: was set up earlier, so this is not new.
\cs_new:Npn \msg_line_number: { \int_use:N \tex_inputlineno:D }
\cs_gset:Npn \msg_line_context: {
  \c__msg_on_line_text_tl \c_space_tl \msg_line_number: }
\end{verbatim}

(End definition for \msg_line_number: and \msg_line_context: These functions are documented on page 77.)
48.4 Showing messages: low level mechanism

The low-level interruption macro is rather opaque, unfortunately. Depending on the availability of more information there is a choice of how to set up the further help. We feed the extra help text and the message itself to a wrapping auxiliary, in this order because we must first setup \TeX’s \texttt{errhelp} register before issuing an \texttt{errmessage}. To deal with the various cases of critical or fatal errors with and without help text, there is a bit of argument-passing to do.

First setup \TeX’s \texttt{errhelp} register with the extra help #1, then build a nice-looking error message with #2. Everything is done using \texttt{x}-type expansion as the new line markers are different for the two type of text and need to be correctly set up. The auxiliary \texttt{\_msg_interrupt_more_text:n} receives its argument as a line-wrapped string, which is thus unaffected by expansion. We ave to split the main text into two parts as only the “message” itself is wrapped with a leader: the generic help is wrapped at full width. We also have to allow for the two characters used by \texttt{errmessage} itself.
\_msg\_interrupt:n

The business end of the process starts by producing some visual separation of the message from the main part of the log. The error message needs to be printed with everything made “invisible”: \TeX’s own information involves the macro in which \errmessage is called, and the end of the argument of the \errmessage, including the closing brace. We use an active ! to call the \errmessage primitive, and end its argument with \use_\_none:n {⟨spaces⟩} which fills the output with spaces. Two trailing closing braces are turned into spaces to hide them as well. The group in which we alter the definition of the active ! is closed before producing the message: this ensures that tokens inserted by typing \texttt{I} in the command-line are inserted after the message is entirely cleaned up.

The \_kernel_iow with:N\_nn\_auxiliary, defined in l3file, expects an ⟨integer variable⟩, an integer ⟨value⟩, and some ⟨code⟩. It runs the ⟨code⟩ after ensuring that the ⟨integer variable⟩ takes the given ⟨value⟩, then restores the former value of the ⟨integer variable⟩ if needed. We use it to ensure that the newlinechar is 10, as needed for \iow\_newline: to work, and that \errorcontextlines is −1, to avoid showing irrelevant context. Note that restoring the former value of these integers requires inserting tokens after the \errmessage, which go in the way of tokens which could be inserted by the user. This is unavoidable.
\textbf{48.5 Displaying messages}

\LaTeX{} is handling error messages and so the \TeX{} ones are disabled.
A function for issuing messages: both the text and order could in principle vary. The module name may be empty for kernel messages, hence the slightly contorted code path for a space.
\__msg_class_new:nn
\group_begin:
\cs_set_protected:Npn \__msg_class_new:nn #1#2
{ \prop_new:c { l__msg_redirect_ #1 _prop } \cs_new_protected:cpn { __msg_ #1 _code:nnnnnn } ##1##2##3##4##5##6 {#2} \cs_new_protected:cpn { msg_ #1 :nnnnnn } ##1##2##3##4##5##6 {#2} \cs_new_protected:cpn { msg_ #1 :nnnnnn } ##1##2##3##4##5##6 {#2} \use:x
\group_end:

(End definition for \msg_fatal_text:n and others. These functions are documented on page 77.)

\g_msg_module_name_prop
\g_msg_module_type_prop

For storing public module information: the kernel data is set up in advance.
\prop_new:N \g_msg_module_name_prop
\prop_gput:Nnn \g_msg_module_name_prop { LaTeX } { LaTeX3 }
\prop_new:N \g_msg_module_type_prop
\prop_gput:Nnn \g_msg_module_type_prop { LaTeX } { }

(End definition for \g_msg_module_name_prop and \g_msg_module_type_prop. These variables are documented on page 76.)

\msg_module_type:n
\msg_see_documentation_text:n

Contextual footer information, with the potential to give modules an alternative name.
\cs_new:Npn \msg_module_type:n #1
{ \prop_if_in:NnTF \g_msg_module_type_prop {#1} { \prop_item:Nn \g_msg_module_type_prop {#1} } Package }

(End definition for \msg_module_type:n. This function is documented on page 76.)

\msg_module_name:n
\msg_see_documentation_text:n

Contextual footer information, with the potential to give modules an alternative name.
\cs_new:Npn \msg_module_name:n #1
{ \prop_if_in:NnTF \g_msg_module_name_prop {#1} { \prop_item:Nn \g_msg_module_name_prop {#1} } {#1} }
\cs_new:Npn \msg_see_documentation_text:n #1
{ See~the~ \msg_module_name:n {#1} ~ documentation~for~further~information.
}

(End definition for \msg_module_name:n and \msg_see_documentation_text:n. These functions are documented on page 76.)
\msg_fatal:nnnnnn For fatal errors, after the error message \TeX{} bails out. We force a bail out rather than using \texttt{end} as this means it does not matter if we are in a context where normally the run cannot end.

\msg_fatal:nnnnx \msg_fatal:nnnnn \msg_fatal:nnnn \msg_fatal:nnxn \msg_fatal:nnxn \msg_fatal:nnnx \msg_fatal:nnx \msg_fatal:nnx \msg_fatal:nnx

(End definition for \emph{\_\textbackslash msg\_class\_new:nn}.)

\msg_critical:nnnnnn \msg_critical:nnnnx \msg_critical:nnnnn \msg_critical:nnnn \msg_critical:nnxn \msg_critical:nnxn \msg_critical:nnx \msg_critical:nnx \msg_critical:nnx

Not quite so bad: just end the current file.

(End definition for \emph{\_\textbackslash msg\_class\_new:nn} and others. These functions are documented on page 78.)
\msg_error:nnnnnn
\msg_error:nnxxxx
\msg_error:nnnnn
\msg_error:nnnn
\msg_error:nnxx
\msg_error:nnn
\msg_error:nnx
\msg_error:nn

For an error, the interrupt routine is called. We check if there is a "more text" by comparing that control sequence with a permanently empty text. We have to undefine the bootstrap versions here.

\cs_undefine:N \msg_error:nnxx
\cs_undefine:N \msg_error:nnx
\cs_undefine:N \msg_error:nn
\__msg_class_new:nn { error } { 
\__msg_interrupt:NnnnN
\msg_error_text:n {#1} {#2}
{ {#3} {#4} {#5} {#6} }
\c_empty_tl
}
\tex_endinput:D

(End definition for \msg_critical:nnnnnn and others. These functions are documented on page 79.)

\msg_error:nnnnnn
\msg_error:nnxxxx
\msg_error:nnnnn
\msg_error:nnxxx
\msg_error:nnnn
\msg_error:nnxx
\msg_error:nnn
\msg_error:nnx
\msg_error:nn

\__msg_info_aux:NNnnnnnn
\msg_warning:nnnnnn
\msg_warning:nnxxxx
\msg_warning:nnnnn
\msg_warning:nnxxx
\msg_warning:nnnn
\msg_warning:nnxx
\msg_warning:nnn
\msg_warning:nnx
\msg_warning:nn

Warnings and information messages have no decoration. Warnings are printed to the terminal while information can either go to the log or both log and terminal.

\cs_protected:Npn \__msg_info_aux:NNnnnnnn #1#2#3#4#5#6#7#8
{\str_set:Nx \l__msg_text_str { #2 {#3} }
\str_set:Nx \l__msg_name_str { \msg_module_name:n {#3} }
#1 { }
\iow_wrap:nxnN
{
\l__msg_text_str : ~
\use:c { \c__msg_text_prefix_tl #3 / #4 } {#5} {#6} {#7} {#8}
}
( \l__msg_name_str )
\prg_replicate:nn
{\str_count:N \l__msg_text_str
- \str_count:N \l__msg_name_str
} { - }
\__msg_class_new:nn { warning }
\{} \#1
\{} \#1 { }
\__msg_class_new:nn { error }
\{} \#1 { }
}\msg_info:nnnnnn
\msg_info:nnxxxx
\msg_info:nnnnn
\msg_info:nnnnnn
\msg_info:nn

(End definition for \msg_error:nnnnnn and others. These functions are documented on page 79.)
\__msg_class_new:nn { note }
\{ 
\__msg_info_aux:NNnnnnnn \io_term:n \msg_info_text:n
\{#1} {#2} {#3} {#4} {#5} {#6}
\}
\__msg_class_new:nn { info }
\{ 
\__msg_info_aux:NNnnnnnn \io_log:n \msg_info_text:n
\{#1} {#2} {#3} {#4} {#5} {#6}
\}
\__msg_class_new:nn { log }
\{ 
\_msg_log:nnnnn
\_msg_log:nnxxxx
\msg_log:nnnn
\msg_log:nnxxx
\msg_log:nnnn
\msg_log:nnxx
\msg_log:nnn
\msg_log:nnx
\msg_log:nn
\_msg_term:nnnnnnn
\_msg_term:nnxxxx
\msg_term:nnnnnn
\msg_term:nnnnxx
\msg_term:nnnxx
\msg_term:nnxx
\msg_term:nn
\_msg_term:nn
\_msg_term:nnx
\_msg_term:nnxx
\_msg_term:nnxxx
\_msg_term:nnxxxx
\_msg_term:nnnxxxx
\_msg_term:nnxx
\_msg_term:nn
\_msg_class_new:nn { none } { }

\msg_show:nnnnnnn
\msg_show:nnxxxx
\msg_show:nnnnxx
\msg_show:nnnxx
\msg_show:nnxx
\msg_show:nn
\_msg_class_new:nn { show }
\{ 
\_msg_show:n
\_msg_show:w
\_msg_show_dot:w
\_msg_show:nx
\_msg_show:n
\_msg_show:nn
\_msg_show:nnn
\_msg_show:nnxx
\_msg_show:nnxxx
\_msg_show:nnxxxx
\_msg_show:nnnxxxx
\_msg_show:nnxx
\_msg_show:nx
\_msg_show:nn
\_msg_show:nnn
\_msg_show:nnxx
\_msg_show:nnxxx
\_msg_show:nnxxxx
\cs_new_protected:Npn \__msg_show:n #1
\{ 
\tl_if_in:nnTF { ^^J #1 } { ^^J > ~ }
Checking that a message class exists. We build this from \cs_if_free:cTF rather than \cs_if_exist:cTF because that avoids reading the second argument earlier than necessary.

(End definition for \msg_show:nnnnnnn and others. These functions are documented on page 81.)

End the group to eliminate \__msg_class_new:nn.

\group_end: \__msg_class_chk_exist:nT

Support variables needed for the redirection system.

(End definition for \__msg_class_t1 and \__msg_current_class_t1)

For redirection of individually-named messages

(End definition for \__msg_redirect_prop)

During redirection, split the message name into a sequence: \{/module/submodule\}, \{/module\}, and \{\}.
Classes encountered when following redirections to check for loops.

```
\__msg_class_loop_seq
\end{definition}
```

Actually using a message is a multi-step process. First, some safety checks on the message and class requested. The code and arguments are then stored to avoid passing them around. The assignment to `\__msg_use_code:` is similar to `\tl_set:Nn`. The message is eventually produced with whatever `\l__msg_class_tl` is when `\__msg_use_code:` is called. Here is also a good place to suppress tracing output if the trace package is loaded since all (non-expandable) messages go through this auxiliary.

```
\__msg_use:nnnnnnn
\__msg_use_redirect_name:n
\__msg_useHierarchy:nwwN
\__msg_use_redirect_module:n
\__msg_use_code:n
\cs_new_protected:Npn \__msg_use:nnnnnnn #1#2#3#4#5#6#7
{
\cs_if_exist_use:N \conditionally@traceoff
\msg_if_exist:nnTF {#2} {#3}
{
\__msg_class_chk_exist:nT {#1}
{
\tl_set:Nn \l__msg_current_class_tl {#1}
\cs_set_protected:Npx \__msg_use_code:
{
\exp_not:n
\use:c { __msg_ \l__msg_class_tl _code:nnnnnn }
{#2} {#3} {#4} {#5} {#6} {#7}
}
}
\__msg_use_redirect_name:n { #2 / #3 }
}
\cs_if_exist_use:N \conditionally@traceon
}
\cs_new_protected:Npn \__msg_use_code: { }
```

The first check is for a individual message redirection. If this applies then no further redirection is attempted. Otherwise, split the message name into ⟨module⟩, ⟨submodule⟩ and ⟨message⟩ (with an arbitrary number of slashes), and store {/module/submodule}, {/module} and {} into `\l__msg_hierarchy_seq`. We then map through this sequence, applying the most specific redirection.

```
\cs_new_protected:Npn \__msg_use_redirect_name:n #1
{
\prop_get:NnNTF \l__msg_redirect_prop { / #1 } \l__msg_class_tl
{ \__msg_use_code: }
{ }
\seq_clear:N \l__msg_hierarchy_seq
\__msg_use_hierarchy:nwwN { }
\l__msg_mark \__msg_use_hierarchy:nwwN
\l__msg_mark \__msg_use_none_delimit_by_s_stop:w
\l__msg_stop
\__msg_use_redirect_module:n { }
```

593
At this point, the items of \_\_msg\_hierarchy\_seq are the various levels at which we should look for a redirection. Redirections which are less specific than the argument of \_\_msg\_use\_redirect\_module:n are not attempted. This argument is empty for a class redirection, /module for a module redirection, etc. Loop through the sequence to find the most specific redirection, with module \#1. The loop is interrupted after testing for a redirection for \#1 equal to the argument \#1 (least specific redirection allowed). When a redirection is found, break the mapping, then if the redirection targets the same class, output the code with that class, and otherwise set the target as the new current class, and search for further redirections. Those redirections should be at least as specific as \#1.

\msg\_redirect\_name:nnn

Named message always use the given class even if that class is redirected further. An empty target class cancels any existing redirection for that message.
If the target class is empty, eliminate the corresponding redirection. Otherwise, add the redirection. We must then check for a loop: as an initialization, we start by storing the initial class in \l__msg_current_class_tl.

\cs_new_protected:Npn \msg_redirect_class:nn { \__msg_redirect:nnn { } }
\cs_new_protected:Npn \msg_redirect_module:nnn #1 { \__msg_redirect:nnn { / #1 } }
\cs_new_protected:Npn \__msg_redirect:nnn #1#2#3 {\__msg_class_chk_exist:nT {#2} {	l_if_empty:nTF {#3} {\prop_remove:cn { l__msg_redirect_ #2 _prop } {#1}} {\__msg_class_chk_exist:nT {#3} {\prop_put:cnn { l__msg_redirect_ #2 _prop } {#1} {#3}} \tl_set:Nn \l__msg_current_class_tl {#2} \seq_clear:N \l__msg_class_loop_seq \__msg_redirect_loop_chk:nnn {#2} {#3} {#1}}} }

Since multiple redirections can only happen with increasing specificity, a loop requires that all steps are of the same specificity. The new redirection can thus only create a loop with other redirections for the exact same module, \#1, and not submodules. After some initialization above, follow redirections with \l__msg_class_tl, and keep track in \l__msg_class_loop_seq of the various classes encountered. A redirection from a class to itself, or the absence of redirection both mean that there is no loop. A redirection to the initial class marks a loop. To break it, we must decide which redirection to cancel. The user most likely wants the newly added redirection to hold with no further redirection.

We thus remove the redirection starting from \#2, target of the new redirection. Note that no message is emitted by any of the underlying functions: otherwise we may get an infinite loop because of a message from the message system itself.

\cs_new_protected:Npn \__msg_redirect_loop_chk:nnn #1#2#3 { \seq_put_right:Nn \l__msg_class_loop_seq {#1} \prop_get:cnNT { l__msg_redirect_ #1 _prop } {#3} \tl_set:Nn \l__msg_current_class_tl {#1} \seq_clear:N \l__msg_class_loop_seq \__msg_redirect_loop_chk:nnn {#2} {#3} {#1} \tl_if_eq:NNTF \l__msg_class_tl \l__msg_current_class_tl {\prop_put:cnn { l__msg_redirect_ #2 _prop } {#3} {#2} \msg_warning:nnxxxx { msg } { redirect-loop } }
48.6 Kernel-specific functions

These are all retained purely for older \texttt{xparse} support.

\begin{verbatim}
\cs_new_protected:Npn \__kernel_msg_new:nnnn #1 { \msg_new:nnnn { LaTeX / #1 } }
\cs_new_protected:Npn \__kernel_msg_new:nnn #1 { \msg_new:nnn { LaTeX / #1 } }
\end{verbatim}

(End definition for \texttt{\__kernel_msg_new:nnnn} and \texttt{\__kernel_msg_new:nnn}.)

\begin{verbatim}
\cs_new_protected:Npn \__kernel_msg_info:nnxx #1 { \msg_info:nnxx { LaTeX / #1 } }
\cs_new_protected:Npn \__kernel_msg_warning:nnx #1 { \msg_warning:nnx { LaTeX / #1 } }
\cs_new_protected:Npn \__kernel_msg_warning:nnxx #1 { \msg_warning:nnxx { LaTeX / #1 } }
\cs_new_protected:Npn \__kernel_msg_error:nnx #1 { \msg_error:nnx { LaTeX / #1 } }
\cs_new_protected:Npn \__kernel_msg_error:nnxx #1 { \msg_error:nnxx { LaTeX / #1 } }
\cs_new_protected:Npn \__kernel_msg_error:nnxxx #1 { \msg_error:nnxxx { LaTeX / #1 } }
\end{verbatim}

(End definition for \texttt{\__kernel_msg_info:nnxx} and others.)

\begin{verbatim}
\cs_new:Npn \__kernel_msg_expandable_error:nnn #1 { \msg_expandable_error:nnn { LaTeX / #1 } }
\cs_new:Npn \__kernel_msg_expandable_error:nnf #1 { \msg_expandable_error:nnf { LaTeX / #1 } }
\cs_new:Npn \__kernel_msg_expandable_error:nnff #1 { \msg_expandable_error:nnff { LaTeX / #1 } }
\end{verbatim}

(End definition for \texttt{\__kernel_msg_expandable_error:nnn} and \texttt{\__kernel_msg_expandable_error:nnf}.)
48.7 Internal messages

Error messages needed to actually implement the message system itself.

```
\msg_new:nnnn { msg } { already-defined }
{ Message-’#2’-for-module-’#1’-already-defined. }
{ LaTeX-was-asked-to-define-a-new-message-called-’#2’\"
by-the-module-’#1’:-this-message-already-exists.
LaTeX_return_text_tl
}
\msg_new:nnnn { msg } { unknown }
{ Unknown-message-’#2’-for-module-’#1’. }
{ LaTeX-was-asked-to-display-a-message-called-’#2’\"
by-the-module-’#1’:-this-message-does-not-exist.
LaTeX_return_text_tl
}
\msg_new:nnnn { msg } { class-unknown }
{ Unknown-message-class-’#1’. }
{ LaTeX-was-asked-to-redirect-messages-to-a-class-’#1’:\"
this-was-never-defined.
LaTeX_return_text_tl
}
\msg_new:nnnn { msg } { redirect-loop }
{ Message-redirection-loop-caused-by- {#1} -\rightarrow- {#2}
\tl_if_empty:nF {#3} { -for-module- ’ \use_none:n #3 ’ } .
}
{ Adding-the-message-redirection- {#1} -\rightarrow- {#2}
\tl_if_empty:nF {#3} { -for-the-module- ’ \use_none:n #3 ’ } -
created-an-infinite-loop\\\\
LaTeX_return_text_tl
\tl_when条件:n { #4 \\\\\
}

Messages for earlier kernel modules plus a few for l3keys which cover coding errors.
```
\msg_new:nnnn { kernel } { bad-number-of-arguments }
{ Function-’#1’-cannot-be-defined-with-#2-arguments. }
{ LaTeX-was-asked-to-define-a-function-’#1’-with- 
#2-arguments.- 
TeX-allow-between-0-and-9-arguments-for-a-single-function.
LaTeX_return_text_tl
}
\msg_new:nnnn { kernel } { command-already-defined }
{ Control-sequence-’#1’-already-defined. }
{ LaTeX-was-asked-to-create-a-new-control-sequence-’#1’-
but-this-name-has-already-been-used-elsewhere. \ \ \
The-current-meaning-is:\}
```
\msg_new:nnnn { kernel } { command-not-defined }
{ Control-sequence '#1' undefined. }
{ }
\c__msg_coding_error_text_tl
LaTeX has been asked to use a control sequence - `'#1':\\
this has not been defined yet.
}
\msg_new:nnnn { kernel } { empty-search-pattern }
{ Empty-search-pattern. }
{ }
\c__msg_coding_error_text_tl
LaTeX has been asked to replace an empty pattern by `'#1': that
would lead to an infinite loop!
}
\msg_new:nnnn { kernel } { non-base-function }
{ Function - `'#1' - is not a base function }
{ }
\c__msg_coding_error_text_tl
Functions defined through \cs_new:Nn \cs_new:Nn must have
a signature consisting of only normal arguments - `N' and `n'. -
The signature - `'#2':of `'#1' - contains other arguments - `'#3'. -
To define variants use \cs_generate_variant:Nn \cs_generate_variant:NN
and to define other functions use \cs_new:Npn.
}
\msg_new:nnnn { kernel } { missing-colon }
{ Function - `'#1' - contains no `':' . }
{ }
\c__msg_coding_error_text_tl
Code-level functions must contain `':' to separate the
argument specification from the function name. This is
needed when defining conditionals or variants, or when building a
parameter text from the number of arguments of the function.
}
\msg_new:nnnn { kernel } { overflow }
{ Integers larger than 2^{30} - 1 cannot be stored in arrays. }
{ }
\tl_if_empty:nF {#2} { at-position `#2' in-the-array `'#1'. -
The largest allowed value `#4' will be used instead.
}
\msg_new:nnnn { kernel } { out-of-bounds }
{ Access-to-an-entry beyond an array's bounds. }
{ }
\tl_if_empty:nF {#2} { at-position `#2' of-the-array `'#1',
but this array has entries at positions from 1 to `#3'.
}
\msg_new:nnnn { kernel } { protected-predicate }
{ Predicate `'#1' must be expandable. }
{ }
\c__msg_coding_error_text_tl
LaTeX has been asked to define `'#1' as a protected predicate. -
Only expandable tests can have a predicate version.
\msg_new:nnn { kernel } { randint-backward-range }
{ Wrong-order-of-bounds-in-\low_char:N\int_rand:nn{#1}{#2}. }
\msg_new:nnn { kernel } { conditional-form-unknown }
{ Conditional-form-’#1’-for-function-’#2’-unknown. }
\c__msg_coding_error_text_tl
LaTeX-has-been-asked-to-define-the-conditional-form-’#1’-of-
the-function-’#2’,-but-only-’TF’,-’T’,-’F’,-and-’p’-forms-exist.
\msg_new:nnn { kernel } { variant-too-long }
{ Variant-form-’#1’-longer-than-base-signature-of-’#2’. }
\c__msg_coding_error_text_tl
LaTeX-has-been-asked-to-create-a-variant-of-the-function-’#2’-
with-a-signature-starting-with-’#1’,-but-that-is-longer-than-
the-signature-(part-after-the-colon)-of-’#2’.
\msg_new:nnn { kernel } { invalid-variant }
{ Variant-form-’#1’-invalid-for-base-form-’#2’. }
\c__msg_coding_error_text_tl
LaTeX-has-been-asked-to-create-a-variant-of-the-function-’#2’-
with-a-signature-starting-with-’#1’,-but-cannot-change-an-argument-
from-type-’#3’-to-type-’#4’.
\msg_new:nnn { kernel } { invalid-exp-args }
{ Invalid-variant-specifier-’#1’-in-’#2’. }
\c__msg_coding_error_text_tl
LaTeX-has-been-asked-to-create-an-\low_char:N\exp_args:N...-
function-with-signature-’N#2’-but-’#1’-is-not-a-valid-argument-
specifier.
\msg_new:nnn { kernel } { deprecated-variant }
{ Variant-form-’#1’-deprecated-for-base-form-’#2’.-
One-should-not-change-an-argument-from-type-’#3’-to-type-’#4’
\str_case:nnF {#3}
{ n } { :-use-a-’\token_if_eq_charcode:NNTF #4 c v V’-variant? }
{ N } { :-base-form-only-accepts-a-single-token-argument. }
{ #4 } { :-base-form-is-already-a-variant. }
} .
\msg_new:nnn { char } { active }
\msg_new:nnn { char } { invalid-catcode }
\msg_new:nnn { char } { invalid-catcode-for-char-generation. }
\msg_new:nnn { char } { null-space }
\msg_new:nnn { char } { out-of-range }
\msg_new:nnn { char } { space }

599
\begin{verbatim}
{ \TeX\-has- \texttt{\int_eval:n \{ \c_max_register_int + 1 \}} \- toks\,-\,registers:\,-\,this\,-\,only\,-\,allows\,-\,to\,-\,shuffle\,-\,up\,-\,to\,-
{ \int_use:N \c_max_register_int \,items.\,-
The\,-\,list\,-\,will\,-\,not\,-\,be\,-\,shuffled.}
}
\msg_new:nnnn { kernel } { variable-not-defined }
{ Variable~"#1"~undefined. }
{ \c__msg_coding_error_text_tl
LaTeX\,-\,has\,-\,been\,-\,asked\,-\,to\,-\,show\,-\,a\,-\,variable\,-\,"#1\,"\,-\,but\,-\,this\,-\,has\,-\,not\,-
been\,-\,defined\,-\,yet.
}
\msg_new:nnn { kernel } { bad-type }
{ Variable\,-\,"#1\,"\,-\,is\,-\,not\,-\,a\,-\,valid\,-\,"#3\." }
{ \c__msg_coding_error_text_tl
The\,-\,variable\,-\,"#1\,"\,-\,with\,-\,\texttt{\tl_if_empty:nTF \{#4\} \{meaning\} \{value\}}\\
\texttt{\iow_indent:n \{#2\}}\\
should\,-\,be\,-\,a\,-\,"#3\,"\,-\,variable,\,-\,but\,-
\texttt{\tl_if_empty:nTF \{#4\}}
{ it\,-\,is\,-\,not \texttt{\str_if_eq:nnF \{#3\} \{bool\} \{-a\,-\,short\,-\,macro\}} . }
{ it\,-\,does\,-\,not\,-\,have\,-\,the\,-\,correct\,-
\texttt{\str_if_eq:nTF \{#2\} \{#4\}}
{ category\,-\,codes. }
{ internal\,-\,structure:\\\\texttt{\iow_indent:n \{#4\}} }
}
\msg_new:nnnn { clist } { non-clist }
{ Variable\,-\,"#1\,"\,-\,is\,-\,not\,-\,a\,-\,valid\,-\,clist. }
{ \c__msg_coding_error_text_tl
The\,-\,variable\,-\,"#1\,"\,-\,with\,-\,value\\
\texttt{\iow_indent:n \{#2\}}\\
should\,-\,be\,-\,a\,-\,clist\,-\,variable,\,-\,but\,-\,it\,-\,includes\,-\,empty\,-\,or\,-\,blank\,-\,items\,-
without\,-\,braces.
}
\end{verbatim}

Some errors are only needed in package mode if debugging is enabled by one of the options \texttt{enable-debug}, \texttt{check-declarations}, \texttt{log-functions}, or on the contrary if debugging is turned off. In format mode the error is somewhat different.

\begin{verbatim}
\msg_new:nnnn { debug } { enable-debug }
{ To\,-\,use\,-\,"#1\,"\,-\,set\,-\,the\,-\,'enable-debug'\,-\,option. }
{ \c__msg_coding_error_text_tl
The\,-\,function\,-\,"#1\,"\,-\,will\,-\,be\,-\,ignored\,-\,because\,-\,it\,-\,can\,-\,only\,-\,work\,-\,if\,-
some\,-\,internal\,-\,functions\,-\,in\,-\,expl3\,-\,have\,-\,been\,-\,appropriately\,-
declared.\,-\,This\,-\,only\,-\,happens\,-\,if\,-\,one\,-\,of\,-\,the\,-\,options\,-
'enable-debug',\,-\,'check-declarations',\,-\,'log-functions'\,-\,was\,-
given\,-\,as\,-\,an\,-\,option:\,-\,see\,-\,the\,-\,main\,-\,expl3\,-\,documentation.
}
\end{verbatim}

Some errors only appear in expandable settings, hence don't need a “more-text” argument.

\begin{verbatim}
\msg_new:nn { kernel } { bad-exp-end-f }
\end{verbatim}

601
{ Misused-\exp_end_continue_f:w or:-:nw }
\msg_new:nnn { kernel } { bad-variable }
{ Erroneous-variable-#1 used! }
\msg_new:nnn { seq } { misused }
{ A-sequence-was-misused. }
\msg_new:nnn { prop } { misused }
{ A-property-list-was-misused. }
\msg_new:nnn { prg } { negative-replication }
{ Negative-argument-for-\io_char:N\prg_replicate:nn. }
\msg_new:nnn { prop } { prop-keyval }
{ Missing-‘=’-in-‘#1’-(in-‘..._keyval:Nn’) }
\msg_new:nnn { kernel } { unknown-comparison }
{ Relation-‘#1’-not-among-\',>=,<=,!=,\',>. }
\msg_new:nnn { kernel } { zero-step }
{ Zero-step-size-for-function-#1. }
\cs_if_exist:NF \tex_expanded:D
{ \msg_new:nnn { kernel } { e-type }
  { #1 - in-e-type-argument } }
} Messages used by the “show” functions.
\msg_new:nnn { clist } { show }
{ The-comma-list- \tl_if_empty:nF {#1} { #1 - }
  \tl_if_empty:nTF {#2}
  { is-empty \> . } }
{ contains-the-items-(without-outer-braces): #2 . }
\msg_new:nnn { intarray } { show }
{ The-integer-array-#1-contains-#2-items: \ #3 . }
\msg_new:nnn { prop } { show }
{ The-property-list-#1-
  \tl_if_empty:nTF {#2}
  { is-empty \> . } }
{ contains-the-pairs-(without-outer-braces): #2 . }
\msg_new:nnn { seq } { show }
{ The-sequence-#1-
  \tl_if_empty:nTF {#2}
  { is-empty \> . } }
{ contains-the-items-(without-outer-braces): #2 . }
\msg_new:nnn { kernel } { show-streams }
{ \tl_if_empty:nTF {#2} { No - } { The-following- }
  \str_case:nn {#1}
  { ior } { input - }
  { iow } { output - }
}
streams-are-
\tl_if_empty:nTF {#2} { open } { in-use: #2 . }

602
System layer messages

\msg_new:nnnn { sys } { backend-set }
{ Backend-configuration-already-set. }
{ Run-time-backend-selection-may-only-be-carried-out-once-during-a-run.-
This-second-attempt-to-set-them-will-be-ignored. }
\msg_new:nnnn { sys } { wrong-backend }
{ Backend-request-inconsistent-with-engine:-using-’#{2}’-backend. }
{ You-have-requested-backend-’#{1}’,-but-this-is-not-suitable-for-use-with-the-
active-engine.-LaTeX3-will-use-the-’#{2}’-backend-instead. }

48.8 Expandable errors

In expansion only context, we cannot use the normal means of reporting errors. Instead, we rely on a low-level \TeX error caused by expanding a macro \texttt{??} with parameter text "?" (this could be any token) which we used followed by something else (here, a space). This shows the context, which thanks to the odd-looking \texttt{\use:n} is

\begin{verbatim}
<argument> \???
mypkg Error: The error message.
\end{verbatim}

In other words, \TeX is processing the argument of \texttt{\use:n}, which is \???

\begin{verbatim}
\cs_set_protected:Npn \__msg_tmp:w #1
{ \cs_new:Npn #1 ? { } \cs_new:Npn \__msg_expandable_error:nn ##1##2
{ \exp_after:wN \exp_after:wN \__msg_use_none_delimit_by_s_stop:w
\use:n { #1 ~ ! ~ ##2 : ~ ##1 } \s__msg_stop }
}
\exp_args:Nc \__msg_tmp:w { ??? }
\end{verbatim}

The command built from the csname \texttt{\__msg_text_prefix_tl #1 / #2} takes four arguments and builds the error text, which is fed to \texttt{\__msg_expandable_error:n} with appropriate expansion: just as for usual messages the arguments are first turned to strings, then the message is fully expanded. The module name also has to be determined.
48.9 Message formatting

\begin{verbatim}
\prop_gput:Nn \g_msg_module_name_prop { kernel } { LaTeX3 }
\prop_gput:Nn \g_msg_module_type_prop { kernel } { }
\clist_map_inline:nn
{ char , clist , coffin , debug , deprecation , msg ,
  quark , prg , prop , scanmark , seq , sys }
\prop_gput:Nn \g_msg_module_name_prop {#1} { LaTeX3 }
\prop_gput:Nn \g_msg_module_type_prop {#1} { }
\end{verbatim}

(End definition for $\texttt{msg\_expandable\_error:nnnnnn}$ and others. These functions are documented on page 81.)
Chapter 49

\textbf{l3file implementation}

The following test files are used for this code: \texttt{m3file001}.

\begin{itemize}
\item \texttt{\textbackslash package}
\end{itemize}

49.1 Input operations

\begin{itemize}
\item \texttt{\textbackslash@@=ior}
\end{itemize}

49.1.1 Variables and constants

$\texttt{\l_\_ior\_internal\_tl}$ Used as a short-term scratch variable.

\begin{verbatim}
\tl_new:N \l_\_ior\_internal\_tl
\end{verbatim}

(End definition for $\l_\_ior\_internal\_tl$.)

$\texttt{\c_\_ior\_term\_ior}$ Reading from the terminal (with a prompt) is done using a positive but non-existent stream number. Unlike writing, there is no concept of reading from the log.

\begin{verbatim}
\int_const:Nn \c_\_ior\_term\_ior { 16 }
\end{verbatim}

(End definition for $\c_\_ior\_term\_ior$.)

$\texttt{\g_\_ior\_streams\_seq}$ A list of the currently-available input streams to be used as a stack.

\begin{verbatim}
\seq_new:N \g_\_ior\_streams\_seq
\end{verbatim}

(End definition for $\g_\_ior\_streams\_seq$.)

$\texttt{\l_\_ior\_stream\_tl}$ Used to recover the raw stream number from the stack.

\begin{verbatim}
\tl_new:N \l_\_ior\_stream\_tl
\end{verbatim}

(End definition for $\l_\_ior\_stream\_tl$.)

$\texttt{\g_\_ior\_streams\_prop}$ The name of the file attached to each stream is tracked in a property list. To get the correct number of reserved streams in package mode the underlying mechanism needs to be queried. For \LaTeX{} and plain \TeX{} this data is stored in $\texttt{\count16}$: with the \texttt{etex} package loaded we need to subtract 1 as the register holds the number of the next stream to use. In Con\TeXt{}, we need to look at $\texttt{\count38}$ but there is no subtraction: like the original plain \TeX/\LaTeX{} mechanism it holds the value of the last stream allocated.

\begin{verbatim}
\prop_new:N \g_\_ior\_streams\_prop
\end{verbatim}
49.1.2 Stream management

Reserving a new stream is done by defining the name as equal to using the terminal.

\ior_new:N
\ior_new:c

```
\cs_new_protected:Npm \ior_new:N \cs_new_eq:NN \cs__ior_term_ior
\cs_generate_variant:Nn \ior_new:N { c }
```

(End definition for \ior_new:N. This function is documented on page \pageref{ior-new}.)

\ior_open:Nn \ior_open:cn

The usual scratch space.

\g_tmpa_ior \g_tmpb_ior

```
\ior_new:N \g_tmpa_ior \ior_new:N \g_tmpb_ior
```

(End definition for \g_tmpa_ior and \g_tmpb_ior. These variables are documented on page \pageref{g-tmpa-ior} and \pageref{g-tmpb-ior}.)

\l__ior_file_name_tl

Data storage.

```
\tl_new:N \l__ior_file_name_tl
```

(End definition for \l__ior_file_name_tl.)

\ior_open:NnTF \ior_open:cnTF

An auxiliary searches for the file in the \TeX, \LaTeXe\ 2\epsilon and \LaTeX3 paths. Then pass the
file found to the lower-level function which deals with streams. The \full_name is empty
when the file is not found.

```
\prg_new_protected_conditional:Npnn \ior_open:Nn \ior_open:Nn \ior_open:cn { T , F , TF } { \file_get_full_name:NTF {\#2} \l__ior_file_name_tl
\cs__kernel_ior_open:No \l__ior_file_name_tl
\prg_return_true:
\prg_generate_conditional_variant:Nnn \ior_open:Nn { c } { T , F , TF }
```

(End definition for \ior_open:NnTF and \ior_open:cnTF.)
Streams are reserved using \newread before they can be managed by ior. To prevent ior from being affected by redefinitions of \newread (such as done by the third-party package morewrites), this macro is saved here under a private name. The complicated code ensures that \__ior_new:N is not \outer despite plain \TeX’s \newread being \outer. For Con\TeXt, we have to deal with the fact that \newread works like our own: it actually checks before altering definition.

```latex
\cs_new_protected:Npn \__ior_new:N #1
\{ \cs_undefine:N #1 \__ior_new_aux:N #1 \}
\cs_set_protected:Npn \__ior_new:N #1
\{ \cs_undefine:N #1 \__ior_new_aux:N #1 \}
```

The stream allocation itself uses the fact that there is a list of all of those available. Life gets more complex as it’s important to keep things in sync. That is done using a two-part approach: any streams that have already been taken up by ior but are now free are tracked, so we first try those. If that fails, ask plain \TeX or \LAT\EX\2ε for a new stream and use that number (after a bit of conversion).

```latex
\cs_new_protected:Npx \__ior_open_stream:Nn #1#2
\{ \tex_global:D \tex_chardef:D #1 = \exp_not:N \l__ior_stream_tl \scan_stop:
\prop_gput:NVn \exp_not:N \g__ior_streams_prop #1 {#2} \}
\tex_openin:D #1
\sys_if_engine_luatex:TF
\{ \exp_not:N \__kernel_file_name_quote:n {#2} \scan_stop: \}
\}
```

Here, we act defensively in case Lua\TeX is in use with an extensionless file name.

```latex
\cs_generate_variant:Nn \__ior_open_stream:Nn { No }
```

(End definition for \ior_open:NnTF. This function is documented on page 84.)
Closing a stream means getting rid of it at the \TeX level and removing from the various data structures. Unless the name passed is an invalid stream number (outside the range $[0, 15]$), it can be closed. On the other hand, it only gets added to the stack if it was not already there, to avoid duplicates building up.

\begin{verbatim}
\ior_close:N
\ior_close:c

Closing a stream means getting rid of it at the \TeX level and removing from the various data structures. Unless the name passed is an invalid stream number (outside the range $[0, 15]$), it can be closed. On the other hand, it only gets added to the stack if it was not already there, to avoid duplicates building up.

\end{verbatim}
49.1.3 Reading input

The primitive conditional \texttt{\if_eof:w}

\begin{verbatim}
\cs_new_eq:NN \if_eof:w \tex_ifeof:D
\end{verbatim}

(End definition for \texttt{\if_eof:w}. This function is documented on page 91.)

\texttt{\ior_if_eof_p:N, \ior_if_eof:NTF}

To test if some particular input stream is exhausted the following conditional is provided.

The primitive test can only deal with numbers in the range $[0, 15]$ so we catch outliers (they are exhausted).

\begin{verbatim}
\prg_new_conditional:Npnn \ior_if_eof:N #1 { p , T , F , TF } 
\begin{verbatim}
\if_int_compare:w -1 < #1 \else:\fi:
\else:\fi:
\else:\fi:
\else:\fi:
\end{verbatim}
\end{verbatim}

(End definition for \texttt{\ior_if_eof:NTF}. This function is documented on page 88.)

\texttt{\ior_get:NN, \ior_get:NN, \ior_get:NTF}

And here we read from files.

\begin{verbatim}
\cs_new_protected:Npn \ior_get:NN #1#2 \ior_get:NN #1#2 { \tl_set:Nn #2 { \q_no_value } } \cs_new_protected:Npn \__ior_get:NN #1#2 \tex_read:D #1 to #2 \prg_new_protected_conditional:Npnn \ior_get:NN #1#2 { T , F , TF } \begin{verbatim}
\ior_if_eof:NTF #1 \else:\fi:
\__ior_get:NN #1#2 \prg_return_true:
\end{verbatim}
\end{verbatim}

(End definition for \texttt{\ior_get:NN, \ior_get:NN, and \ior_get:NTF}. These functions are documented on page 85.)
Reading as strings is a more complicated wrapper, as we wish to remove the endline character and restore it afterwards.

\begin{verbatim}
\ior_str_get:NN \__ior_str_get:NN \ior_str_get:NNTF
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \ior_str_get:NN #1#2
\ior_str_get:NNF #1 #2 { \tl_set:Nn #2 { \q_no_value } }
\cs_new_protected:Npn \__ior_str_get:NN #1#2
\exp_args:Nno \use:n
\int_set:Nn \tex_endlinechar:D { -1 }
\tex_readline:D #1 to #2
\int_set:Nn \tex_endlinechar:D
\prg_new_protected_conditional:Npnn \ior_str_get:NN #1#2 { T , F , TF }
\ior_if_eof:NTF #1
{ \prg_return_false: }
{ \__ior_str_get:NN #1 #2 \prg_return_true: }
\prg_return_true:
\end{verbatim}

(End definition for \ior_str_get:NN, \__ior_str_get:NN, and \ior_str_get:NNTF. These functions are documented on page 86.)

\begin{verbatim}
c__ior_term_noprompt_iort
\end{verbatim}

\begin{verbatim}
\ior_get_term:nN \ior_str_get_term:nN \__ior_get_term:NnN
\end{verbatim}

Getting from the terminal is better with pretty-printing.

\begin{verbatim}
\cs_new_protected:Npn \ior_get_term:nN #1#2
\__ior_get_term:NnN \__ior_str_get:NN {#1} #2
\cs_new_protected:Npn \ior_str_get_term:nN #1#2
\__ior_get_term:NnN \__ior_str_get:NN {#1} #2
\cs_new_protected:Npn \__ior_get_term:NnN #1#2#3
\group_begin:
\tex_escapechar:D = -1 \scan_stop:
\tl_if_blank:nTF {#2}
{ \exp_args:NNc #1 \c__ior_term_noprompt_iort }
{ \exp_args:NNc #1 \c__ior_term_iort }
\exp_args:NNNv \group_end:
\tl_set:Nn #3 {#2}
\end{verbatim}

(End definition for \ior_get_term:nN, \ior_str_get_term:nN, and \__ior_get_term:NnN. These functions are documented on page 301.)

\begin{verbatim}
\ior_map_break:
\ior_map_break:n
\end{verbatim}

Usual map breaking functions.

\begin{verbatim}
\cs_new:Npn \ior_map_break:
\cs_new:Npn \ior_map_break:n
\end{verbatim}
Mapping over an input stream can be done on either a token or a string basis, hence the set up. Within that, there is a check to avoid reading past the end of a file, hence the two applications of `\ior_if_eof:N` and its lower-level analogue `if_eof:w`. This mapping cannot be nested with twice the same stream, as the stream has only one “current line”.

```
\ior_map_variable:NNn \ior_str_map_variable:NNn \__ior_map_variable:NNNn \__ior_map_variable_loop:NNN

Since the \TeX{} primitive (\texttt{\read} or \texttt{\readline}) assigns the tokens read in the same way as a token list assignment, we simply call the appropriate primitive. The end-of-loop is checked using the primitive conditional for speed.
```

(End definition for \ior_map_break: and \ior_map_break:n. These functions are documented on page 86.)
49.2 Output operations

There is a lot of similarity here to the input operations, at least for many of the basics. Thus quite a bit is copied from the earlier material with minor alterations.

49.2.1 Variables and constants

\l__iow internal_tl Used as a short-term scratch variable.
\tl_new:N \l__iow internal_tl
(End definition for \l__iow internal_tl.)

\c_log_iow \c_term_iow Here we allocate two output streams for writing to the transcript file only (\c_log_iow) and to both the terminal and transcript file (\c_term_iow). Recent Lua\TeX{} provide 128 write streams; we also use \c_term_iow as the first non-allowed write stream so its value depends on the engine.
\int_const:Nn \c_log_iow \{ -1 \}
\int_const:Nn \c_term_iow \{ \}
\bool_lazy_and:nnTF \sys_if_engine_luatex_p: \int_compare_p:nNn \tex_luatexversion:D > \{ 80 \} \{ 128 \} \{ 16 \}

(End definition for \c_log_iow and \c_term_iow. These variables are documented on page 91.)

\g__iow streams_seq A list of the currently-available output streams to be used as a stack.
\seq_new:N \g__iow streams_seq
(End definition for \g__iow streams_seq.)

\l__iow stream_tl Used to recover the raw stream number from the stack.
\tl_new:N \l__iow stream_tl
(End definition for \l__iow stream_tl.)

\g__iow streams_prop As for reads with the appropriate adjustment of the register numbers to check on.
\prop_new:N \g__iow streams_prop
\int_step_inline:nnn \{ 0 \}
\{ \cs_if_exist:NTF \normalend \{ \tex_count:D 39 ~ \} \{ \tex_count:D 17 ~ \}
49.2.2 Internal auxiliaries

\texttt{\_io\_use\_i\_delimit\_by\_s\_stop:nw} Functions to gobble up to a scan mark.

\texttt{\__io\_new:N} Reserving a new stream is done by defining the name as equal to writing to the terminal: odd but at least consistent.

\texttt{\__io\_new:N} As for read streams, copy \texttt{\textbackslash newwrite}, making sure that it is not \texttt{\textbackslash outer}.

\texttt{\l\_io\_file\_name\_tl} Data storage.

\texttt{\g\_tmp\_iow} \texttt{\g\_tmpb\_iow} The usual scratch space. These variables are documented on page \texttt{91}.

\texttt{\l\_io\_new:N} Data storage.
\iow_open:Nn
\iow_open:cn
\_iow_open_stream:Nn
\_iow_open_stream:NV
The same idea as for reading, but without the path and without the need to allow for a
conditional version.
\begin{verbatim}
\cs_new_protected:Npn \iow_open:Nn #1#2
\{ \__kernel_tl_set:Nx \l__iow_file_name_tl { \__kernel_file_name_sanitize:n {#2} } \iow_close:N #1 \seq_gpop:NNTF \g__iow_streams_seq \l__iow_stream_tl \{ \_iow_open_stream:NV #1 \l__iow_file_name_tl \} \{ \_iow_new:N #1 \__kernel_tl_set:Nx \l__iow_stream_tl { \int_eval:n {#1} } \_iow_open_stream:NV #1 \l__iow_file_name_tl \} \}
\cs_generate_variant:Nn \iow_open:Nn { c }
\cs_new_protected:Npn \__iow_open_stream:Nn #1#2
\{ \tex_global:D \tex_chardef:D #1 = \l__iow_stream_tl \scan_stop: \prop_gput:NVn \g__iow_streams_prop #1 {#2} \tex_immediate:D \tex_openout:D #1 \__kernel_file_name_quote:n {#2} \scan_stop: \}
\cs_generate_variant:Nn \__iow_open_stream:Nn { NV }
\end{verbatim}

\_iow_open_stream:NV
\begin{verbatim}(End definition for \iow_open:Nn and \_iow_open_stream:Nn. This function is documented on page 84.)
\iow_close:N
\iow_close:c
Closing a stream is not quite the reverse of opening one. First, the close operation is
easier than the open one, and second as the stream is actually a number we can use it
directly to show that the slot has been freed up.
\begin{verbatim}
\cs_new_protected:Npn \iow_close:N #1
\{ \int_compare:nT { - \c_log_iow < #1 < \c_term_iow } \{ \tex_immediate:D \tex_closeout:D #1 \prop_gremove:NV \g__iow_streams_prop #1 \seq_if_in:NVF \g__iow_streams_seq #1 \{ \seq_gpush:NV \g__iow_streams_seq #1 \} \cs_gset_eq:NN #1 \c_term_iow \}
\cs_generate_variant:Nn \iow_close:N { c }
\end{verbatim}

\_iow_close:N
\begin{verbatim}(End definition for \iow_close:N. This function is documented on page 84.)
\iow_show:N
\iow_log:N
\_iow_show:NN
Seek the stream in the \g__iow_streams_prop list, then show the stream as open or
closed accordingly.
\begin{verbatim}
\cs_new_protected:Npn \iow_show:N \_iow_show:NN \tl_show:n
\cs_generate_variant:Nn \iow_show:N { c }
\cs_new_protected:Npn \iow_log:N \_iow_show:NN \tl_log:n
\cs_generate_variant:Nn \iow_log:N { c }
\cs_new_protected:Npn \_iow_show:NN #1#2
\{ \cs_gset_eq:NN \c_term_iow #2 \}
\end{verbatim}

614
\_kernel\_chk\_defined:NT #2
{
  \prop\_get:NVNTF \g\_iow\_streams\_prop #2 \l\_iow\_internal\_tl
  {
    \exp\_args:Nx #1
    { \token\_to\_str:N #2 - open: - \l\_iow\_internal\_tl }
  }
  { \exp\_args:Nx #1 { \token\_to\_str:N #2 - closed } }
}

(End definition for \iow\_show:N, \iow\_log:N, and \_\_iow\_show:NN. These functions are documented on page 84.)

\iow\_show\_list:
\iow\_log\_list:
\_\_iow\_list:N
Done as for input, but with a copy of the auxiliary so the name is correct.

\cs\_new\_protected:Npn \iow\_show\_list: { \_\_iow\_list:N \msg\_show:nnxxxx }
\cs\_new\_protected:Npn \iow\_log\_list: { \_\_iow\_list:N \msg\_log:nnxxxx }
\cs\_new\_protected:Npn \_\_iow\_list:N #1
  { #1 { kernel } { show-streams }
    { iow }
    {
      \prop\_map\_function:NN \g\_iow\_streams\_prop
        \msg\_show\_item\_unbraced:nn
    }
    { } { }
  }
(End definition for \iow\_show\_list:, \iow\_log\_list:, and \_\_iow\_list:N. These functions are documented on page 85.)

49.3.1 Deferred writing

\_\_kernel\_iow\_with:Nnn
\_\_iow\_with:nNnn
If the integer #1 is equal to #2, just leave #3 in the input stream. Otherwise, pass the old value to an auxiliary, which sets the integer to the new value, runs the code, and restores the integer.

\cs\_new\_protected:Npn \_\_kernel\_iow\_with:Nnn #1#2
{

(End definition for \_\_kernel\_iow\_with:Nnn. This function is documented on page 85.)

49.3.2 Immediate writing

\_\_kernel\_show\_with:Nnn
\_\_show\_with:nNnn
With \e\-\text, available deferred writing without expansion is easy.

\cs\_new\_protected:Npn \_\_\_show\_with:Nnn #1#2
{
  \tex\_write:D #1 {#2} }
\cs\_generate\_variant:Nn \_\_\_show\_with:Nnn { c, N, cx }
(End definition for \_\_\_show\_with:Nnn. This function is documented on page 88.)

\_\_kernel\_iow\_with:Nnn
\_\_iow\_with:nNnn
If the integer #1 is equal to #2, just leave #3 in the input stream. Otherwise, pass the old value to an auxiliary, which sets the integer to the new value, runs the code, and restores the integer.

\cs\_new\_protected:Npn \_\_kernel\_iow\_with:Nnn #1#2
{

(End definition for \_\_kernel\_iow\_with:Nnn. This function is documented on page 85.)
\int_compare:nNnTF {#1} = {#2}
\{ \use:n \}
\{ \exp_args:No \_\_iow_with:nNnn \{ \int_use:N #1 \} #1 \{#2\} \}
\}
\cs_new_protected:Npn \_\_iow_with:nNnn #1#2#3#4
\{
\int_set:Nn #2 {#3}
#4
\int_set:Nn #2 {#1}
\}
\iow_now:Nn \iow_now:Nx \iow_now:cn \iow_now:cx
This routine writes the second argument onto the output stream without expansion. If this stream isn’t open, the output goes to the terminal instead. If the first argument is no output stream at all, we get an internal error. We don’t use the expansion done by \write to get the \texttt{x} variant, because it differs in subtle ways from \texttt{x}-expansion, namely, macro parameter characters would not need to be doubled. We set the \texttt{newlinechar} to 10 using \_\_kernel_iow_with:Nnn to support formats such as plain \TeX; otherwise, \texttt{iow_newline}: would not work. We do not do this for \texttt{iow_shipout:Nn} or \texttt{iow_shipout_x:Nn}, as \TeX looks at the value of the \texttt{newlinechar} at shipout time in those cases.
\cs_new_protected:Npn \iow_newline: { \^\^J }
\cs_set_protected:Npn \iow_log:x { \iow_now:Nx \c_log_iow }\cs_new_protected:Npn \iow_log:n { \iow_now:Nn \c_log_iow }\cs_set_protected:Npn \iow_term:x { \iow_now:Nx \c_term_iow }\cs_new_protected:Npn \iow_term:n { \iow_now:Nn \c_term_iow }
(End definition for \_\_kernel_iow_with:Nnn and \_\_iow_with:nNnn.)
\iow_log:n Writing to the log and the terminal directly are relatively easy.
\iow_log:x \iow_term:n \iow_term:x
(End definition for \iow_log:n and \iow_term:n. These functions are documented on page 88.)
\texttt{iow_newline}: Global variable holding the character that forces a new line when something is written to an output stream.
\cs_new:Npn \iow_newline: { \^\^J }
(End definition for \iow_newline:. This function is documented on page 89.)
\iow_char:N Function to write any escaped char to an output stream.
\cs_new_eq:NN \iow_char:N \cs_to_str:N
(End definition for \iow_char:N. This function is documented on page 89.)
49.3.4 Hard-wrapping lines to a character count

The code here implements a generic hard-wrapping function. This is used by the messaging system, but is designed such that it is available for other uses.

\l_iow_line_count_int

This is the “raw” number of characters in a line which can be written to the terminal. The standard value is the line length typically used by TeX Live and MiKTeX.

\l_iow_newline_tl

The token list inserted to produce a new line, with the (run-on text).

\l_iow_line_target_int

This stores the target line count: the full number of characters in a line, minus any part for a leader at the start of each line.

\l__iow_one_indent_tl \l__iow_one_indent_int

The \_iow_one_indent variables hold one indentation marker and its length. The \_iow_unindent:w auxiliary removes one indentation. The function \_iow_set_indent:n (that could possibly be public) sets the indentation in a consistent way. We set it to four spaces by default.

\l__iow_line_tl \l__iow_line_part_tl

These hold the current line of text and a partial line to be added to it, respectively.
\l__iow_line_break_bool Indicates whether the line was broken precisely at a chunk boundary.

\bool_new:N \l__iow_line_break_bool
(End definition for \l__iow_line_break_bool.)

\l__iow_wrap_tl
Used for the expansion step before detokenizing, and for the output from wrapping text: fully expanded and with lines which are not overly long.

\tl_new:N \l__iow_wrap_tl
(End definition for \l__iow_wrap_tl.)

\c__iow_wrap_marker_tl \c__iow_wrap_end_marker_tl \c__iow_wrap_newline_marker_tl \c__iow_wrap_allow_break_marker_tl \c__iow_wrap_indent_marker_tl \c__iow_wrap_unindent_marker_tl
Every special action of the wrapping code is starts with the same recognizable string, \c__iow_wrap_marker_tl. Upon seeing that “word”, the wrapping code reads one space-delimited argument to know what operation to perform. The setting of \escapechar here is not very important, but makes \c__iow_wrap_marker_tl look marginally nicer.

\group_begin:
\int_set:Nn \tex_escapechar:D { -1 }
\tl_const:Nx \c__iow_wrap_marker_tl { \tl_to_str:n { \^^I \^^O \^^W \^^_ \^^W \^^R \^^A \^^P } }
\group_end:
\tl_map_inline:nn { { end } { newline } { allow_break } { indent } { unindent } }
\tl_const:cx { c__iow_wrap_ #1_marker_tl }
\tl_const:cx { c__iow_wrap_end Marker_tl }
\tl_const:cx { c__iow_wrap_newline Marker_tl }
\tl_const:cx { c__iow_wrap_allow_break Marker_tl }
\tl_const:cx { c__iow_wrap_indent Marker_tl }
\tl_const:cx { c__iow_wrap_unindent Marker_tl }
(End definition for \c__iow_wrap_marker_tl and others.)

\iow_allow_break: \__iow_allow_break: \__iow_allow_break_error:
We set \iow_allow_break:n to produce an error when outside messages. Within wrapped message, it is set to \__iow_allow_break: when valid and otherwise to \__iow_allow_break_error:. The second produces an error expandably.

\cs_new_protected:Npm \iow_allow_break:
{ }
\msg_error:nnnn { kernel } { iow-indent }
\cs_new_protected:Npm \__iow_allow_break:
{ }
(End definition for \iow_allow_break:, \__iow_allow_break:, and \__iow_allow_break_error:. This function is documented on page 301.)
We set \texttt{\iow_indent:n} to produce an error when outside messages. Within wrapped message, it is set to \texttt{\__iow_indent:n} when valid and otherwise to \texttt{\__iow_indent_error:n}. The first places the instruction for increasing the indentation before its argument, and the instruction for unindenting afterwards. The second produces an error expandably. Note that there are no forced line-break, so the indentation only changes when the next line is started.

\begin{verbatim}
\cs_new_protected:Npn \iow_indent:n #1
\{ \msg_error:nnnnn { kernel } { iow-indent } { \iow_wrap:nnnN } { \iow_indent:n } {#1} \}

\cs_new:Npx \__iow_indent:n #1
\{ \c__iow_wrap_indent_marker_tl \ #1 \ \c__iow_wrap_unindent_marker_tl \}

\cs_new:Npn \__iow_indent_error:n #1
\{ \msg_expandable_error:nnnnn { kernel } { iow-indent } { \iow_wrap:nnnN } { \iow_indent:n } {#1} \}
\end{verbatim}

(End definition for \texttt{\iow_indent:n}, \texttt{\__iow_indent:n}, and \texttt{\__iow_indent_error:n}. This function is documented on page 90.)

\texttt{\iow_wrap:nnnN} \texttt{\iow_wrap:nxnN}

The main wrapping function works as follows. First give $\backslash\backslash$, $\backslash\backslash$ and other formatting commands the correct definition for messages and perform the given setup \texttt{#3}. The definition of $\backslash\backslash$ uses an “other” space rather than a normal space, because the latter might be absorbed by \TeX{} to end a number or other \texttt{f}-type expansions. Use \texttt{\conditionally@traceoff} if defined; it is introduced by the trace package and suppresses uninteresting tracing of the wrapping code.

\begin{verbatim}
\cs_new_protected:Npn \iow_wrap:nnnN #1#2#3#4
\{ \group_begin: \cs_if_exist_use:N \conditionally@traceoff \int_set:Nn \tex_escapechar:D { -1 } \cs_set:Npx \{ { \token_to_str:N \{ } \cs_set:Npx \# { \token_to_str:N \# } \cs_set:Npx \} { \token_to_str:N \} } \cs_set_eq:NN \% { \token_to_str:N \~ } \int_set:Nn \tex_escapechar:D { 92 } \cs_set_eq:NN \\ \iow_newline: \cs_set_eq:NN \ \c_catcode_other_space_tl \cs_set_eq:NN \ \iow_allow_break: \\ \iow_allow_break: \\ \iow_indent:n \ \__iow_indent:n \\ #3 \}
\end{verbatim}

Then fully-expand the input: in package mode, the expansion uses \LaTeX{}\texttt{2}$\epsilon$’s \texttt{\protect} mechanism in the same way as \texttt{\typeout}. In generic mode this setting is useless but
harmless. As soon as the expansion is done, reset \iow_indent:n to its error definition: it only works in the first argument of \iow_wrap:nnnN.

Afterwards, set the newline marker (two assignments to fully expand, then convert to a string) and initialize the target count for lines (the first line has target count \l_iow_line_count_int instead).

Sanity check.

There is then a loop over the input, which stores the wrapped result in \l__iow_wrap_tl. After the loop, the resulting text is passed on to the function which has been given as a post-processor. The \tl_to_str:N step converts the “other” spaces back to normal spaces. The f-expansion removes a leading space from \l__iow_wrap_tl.

Escape spaces and change newlines to \c__iow_wrap_newline_marker_tl. Set up a few variables, in particular the initial value of \l__iow_wrap_tl: the space stops the f-expansion of the main wrapping function and \use_none:n removes a newline marker inserted by later code. The main loop consists of repeatedly calling the chunk auxiliary to wrap chunks delimited by (newline or indentation) markers.
The chunk and next auxiliaries are defined indirectly to obtain the expansions of \c_catcode_other_space_tl and \c__iow_wrap_marker_tl in their definition. The next auxiliary calls a function corresponding to the type of marker (its #2), which can be newline or indent or unindent or end. The first argument of the chunk auxiliary is a target number of characters and the second is some string to wrap. If the chunk is empty simply call next. Otherwise, set up a call to \_iow_wrap_line:nw, including the indentation if the current line is empty, and including a trailing space (#1) before the \_iow_wrap_end_chunk:w auxiliary.

\_iow_wrap_chunk:nw  \_iow_wrap_next:nw
This is followed by \{⟨string⟩\} \{interp⟩\}. It stores the ⟨string⟩ and up to ⟨interp⟩ characters from the current chunk into \__iow_line_part_tl. Characters are grabbed 8 at a time and left in \__iow_line_part_tl by the line_loop auxiliary. When \(k < 8\) remain to be found, the line_aux auxiliary calls the line_end auxiliary followed by (the single digit) \(k\), then \(7 - k\) empty brace groups, then the chunk’s remaining characters. The line_end auxiliary leaves \(k\) characters from the chunk in the line part, then ends the assignment. Ignore the \use_none:nnnnn line for now. If the next character is a space the line can be broken there: store what we found into the result and get the next line. Otherwise some work is needed to find a break-point. So far we have ignored what happens if the chunk is shorter than the requested number of characters: this is dealt with by the end_chunk auxiliary, which gets treated like a character by the rest of the code. It ends up being called either as one of the arguments \#2–\#9 of the line_loop auxiliary or as one of the arguments \#2–\#8 of the line_end auxiliary. In both cases stop the assignment and work out how many characters are still needed. Notice that when we have exactly seven arguments to clean up, a \exp_stop_f: has to be inserted to stop the \exp:w. The weird \use_none:nnnnn ensures that the required data is in the right place.
Functions here are defined indirectly: \_\_iow_wrap_line:nw is eventually called with an “other” space as its argument. The goal is to remove from \l__iow_line_part_tl the part after the last space. In most cases this is done by repeatedly calling the break_loop auxiliary, which leaves “words” (delimited by spaces) until it hits the trailing space: then its argument ##3 is ? \_\_iow_wrap_break_end:w instead of a single token, and that break_end auxiliary leaves in the assignment the line until the last space, then calls \_\_iow_wrap_line_end:nw to finish up the line and move on to the next. If there is no space in \_\_iow_line_part_tl then the break_first auxiliary calls the break-none auxiliary. In that case, if the current line is empty, the complete word (including ##4, characters beyond what we had grabbed) is added to the line, making it over-long. Otherwise, the word is used for the following line (and the last space of the line so far is removed because it was inserted due to the presence of a marker).
The special case where the end of a line coincides with the end of a chunk is detected here, to avoid a spurious empty line. Otherwise, call \_\_iow_wrap_line:nw to find characters for the next line (remembering to account for the indentation).

\_\_iow_wrap_allow_break:n
This is called after a chunk has been wrapped. The \_\_iow_line_part_tl typically ends with a space (except at the beginning of a line?), which we remove since the allow_break marker should not insert a space. Then move on with the next chunk, making sure to adjust the target number of characters for the line in case we did remove a space.
These functions are called after a chunk has been wrapped, when encountering \texttt{indent/unindent} markers. Add the line part (last line part of the previous chunk) to the line so far and reset a boolean denoting the presence of a line-break. Most importantly, add or remove one indent from the current indent (both the integer and the token list). Finally, continue wrapping.

\begin{verbatim}
\cs_new_protected:Npn \__iow_wrap_indent:n #1
\tl_put_right:Nx \l__iow_line_tl { \l__iow_line_part_tl }
\bool_set_false:N \l__iow_line_break_bool
\int_add:Nn \l__iow_indent_int { \l__iow_one_indent_int }
\tl_put_right:No \l__iow_indent_tl { \l__iow_one_indent_tl }
\__iow_wrap_chunk:nw {#1}
\end{verbatim}

(End definition for \texttt{\__iow_wrap_indent:n} and \texttt{\__iow_wrap_unindent:n}.)

These functions are called after a chunk has been line-wrapped, when encountering a \texttt{newline/end} marker. Unless we just took a line-break, store the line part and the line so far into the whole \texttt{\l__iow_wrap_tl}, trimming a trailing space. In the \texttt{newline} case look for a new line (of length \texttt{\l__iow_line_target_int}) in a new chunk.

\begin{verbatim}
\cs_new_protected:Npn \__iow_wrap_newline:n #1
\bool_if:NF \l__iow_line_break_bool
{ \__iow_wrap_store_do:n { \__iow_wrap_trim:N } }
\bool_set_false:N \l__iow_line_break_bool
\__iow_wrap_chunk:nw { \l__iow_line_target_int }
\end{verbatim}

(End definition for \texttt{\__iow_wrap_newline:n} and \texttt{\__iow_wrap_end:n}.)

First add the last line part to the line, then append it to \texttt{\l__iow_wrap_tl} with the appropriate new line (with “run-on” text), possibly with its last space removed (#1 is empty or \texttt{\__iow_wrap_trim:N}).
\cs_new_protected:Npn \__iow_wrap_store_do:n #1
  {
    \__kernel_tl_set:Nx \l__iow_line_tl
    { \l__iow_line_tl \l__iow_line_part_tl }
    \__kernel_tl_set:Nx \l__iow_wrap_tl
    { \l__iow_wrap_tl \l__iow_newline_tl
      \exp_after:wN \__iow_wrap_trim:w \s__iow_mark #1 \l__iow_line_tl
    }
    \tl_clear:N \l__iow_line_tl
  }
\enddef \__iow_wrap_store_do:n

\__iow_wrap_trim:N
\__iow_wrap_trim:w
\__iow_wrap_trim_aux:w
\cs_set_protected:Npn \__iow_tmp:w #1
  {
    \cs_new:Npn \__iow_wrap_trim:N ##1
      { \exp_after:wN \__iow_wrap_trim:w ##1 \s__iow_mark \s__iow_stop }
    \cs_new:Npn \__iow_wrap_trim:w ##1 #1 \s__iow_mark
      { \__iow_wrap_trim_aux:w ##1 \s__iow_mark }
    \cs_new:Npn \__iow_wrap_trim_aux:w ##1 \s__iow_mark ##2 \s__iow_stop
    { ##1 }
  }
\exp_args:NV \__iow_tmp:w \c_catcode_other_space_tl
\enddef \__iow_wrap_trim:N, \__iow_wrap_trim:w, and \__iow_wrap_trim_aux:w

49.4 File operations

\l__file_internal_tl
\enddef \l__file_internal_tl

\g__file_stack_seq
\enddef \g__file_stack_seq

\g_file_curr_dir_str
\g_file_curr_ext_str
\g_file_curr_name_str
\enddef \g_file_curr_dir_str, \g_file_curr_ext_str, and \g_file_curr_name_str. These
variables are documented on page 91.)

The input list of files is stored as a sequence stack. In package mode we can recover
the information from the details held by \LaTeX\ 2ε (we must be in the preamble and loaded
using \usepackage or \ RequirePackage). As \LaTeX\ 2ε doesn’t store directory and name
separately, we stick to the same convention here. In pre-loading, \@currnamestack
is empty so is skipped.
\seq_new:N \g__file_stack_seq
\group_begin:
\cs_set_protected:Npn \__file_tmp:w #1#2#3
\g__file_stack_seq
\g__file_record_seq
\l__file_base_name_tl
\l__file_full_name_tl
\l__file_dir_str
\l__file_ext_str
\l__file_name_str
\l_file_search_path_seq
\l__file_tmp_seq

The total list of files used is recorded separately from the current file stack, as nothing is ever popped from this list. The current file name should be included in the file list! We will eventually copy the contents of \@filelist.

For storing the basename and full path whilst passing data internally.

Used in parsing a path into parts: in contrast to the above, these are never used outside of the current module.

The current search path.

Scratch space for comma list conversion.

(End definition for \g__file_stack_seq.)

(End definition for \g__file_record_seq.)

(End definition for \l__file_base_name_tl and \l__file_full_name_tl.)

(End definition for \l__file_dir_str, \l__file_ext_str, and \l__file_name_str.)

(End definition for \l_file_search_path_seq. This variable is documented on page 92.)

(End definition for \l__file_tmp_seq.)
49.4.1 Internal auxiliaries

\s__file_stop
Internal scan marks.
\scan_new:N \s__file_stop
(End definition for \s__file_stop.)

\q__file_nil
Internal quarks.
\quark_new:N \q__file_nil
(End definition for \q__file_nil.)

\__file_quark_if_nil_p:n
\__file_quark_if_nil:n
Branching quark conditional.
\__kernel_quark_new_conditional:Nn \__file_quark_if_nil:n { TF }
(End definition for \__file_quark_if_nil:nTF.)

\q__file_recursion_tail
\q__file_recursion_stop
Internal recursion quarks.
\quark_new:N \q__file_recursion_tail
\quark_new:N \q__file_recursion_stop
(End definition for \q__file_recursion_tail and \q__file_recursion_stop.)

\__file_if_recursion_tail_break:NN
\__file_if_recursion_tail_stop_do:Nn
Functions to query recursion quarks.
\__kernel_quark_new_test:N \__file_if_recursion_tail_break:N
\__kernel_quark_new_test:N \__file_if_recursion_tail_stop_-
do:Nn
(End definition for \__file_if_recursion_tail_break:NN and \__file_if_recursion_tail_stop-
do:Nn.)

\__kernel_file_name_sanitize:n
\__file_name Expand:n
\__file_name_expand_cleanup:Nw
\__file_name_expand_cleanup:w
\__file_name Expand_end:
\__file_name Expand_error:Nw
\__file_name Strip_quotes:n
\__file_name Strip_quotes:nnn
\__file_name Strip_spaces:n
\__file_name Strip_spaces:nnn
\__file_name Trim_spaces:n
\__file_name Trim_spaces:nnn
\__file_name Trim_spaces_aux:n
\__file_name Trim_spaces_aux:nnn
Expanding the file name uses a \csname-based approach, and relies on active characters
(for example from UTF-8 characters) being properly set up to expand to a expansion-
safe version using \ifcsname. This is less conservative than the token-by-token approach
used before, but it is much faster.
\cs_new:Npn \__kernel_file_name_sanitize:n #1
{ \exp_args:Ne \__file_name_trim_spaces:n
\exp_args:Ne \__file_name_strip_quotes:n
\__file_name_expand:n {#1} }
\cs_new:Npn \__file_name_expand:n #1
{ \exp_after:wN \__file_name_expand_cleanup:Nw
\cs:w __file_name = #1 \cs_end:
\__file_name_expand_end:
}

We’ll use \cs:w to start expanding the file name, and to avoid creating csnames
equal to \relax with “common” names, there’s a prefix \__file_name= to the csnames.
There’s also a guard token at the end so we can check if there was an error during the
process and (try to) clean up gracefully.
\cs_new:Npn \__file_name_expand:n #1
{ \exp_after:wN \__file_name_expand_cleanup:Nw
\cs:w __file_name = #1 \cs_end:
\__file_name_expand_end:

With the csname built, we grab it, and grab the remaining tokens delimited by \_\_file_name_expand_end:. If there are any remaining tokens, something bad happened, so we'll call the error procedure \_\_file_name_expand_error:Nw. If everything went according to plan, then use \token_to_str:N on the csname built, and call \_\_file_name_expand_cleanup:w to remove the prefix we added a while back. \_\_file_name_expand_cleanup:w takes a leading argument so we don't have to bother about the value of \tex_escapechar:D.

\begin{verbatim}
\cs_new:Npn \_\_file_name_expand_cleanup:Nw #1 #2 \_\_file_name_expand_end: 
\tl_if_empty:nF {#2} \_\_file_name_expand_error:Nw #2 \_\_file_name_expand_end: \exp_after:wN \_\_file_name_expand_cleanup:w \token_to_str:N #1 \exp_last_unbraced:NNNNo
\cs_new:Npn \_\_file_name_expand_cleanup:w #1 \tl_to_str:n { __file_name = } { }
\end{verbatim}

In non-error cases \_\_file_name_expand_end: should not expand. It will only do so in case there is a \csname too much in the file name, so it will throw an error (while expanding), then insert the missing \cs_end: and yet another \_\_file_name_expand_end: that will be used as a delimiter by \_\_file_name_expand_cleanup:Nw (or that will expand again if yet another \endcsname is missing).

\begin{verbatim}
\cs_new:Npn \_\_file_name_expand_end: 
\msg_expandable_error:nn \cs_end: \_\_file_name_expand_end:
\end{verbatim}

Now to the error case. \_\_file_name_expand_error:Nw adds an extra \cs_end: so that in case there was an extra \csname in the file name, then \_\_file_name_expand_error_aux:Nw throws the error.

\begin{verbatim}
\cs_new:Npn \_\_file_name_expand_error:Nw #1 #2 \_\_file_name_expand_end: 
\exp_after:wN \_\_file_name_expand_error_aux:Nw #1 #2 \_\_file_name_expand_end: 
\cs_new:Npn \_\_file_name_expand_error_aux:Nw #1 #2 \_\_file_name_expand_end: 
\msg_expandable_error:nnff \token_to_str:N #1 \exp_stop_f: #2 
\end{verbatim}

Quoting file name uses basically the same approach as for \luaquotejobname: count the " tokens and remove them.

\begin{verbatim}
\cs_new:Npn \_\_file_name_strip_quotes:n #1 \_\_file_name_strip_quotes:nw { 0 } \_\_file_recursion_tail " \_\_file_recursion_stop {#1}
\cs_new:Npn \_\_file_name_strip_quotes:nw #1#2 " \_\_file_name_strip_quotes_end:wnwn \if_meaning:w \q__file_recursion_tail #2 \_\_file_name_strip_quotes_end: #2
\end{verbatim}

629
Spaces need to be trimmed from the start of the name and from the end of any extension. However, the name we are passed might not have an extension: that means we have to look for one. If there is no extension, we still use the standard trimming function but deliberately prevent any spaces being removed at the end.

(End definition for \texttt{\_\_kernel_file_name_sanitize:n} and others.)

The same idea as the marker for rescanning token lists: this pair of tokens cannot appear in a file that is being input.

(End definition for \texttt{\_\_kernel_file_name_quote:n} and \texttt{\_\_file_name_quote:nw}.)

The approach here is similar to that for \texttt{\textbackslash tl_set_rescan:nN}. The file contents are grabbed as an argument delimited by \texttt{\_\_file_marker_tl}. A few subtleties: braces in \texttt{\_\_file_get:nN} to deal with possible alignment tabs, \texttt{\_\_file_recursion_stop} to avoid
a warning about a group being closed inside the `\scantokens`, and `\prg_return_true:` is placed after the end-of-file marker.

\cs_new_protected:Npn \file_get:nnN \#1\#2\#3
\{ 
  \file_get:nnNF \#1 \{\#2\} \#3 
  { \tl_set:Nn \#3 \{ \q_no_value \} } 
\}

\prg_new_protected_conditional:Npnn \file_get:nnN \#1\#2\#3 { T , F , TF }
\{ 
  \file_get_full_name:nNTF \#1 \l__file_full_name_tl 
  \exp_args:NV \__file_get_aux:nnN \l__file_full_name_tl \#2 \#3 
  \prg_return_true: 
\}
{ \prg_return_false: }

\cs_new_protected:Npx \__file_get_aux:nnN \#1\#2\#3
\{ 
  \exp_not:N \if_false: { \exp_not:N \fi: 
    \group_begin: 
      \int_set_eq:NN \tex_tracingnesting:D \c_zero_int 
      \exp_not:N \exp_args:No \tex_everyeof:D 
      \{ \exp_not:N \c__file_marker_tl \} 
      #2 \scan_stop: 
      \exp_not:N \exp_after:wN \exp_not:N \__file_get_do:Nw \l__file_full_name_tl \#2 
      \exp_not:N \exp_after:wN \exp_not:N \prg_do_nothing: 
      \sys_if_engine_luatex:TF 
      { \{\#1\} } 
      { \exp_not:N \__kernel_file_name_quote:n \{\#1\} \scan_stop: } 
    \exp_not:N \if_false: \exp_not:N \fi: 
  \}
\exp_args:Nno \use:nn 
  \{ \cs_new_protected:Npn \__file_get_do:Nw \#1\#2 \} 
  \{ \c__file_marker_tl \} 
  \group_end: 
  \tl_set:No \#1 \{\#2\} 
\}

(End definition for `\file_get:nnNTF` and others. These functions are documented on page 92.)

`\__file_size:n` A copy of the primitive where it’s available.

\cs_new_eq:NN \__file_size:n \tex_filesize:D

(End definition for `\__file_size:n`) File searching can be carried out if the `\pdffilesize` primitive or an equivalent is available. That of course means we need to arrange for everything else to here to be done by expansion too. We start off by sanitizing the name and quoting if required: we may need to remove those quotes, so the raw name is passed too.
First, we check of the file is just here: no mapping so we do not need the break part of the broader auxiliary. We are using the fact that the primitive here returns nothing if the file is entirely absent. To avoid unnecessary filesystem lookups, the result of `\pdffilesize` is kept available as an argument. For package mode, `\input@path` is a token list not a sequence.

Two pars to the auxiliary here so we can avoid doing quoting twice in the event we find the right file.

As \TeX automatically adds `.tex` if there is no extension, there is a little clean up to do here. First, make sure we are not in the directory part, saving that. Then check for an extension.
\cs_new:Npn \__file_ext_check:nn #1 #2
{ \__file_ext_check:nnw { #2 } { / } #1 / \q__file_nil / \s__file_stop }
\cs_new:Npn \__file_ext_check:nnw #1 #2 #3 / #4 / #5 \s__file_stop
{ \__file_quark_if_nil:nTF { #4 }
  { \exp_args:No \__file_ext_check:nnnw
    { \use_none:n #2 } { #1 } { #3 } #3 . \q__file_nil . \s__file_stop }
  { \__file_ext_check:nnw { #1 } { #2 #3 / } #4 / #5 \s__file_stop }
}
\cs_new:Npx \__file_ext_check:nnnw #1 #2 #3 #4 . #5 . #6 \s__file_stop
{ \exp_not:N \__file_quark_if_nil:nTF { #5 }
  { \exp_not:N \__file_ext_check:nnn
    { #1 #3 \tl_to_str:n { .tex } } { #1 #3 } { #2 }
  } #1 #3
}
\cs_new:Npn \__file_ext_check:nnn #1
{ \exp_args:Nne \__file_ext_check:nnnn { #1 } { \__file_size:n { #1 } } }
\cs_new:Npn \__file_ext_check:nnnn #1 #2 #3 #4
{ \tl_if_blank:nTF { #2 }
  {#3}
  { \int_compare:nNnTF
    { #4 } = { #2 }
    { #1 }
    { #3 }
  }
}
\cs_if_exist:NF \tex_filesize:D
{ \cs_gset:Npn \file_full_name:n #1
  { \msg_expandable_error:nnn { kernel } { primitive-not-available } }
  { \pdf)filesize }
}
\msg_new:nnnn { kernel } { primitive-not-available }
{ Primitive~\token_to_str:N #1 not~available }
{ The~version~of~your~\TeX~engine~does~not~provide~functionality~equivalent~to~the~#1~primitive. }

(End definition for \file_full_name:n and others. This function is documented on page 92.)

These functions pre-date using \tex_filesize:D for file searching, so are get functions with protection. To avoid having different search set ups, they are simply wrappers
around the code above.

If \texttt{\textproc{texfilesize:D}} is not available, the way to test if a file exists is to try to open it: if it does not exist then \TeX{} reports end-of-file. A search is made looking at each potential path in turn (starting from the current directory). The first location is of course treated as the correct one: this is done by jumping to \texttt{\prg\prg{break point}}. If nothing is found, \texttt{\l_2} is returned empty. A special case when there is no extension is that once the first location is found we test the existence of the file with \texttt{.tex} extension in that directory, and if it exists we include the \texttt{.tex} extension in the result.
\ior_if_eof:NF \g__file_internal_ior
{ \tl_put_right:Nn \l__file_full_name_tl {.tex} }
\ior_close:N \g__file_internal_ior
\tl_set_eq:NN \l_@file_full_name_tl \l__file_full_name_tl
\prg_return_true:

\ior_if_eof:NF \g__file_internal_ior
{ \tl_set_eq:NN \l__file_full_name_tl \file_full_name:n }
\ior_close:N \g__file_internal_ior
\prg_return_true:

\cs_new_protected:Npn \__file_get_full_name_search:nN #1#2
{ \__kernel_tl_set:Nx \l__file_full_name_tl {\tl_to_str:n {#1} \l__file_base_name_tl}
\__kernel_ior_open:No \g__file_internal_ior \l__file_full_name_tl
\ior_if_eof:NF \g__file_internal_ior { #2 { \prg_break: } } }

(End definition for \file_get_full_name:nN, \file_get_full_name:nNTF, and \__file_get_full_name_search:nN. These functions are documented on page 92.)

\g__file_internal_ior
A reserved stream to test for file existence (if required), and for opening a shell.

\ior_new:N \g__file_internal_ior
(End definition for \g__file_internal_ior.)

\file_mdfive_hash:n
\file_size:n
\file_timestamp:n
\__file_details:nn
\__file_details_aux:nn
\__file_mdfive_hash:n

Getting file details by expansion is relatively easy if a bit repetitive. As the MD5 function has a slightly different syntax from the other commands, there is a little cleaning up to do.

\cs_new:Npn \file_size:n #1
{ \__file_details:nn {#1} { size } }
\cs_new:Npn \file_timestamp:n #1
{ \__file_details:nn {#1} { moddate } }
\cs_new:Npn \__file_details:nn #1#2
{ \exp_args:Ne \__file_details_aux:nn
{ \file_full_name:n {#1} } {#2} }
\cs_new:Npn \__file_details_aux:nn #1#2
{ \tl_if_blank:nF {#1}
{ \use:c { tex_file #2 :D } {#1} }
}
\cs_new:Npn \file_mdfive_hash:n #1
{ \exp_args:Ne \__file_mdfive_hash:n
{ \file_full_name:n {#1} } }
\cs_new:Npn \__file_mdfive_hash:n #1
{ \tex_mdfivesum:D file {#1} }

(End definition for \file_mdfive_hash:n and others. These functions are documented on page 94.)

\file_hex_dump:nnn
\__file_hex_dump_auxi:nnn
\__file_hex_dump_auxii:nnnn
\__file_hex_dump_auxiii:nnnn
\__file_hex_dump_auxiv:nnnn
\__file_hex_dump:n
\file_hex_dump:n

These are separate as they need multiple arguments or the file size. For LuaTeX, the emulation does not need the file size so we save a little on expansion.

\cs_new:Npn \file_hex_dump:nnn #1#2#3
{ \exp_args:Ne \__file_hex_dump_auxi:nnn
{ \file_full_name:n {#1} } }
\cs_new:Npn \__file_hex_dump_auxi:nnn
{ \file_hex_dump:n {#1} }

635
\cs_new:Npn \__file_hex_dump_auxi:nnn \#1\#2\#3
\{
  \bool_lazy_any:nF
  { \tl_if_blank_p:n \#1 }
  { \int_compare_p:nNn \#2 = 0 }
  { \int_compare_p:nNn \#3 = 0 }
  {\exp_args:Ne \__file_hex_dump_auxii:nnnn
   \{ \__file_details_aux:nn \#1 \{ size \} \}
    \#1 \#2 \#3}
\}
\cs_new:Npn \__file_hex_dump_auxii:nnnn \#1\#2\#3\#4
\{
  \int_compare:nNnTF \#3 > 0
  { \__file_hex_dump_auxiii:nnnn \#3 }
  { \exp_args:Ne \__file_hex_dump_auxiii:nnnn
    \{ \int_eval:n \{ \#1 + \#3 \} \}
     \#1 \#2 \#4}
\}
\cs_new:Npn \__file_hex_dump_auxiii:nnnn \#1\#2\#3\#4
\{
  \int_compare:nNnTF \#4 > 0
  { \__file_hex_dump_auxiv:nnn \#4 }
  { \exp_args:Ne \__file_hex_dump_auxiv:nnn
    \{ \int_eval:n \{ \#2 + \#4 \} \}
     \#1 \#3}
\}
\cs_new:Npn \__file_hex_dump_auxiv:nnn \#1\#2\#3
\{
  \tex_filedump:D
  offset ~ \int_eval:n \{ \#2 - 1 \} ~
  length ~ \int_eval:n \{ \#1 - \#2 + 1 \}
  \#3}
\cs_new:Npn \file_hex_dump:n \#1
\{
  \exp_args:Ne \__file_hex_dump:n \{ \file_full_name:n \#1 \}
  \sys_if_engine_luatex:TF
  \{
   \cs_new:Npn \__file_hex_dump:n \#1
    \{ \tl_if_blank:nF \#1
      \{ \tex_filedump:D whole \#1 \\#1 \}
    \}
  \}
\}
636
\input{file_hex_dump:nnn}

\begin{quote}
Non-expandable wrappers around the above in the case where appropriate primitive support exists.
\end{quote}

\begin{quote}
Where the primitive is not available, issue an error: this is a little more conservative than absolutely needed, but does work.
\end{quote}
\__file_str_cmp:nn
As we are doing a fixed-length “big” integer comparison, it is easiest to use the low-level behavior of string comparisons.
\__file_compare_timestamp:nnN
Comparison of file date can be done by using the low-level nature of the string comparison functions.
\prg_return_true:
\else:
\prg_return_false:
\fi:
{
\if_int_compare:w
\__file_str_cmp:nn
{ \__file_timestamp:n {#1} }
{ \__file_timestamp:n {#2} }
\prg_return_true:
\else:
\prg_return_false:
\fi:
}
\cs_new_eq:NN \__file_timestamp:n \tex_filemoddate:D
\cs_if_exist:NF \__file_timestamp:n
{ \file_compare_timestamp:nNn #1#2#3
  { \msg_expandable_error:nnn { kernel } { primitive-not-available }
    { \spr{pdf}filemoddate }
    \prg_return_false:
  }
}

(End definition for \file_compare_timestamp:nNnTF, \__file_compare_timestamp:nNF, and \__file_timestamp:n. This function is documented on page 95.)

\file_if_exist:nTF The test for the existence of a file is a wrapper around the function to add a path to a file. If the file was found, the path contains something, whereas if the file was not located then the return value is empty.

\prg_new_protected_conditional:Npnn \file_if_exist:n #1 { T , F , TF }
{ \file_get_full_name:nNTF {#1} \l__file_full_name_tl
{ \prg_return_true: }
{ \prg_return_false: }
}

(End definition for \file_if_exist:nTF. This function is documented on page 92.)

\file_if_exist_input:n The test for the existence of a file is a wrapper around the function to add a path to a file. If the file was found, the path contains something, whereas if the file was not located then the return value is empty.

\prg_new_protected_conditional:Npnn \file_if_exist_input:n #1 { T , F , TF }
{ \file_get_full_name:nNF {#1} \l__file_full_name_tl
{ \prg_return_true: }
{ \prg_return_false: }
}

(End definition for \file_if_exist:nNF. This function is documented on page 92.)

\file_if_exist_input:nF Input of a file with a test for existence. We do not define the T or TF variants because the most useful place to place the (true code) would be inconsistent with other conditionals.

\cs_new_protected:Npn \file_if_exist_input:n #1
{ \file_get_full_name:nNTF {#1} \l__file_full_name_tl
{ \spr{V} \l__file_full_name_tl }
}

\cs_new_protected:Npn \file_if_exist_input:nF #1#2
\file_input\stop: A simple rename.
\cs_new_protected:Npn \file_input\stop: { \tex_endinput:D }
(End definition for \file_input\stop: This function is documented on page 95.)

\kernel_file_missing:n An error message for a missing file, also used in \ior_open:Nn.
\cs_new_protected:Npn \kernel_file_missing:n #1
  { \msg_error:nnx { kernel } { file-not-found } { \kernel_file_name_sanitise:n {#1} } }
(End definition for \kernel_file_missing:n)

\file_input:n Loading a file is done in a safe way, checking first that the file exists and loading only if it does. Push the file name on the \g__file_stack_seq, and add it to the file list, either \g__file_record_seq, or \@filelist in package mode.
\cs_new_protected:Npn \file_input:n #1
  { \file_get_full_name:nNTF {#1} \l__file_full_name_tl
    \__file_input:V \l__file_full_name_tl
    \__kernel_file_missing:n {#1} }
\cs_new_protected:Npx \file_input:n #1
  { \exp_not:N \clist_if_exist:NTF \exp_not:N \@filelist
    { \exp_not:N \@addtofilelist {#1} }
    { \seq_gput_right:Nn \exp_not:N \g__file_record_seq {#1} }
    \exp_not:N \__file_input_push:n {#1}
    \exp_not:N \tex_input:D
    \sys_if_engine_luatex:TF
      { {#1} }
      { \exp_not:N \__kernel_file_name_quote:n {#1} \scan_stop: }
    \exp_not:N \__file_input_pop:
  }
\cs_generate_variant:Nn \file_input:n { V }

Keeping a track of the file data is easy enough: we store the separated parts so we do not need to parse them twice.
\cs_new_protected:Npn \__file_input_push:n #1
  { \exp_not:N \clist_if_exist:NTF \exp_not:N \filelist
    { \exp_not:N \addtolist {#1} }
    { \seq_gput_right:Nn \exp_not:N \g__file_record_seq {#1} }
    \exp_not:N \__file_input_push:n {#1}
    \exp_not:N \tex_input:D
    \sys_if_engine_luatex:TF
      { {#1} }
      { \exp_not:N \__kernel_file_name_quote:n {#1} \scan_stop: }
    \exp_not:N \__file_input_pop:
  }
\cs_generate_variant:Nn \__file_input_push:n { V }

\seq_gpush:Nx \g__file_stack_seq
  { \g_file_curr_dir_str }
  { \g_file_curr_name_str }
  { \g_file_curr_ext_str }

640
The main parsing macro \texttt{\textbackslash file\_parse\_full\_name\_apply:nN} passes the file name \texttt{#1} through \texttt{\_kernel\_file\_name\_sanitize:n} so that we have a single normalised way to treat files internally. \texttt{\textbackslash file\_parse\_full\_name:n} uses the former, with \texttt{\textbackslash prg\_do\_nothing:} to leave each part of the name within a pair of braces.

\texttt{\textbackslash file\_parse\_full\_name\_area:nw} splits the file name into chunks separated by \texttt{/}, until the last one is reached. The last chunk is the file name plus the extension, and everything before that is the path. When \texttt{\_file\_parse\_full\_name\_area:nw} is done, it leaves the path within braces after the scan mark \texttt{\_file\_stop} and proceeds parsing the actual file name.
\_file\_parse\_full\_name\_base:nw does roughly the same as above, but it separates the chunks at each period. However here there’s some extra complications: In case \#1 is empty, it is assumed that the extension is actually empty, and the file name is \#2. Besides, an extra . has to be added to \#2 because it is later removed in \_file\_parse\_full\_name\_tidy:nnnN. In any case, if there’s an extension, it is returned with a leading ..

\cs\_new:Npn \_file\_parse\_full\_name\_base:nw \#1 \#2 . \#3 \s\_file\_stop

\begin{verbatim}
\tl_if_empty:nTF {#3}
\{\tl_if_empty:nTF {#1}
\{\tl_if_empty:nTF {#2}
\{ \_file\_parse\_full\_name\_tidy:nnnN { } { } }
\{ \_file\_parse\_full\_name\_tidy:nnnN { .#2 } { } }
\}
\{ \_file\_parse\_full\_name\_tidy:nnnN {#1} { .#2 } }
\}
\{ \_file\_parse\_full\_name\_base:nw { #1 . #2 } #3 \s\_file\_stop }
\}
\end{verbatim}

Now we just need to tidy some bits left loose before. The loop used in the two macros above start with a leading / and . in the file path an name, so here we need to remove them, except in the path, if it is a single /, in which case it’s left as is. After all’s done, pass to \#4.

\cs\_new:Npn \_file\_parse\_full\_name\_tidy:nnnN \#1 \#2 \#3 \#4

\begin{verbatim}
\exp_args:Nee #4
\{\str_if_eq:nnF {#3} { / } { \use\_none:n }
#3 \prg\_do\_nothing:
#3
\{ \use\_none:n #1 \prg\_do\_nothing: }
#2
\}
\}
\end{verbatim}

(End definition for \file\_parse\_full\_name:n and others. These functions are documented on page 93.)

\file\_parse\_full\_name:nNNN

\file\_parse\_full\_name:VNNN

\cs\_new\_protected:Npn \file\_parse\_full\_name:nNNN \#1 \#2 \#3 \#4

\begin{verbatim}
\_{file\_parse\_full\_name\_apply:nN {#1}
\_{file\_full\_name\_assign:nnnNN \#2 \#3 \#4 }
\)
\cs\_new\_protected:Npn \_file\_full\_name\_assign:nnnNN \#1 \#2 \#3 \#4 \#5 \#6

\begin{verbatim}
\{\str\_set:Nn \#4 {#1}
\str\_set:Nn \#5 {#2}
\str\_set:Nn \#6 {#3}
\}
\cs\_generate\_variant:Nn \file\_parse\_full\_name:nNNN \{ V \}
\end{verbatim}

(End definition for \file\_parse\_full\_name:nNNN. This function is documented on page 93.)

642
A function to list all files used to the log, without duplicates. In package mode, if \@filelist is still defined, we need to take this list of file names into account (we capture it \AtBeginDocument into \g__file_record_seq), turning it to a string (this does not affect the commas of this comma list).

\cs_new_protected:Npn \file_show_list: { \__file_list:N \msg_show:nnxxxx }
\cs_new_protected:Npn \file_log_list: { \__file_list:N \msg_log:nnxxxx }
\cs_new_protected:Npn \__file_list:N #1
{ \seq_clear:N \l__file_tmp_seq
clist_if_exist:NT \@filelist
{ \exp_args:NNx \seq_set_from_clist:Nn \l__file_tmp_seq
 { \tl_to_str:N \@filelist }
}
\seq_concat:NNN \l__file_tmp_seq \l__file_tmp_seq \g__file_record_seq
\seq_remove_duplicates:N \l__file_tmp_seq
\seq_map_function:NN \l__file_tmp_seq \__file_list_aux:n
}{ }{ }{ }
}
\cs_new:Npn \__file_list_aux:n #1 { \iow_newline: #1 }

(End definition for \file_show_list: and others. These functions are documented on page 95.)

When used as a package, there is a need to hold onto the standard file list as well as the new one here. File names recorded in \@filelist must be turned to strings before being added to \g__file_record_seq.

\cs_new_protected:Npn \GetIdInfo
{ \tl_clear_new:N \ExplFileDescription
\tl_clear_new:N \ExplFileDate
\seq_clear:N \l__file_tmp_seq
clist_if_exist:NT \@filelist
{ \exp_args:NNx \seq_set_from_clist:Nn \l__file_tmp_seq
 { \tl_to_str:N \@filelist }
\seq_concat:NNN \g__file_record_seq \g__file_record_seq \l__file_tmp_seq
\seq_remove_duplicates:N \l__file_tmp_seq
\seq_map_function:NN \l__file_tmp_seq \__file_id_info_auxi:w
\__file_id_info_auxii:w
\__file_id_info_auxiii:w
}{ }{ }{ }
}

49.5 GetIdInfo

As documented in expl3.dtx this function extracts file name etc from an SVN Id line. This used to be how we got version number and so on in all modules, so it had to be defined in \L3شرط. Now it’s more convenient to define it after we have set up quite a lot of tools, and \L3file seems the least unreasonable place for it.

The idea here is to extract out the information needed from a standard SVN Id line, but to avoid a line that would get changed when the file is checked in. Hence the fact that none of the lines here include both a dollar sign and the Id keyword!

\cs_new_protected:Npn \GetIdInfo
{ \tl_clear_new:N \ExplFileDescription
\tl_clear_new:N \ExplFileDate

643
A first check for a completely empty svn field. If that is not the case, there is a second case when a file created using svn cp but has not been checked in. That leaves a special marker -1 version, which has no further data. Dealing correctly with that is the reason for the space in the line to use __file_id_info_auxii:w.

Here, #1 is Id, #2 is the file name, #3 is the extension, #4 is the version, #5 is the check in date and #6 is the check in time and user, plus some trailing spaces. If #4 is the marker -1 value then #5 and #6 are empty.

Convert an svn-style date into a \LaTeX-style one.

This function is responsible for checking if dependencies of the \LaTeXX3 kernel match the version preloaded in the \LaTeXE kernel. If versions don’t match, the function attempts to tell why by searching for a possible stray format file.

The function starts by checking that the kernel date is defined, and if not zero is used to force the error route. The kernel date is then compared with the argument requested.
date (usually the packaging date of the dependency). If the kernel date is less than the
required date, it’s an error and the loading should abort.

\cs_new_protected:Npn \__kernel_dependency_version_check:Nn #1
\exp_args:NV \__kernel_dependency_version_check:nn #1
\cs_new_protected:Npn \__kernel_dependency_version_check:nn #1
\cs_if_exist:NTF \c__kernel_expl_date_tl
\exp_args:NV \__file_kernel_dependency_compare:nnn
\c__kernel_expl_date_tl {#1}
\cs_new_protected:Npn \__file_kernel_dependency_compare:nnn #1 #2 #3
\int_compare:nNnT {\__file_parse_version:w #1 \s__file_stop} <
{\__file_parse_version:w #2 \s__file_stop}
{\__file_mismatched_dependency_error:nn {#2} {#3}}
\cs_new:Npn \__file_parse_version:w #1 - #2 - #3 \s__file_stop {#1#2#3}

If the versions differ, then we try to give the user some guidance. This function starts by
taking the engine name \c_sys_engine_str and replacing \text{tex} by \text{latex}, then building
a command of the form: \text{kpsewhich} –all –engine=⟨engine⟩ ⟨format⟩[-dev].fmt to query
the format files available. A shell is opened and each line is read into a sequence.

\begin{macrocode}
\cs_new_protected:Npn \__file_mismatched_dependency_error:nn #1 #2
\exp_args:NNx \ior_shell_open:Nn \g__file_internal_ior
\kpsewhich ~ --all ~ --engine = \c_sys_engine_exec_str
\c_space_tl \c_sys_engine_format_str
\bool_lazy_and:nnT {\tl_if_exist_p:N \development@branch@name }
{ ! \tl_if_empty_p:N \development@branch@name }
{-dev } .fmt
\seq_clear:N \l__file_tmp_seq
\ior_map_inline:Nn \g__file_internal_ior
{ \seq_put_right:Nn \l__file_tmp_seq {##1} }
\ior_close:N \g__file_internal_ior
\msg_error:nnnn { kernel } { mismatched-support-file } {#1} {#2}
\end{macrocode}

And finish by ending the current file.

\begin{macrocode}
\text_endinput:D
\end{macrocode}
Mismatched LaTeX support files detected. \%
Loading-’#2’-aborted!
\c__kernel_expl_date_tl may not exist, due to an older format, so only print the dates when the sentinel token list exists:
\tl_if_exist:NT \c__kernel_expl_date_tl
{ \%
  \%
The-L3-programming-layer-in-the-LaTeX-format \%
is-dated-\c__kernel_expl_date_tl,-but-in-your-TeX-
tree-the-files-require \% at-least-#1.
}\%
}
}

The sequence containing the format files should have exactly one item: the format file currently being run. If that’s the case, the cause of the error is not that, so print a generic help with some possible causes. If more than one format file was found, then print the list to the user, with appropriate indications of what’s in the system and what’s in the user tree.
\int_compare:nNnTF { \seq_count:N \l__file_tmp_seq } > 1
{ 
  \%
The-cause-seems-to-be-an-old-format-file-in-the-user-tree. \%
LaTeX-found-these-files:
\seq_map_tokens:Nn \l__file_tmp_seq { \%~\use:n } \%
Try-deleting-the-file-in-the-user-tree-then-run-LaTeX-again.
}
{ 
  \%
The-most-likely-causes-are:
  \%\---A-recent-format-generation-failed;
  \%\---A-stray-format-file-in-the-user-tree-which-needs-
to-be-removed-or-rebuilt;
  \%\---You-are-running-a-manually-installed-version-of-#2 \%
  \%\% which-is-incompatible-with-the-version-in-LaTeX. \%
}
\%
LaTeX-will-abort-loading-the-incompatible-support-files-
but-this-may-lead-to \% later-errors.-Please-ensure-that-
your-LaTeX-format-is-correctly-regenerated.
}

(End definition for \_kernel_dependency_version_check:Nn and others.)

49.7 Messages
\msg_new:nnn { kernel } { file-not-found }
{ File-’#1’-not-found. }
{ 
  The-requested-file-could-not-be-found-in-the-current-directory,-
in-the-Tex-search-path-or-in-the-LaTeX-search-path.
}
\msg_new:nnn { kernel } { file-list }
{ >File-List<<

646
\c_sys_platform_str

49.8 Functions delayed from earlier modules

Detecting the platform on LuaTeX is easy: for other engines, we use the fact that the two common cases have special null files. It is possible to probe further (see package platform), but that requires shell escape and seems unlikely to be useful. This is set up here as it requires file searching.
We can now set up the tests.

\clist_map_inline:nn { unix , windows }
{ \__file_const:nn { sys_if_platform_ #1 }
  \if_eq_p:Vn \c_sys_platform_str { #1 }
  \__file_const:nn { sys_if_platform_ #1 }
}

(End definition for \sys_if_platform_unix:TF and \sys_if_platform_windows:TF. These functions are documented on page 72.)
Chapter 50

\texttt{\textsc{l3}luatex} implementation

\section*{50.1 Breaking out to Lua}

\lstinline|\begin{verbatim}\begin{verbatim}
\_lua_escape:n  \_lua_now:n  \_lua_shipout:n
\cs_new_eq:NN \_lua_escape:n  \tex_luaescapestring:D
\cs_new_eq:NN \_lua_now:n  \tex_directlua:D
\cs_new_eq:NN \_lua_shipout:n  \tex_latelua:D
\end{verbatim}\end{verbatim}|

Covers of primitives.
\begin{verbatim}
\cs_new:Npn \lua_now:e \#1 { \_lua_now:n \exp_not:n \#1 }
\cs_new:Npn \lua_now:n \#1 { \lua_now:e \exp_not:n \#1 }
\cs_new_protected:Npn \lua_shipout_e:n \#1 \#1 { \_lua_shipout:n \#1 }
\cs_new_protected:Npn \lua_shipout:n \#1 \#1 { \lua_shipout_e:n \exp_not:n \#1 }
\sys_if_engine_luatex:F {
\clist_map_inline:nn { \_lua_escape:n , \_lua_escape:e , \_lua_now:n , \_lua_now:e }

\sys_if_engine_luatex:F {
\cs_set:Npn \_lua_escape:n \exp_not:n \#1
\cs_set:Npn \_lua_escape:e \exp_not:n \#1
\end{verbatim}\end{verbatim}|

Wrappers around the primitives. As with engines other than \texttt{LuaTEX} these have to be macros, we give them the same status in all cases. When \texttt{LuaTEX} is not in use, simply give an error message/
50.2 Messages

\msg_new:nnnn { luatex } { luatex-required } { LuaTeX~engine~not~in~use!~Ignoring~#1. }
\{ The~feature~you~are~using~is~only~available~with~the~LuaTeX~engine.~LaTeX3~ignored~’#1’. \}
\langle /tex \rangle

50.3 Lua functions for internal use

Most of the emulation of pdfTeX here is based heavily on Heiko Oberdiek’s \texttt{pdftex-cmds} package.

\begin{verbatim}
local io = io
local kpse = kpse
local lfs = lfs
local math = math
local md5 = md5
local os = os
local string = string
local tex = tex
local texio = texio
local tonumber = tonumber
\end{verbatim}

(End definition for \texttt{l3kernel} and \texttt{ltx.utils}. These functions are documented on page 97.)

Local copies of global tables.

\begin{verbatim}
local io = io
local kpse = kpse
local lfs = lfs
local math = math
local md5 = md5
local os = os
local string = string
local tex = tex
local texio = texio
local tonumber = tonumber
\end{verbatim}
Local copies of standard functions.

```plaintext
local abs = math.abs
local byte = string.byte
local floor = math.floor
local format = string.format
local gsub = string.gsub
local lfs_attr = lfs.attributes
local open = io.open
local os_date = os.date
local setcatcode = tex.setcatcode
local sprint = tex.sprint
local cprint = tex.cprint
local write = tex.write
local write_nl = texio.write_nl
local utf8_char = utf8.char

local scan_int = token.scan_int or token.scan_integer
local scan_string = token.scan_string
local scan_keyword = token.scan_keyword
local put_next = token.put_next
local token_create = token.create

Since token.create only returns useful values after the tokens has been added to TeX's hash table, we define a variant which defines it first if necessary.

```plaintext
local token_create_safe
do
  local is_defined = token.is_defined
  local set_char = token.set_char
  local runtoks = tex.runtoks
  local let_token = token_create'let'
  function token_create_safe(s)
    local orig_token = token_create(s)
    if is_defined(s, true) then
      return orig_token
    end
    set_char(s, 0)
    local new_token = token_create(s)
    runtoks(function()
      put_next(let_token, new_token, orig_token)
    end)
    return new_token
  end
  end

local true_tok = token_create_safe'prg_return_true:'
local false_tok = token_create_safe'prg_return_false:'
```

In ConTeXt lmtx token.command_id does not exist, but it can easily be emulated with ConTeXt's tokens.commands.

```plaintext
local command_id = token.command_id
if not command_id and tokens and tokens.commands then
  local id_map = tokens.commands
  function command_id(name)
```
local function deprecated(table, name, func)
    table[name] = function(...) 
        write_nl(format("Calling deprecated Lua function \%s\", name))
        table[name] = func
        return func(...) 
    end 
end 

local function escapehex(str)
    return (gsub(str, ".", function (ch) return format("%02X", byte(ch)) end))
end 

l3kernel.charcat
Creating arbitrary chars using tex.cprint. The alternative approach using token.put_next(token.create(...))
would be about 10% slower.

deprecated(l3kernel, 'charcat', function(charcode, catcode)
    cprint(catcode, utf8_char(charcode))
end) 

l3kernel.elapsedtime
l3kernel.resettimer
Simple timing set up: give the result from the system clock in scaled seconds.

local os_clock = os.clock 
local base_clock_time = 0 
local function elapsedtime()
    local val = (os_clock() - base_clock_time) * 65536 + 0.5 
    if val > 2147483647 then 
        val = 2147483647 
    end 
    write(format("\%d",floor(val)))
end 
l3kernel.elapsedtime = elapsedtime 
local function resettimer()
    base_clock_time = os_clock()
end 
l3kernel.resettimer = resettimer 

652
Similar comments here to the next function: read the file in binary mode to avoid any line-end weirdness.

```lua
local function filedump(name, offset, length)
  local file = kpse_find(name, "tex", true)
  if not file then return end
  local f = open(file, "rb")
  if not f then return end
  if offset and offset > 0 then
    f:seek("set", offset)
  end
  local data = f:read(length or 'a')
  f:close()
  return escapehex(data)
end
ltxutils.filedump = filedump
deprecated(l3kernel, "filedump", function(name, offset, length)
  local dump = filedump(name, tonumber(offset), tonumber(length))
  if dump then
    write(dump)
  end)
```

(End definition for ltx.utils.filedump and l3kernel.filedump. These functions are documented on page 97.)

---

**md5.HEX**

Hash a string and return the hash in uppercase hexadecimal format. In some engines, this is build-in. For traditional LuaTeX, the conversion to hexadecimal has to be done by us.

```lua
local md5_HEX = md5.HEX
if not md5_HEX then
  local md5_sum = md5.sum
  function md5_HEX(data)
    return escapehex(md5_sum(data))
  end
  md5.HEX = md5_HEX
end
```

(End definition for md5.HEX. This function is documented on page ??.)

---

**ltx.utils.filemd5sum**

Read an entire file and hash it: the hash function itself is a built-in. As Lua is byte-based there is no work needed here in terms of UTF-8 (see pdftexcmds and how it handles strings that have passed through LuaTeX). The file is read in binary mode so that no line ending normalisation occurs.

```lua
local function filemd5sum(name)
  local file = kpse_find(name, "tex", true) if not file then return end
  local f = open(file, "rb") if not f then return end
  local data = f:read("*a")
  f:close()
  return md5.HEX(data)
end
ltxutils.filemd5sum = filemd5sum
deprecated(l3kernel, "filemdfivesum", function(name)
  local hash = filemd5sum(name)
```
if hash then
  write(hash)
end

(End definition for \texttt{ltx.utils.filemd5sum} and \texttt{l3kernel.filemdfivesum}. These functions are documented on page 97.)

\texttt{ltx.utils.filemoddate}
\texttt{l3kernel.filemoddate}

There are two cases: If the C standard library is C99 compliant, we can use \texttt{%z} to get the timezone in almost the right format. We only have to add primes and replace a zero or missing offset with \texttt{Z}.

Of course this would be boring, so Windows does things differently. There we have to manually calculate the offset. See procedure \texttt{makepdftime} in \texttt{utils.c} of \texttt{pdfTeX}.

\begin{verbatim}
local filemoddate
if os_date('%z':match'^[+-]%d%d%d%d' then
  local pattern = lpeg.Cs(16 * (lpeg.Cg(lpeg.S'+-') * '0000' * lpeg.Cc'Z')
    + 3 * lpeg.Cc"" * 2 * lpeg.Cc"
    + lpeg.Cc'Z')
  function filemoddate(name)
    local file = kpse_find(name, "tex", true)
    if not file then return end
    local date = lfs_attr(file, "modification")
    if not date then return end
    return pattern:match(os_date("D:%Y%m%d%H%M%S%z", date))
  end
else
  local function filemoddate(name)
    local file = kpse_find(name, "tex", true)
    if not file then return end
    local date = lfs_attr(file, "modification")
    if not date then return end
    local d = os_date("*t", date)
    local u = os_date("!*t", date)
    local off = 60 * (d.hour - u.hour) + d.min - u.min
    if d.year ~= u.year then
      if d.year > u.year then
        off = off + 1440
      elseif d.yday ~= u.yday then
        if d.yday > u.yday then
          off = off + 1440
        elseif d.yday < u.yday then
          off = off - 1440
        end
      end
    end
    local timezone
    if off == 0 then
      timezone = "Z"
    else
      if off < 0 then
        timezone = "-"..off
      elseif off > 0 then
        timezone = "+"..off
      end
    end
  end
end

end

\end{verbatim}
timezone = ":-"
off = -off
else
timezone = "+"
end
timezone = format("%s%02d'\%02d'\", timezone, hours // 60, hours % 60)
end
return format("D:%04d%02d%02d%02d%02d%02d%s", d.year, d.month, d.day, d.hour, d.min, d.sec, timezone)
end
end
ltxutils.filemoddate = filemoddate
deprecated(l3kernel, "filemoddate", function(name)
local hash = filemoddate(name)
if hash then
write(hash)
end
end)

(End definition for ltx.utils.filemoddate and l3kernel.filemoddate. These functions are documented on page 98.)

ltx.utils.filesize  A simple disk lookup.
l3kernel.filesize

local function filesize(name)
local file = kpse_find(name, "tex", true)
if file then
local size = lfs_attr(file, "size")
if size then
return size
end
end
end
ltxutils.filesize = filesize
deprecated(l3kernel, "filesize", function(name)
local size = filesize(name)
if size then
write(size)
end
end)

(End definition for ltx.utils.filesize and l3kernel.filesize. These functions are documented on page 98.)

l3kernel.strcmp  String comparison which gives the same results as pdfTeX’s \pdfstrcmp, although the ordering should likely not be relied upon!

deprecated(l3kernel, "strcmp", function (A, B)
if A == B then
write("0")
elseif A < B then
write("-1")
else
write("1")
end
end)
l3kernel.shellescape

Replicating the pdfTeX log interaction for shell escape.

```lua
local os_exec = os.execute

deprecated(l3kernel, "shellescape", function(cmd)
  local status, msg = os_exec(cmd)
  if status == nil then
    write_nl("log","runsystem(" .. cmd .. ")...(" .. msg .. ")\n")
  elseif status == 0 then
    write_nl("log","runsystem(" .. cmd .. ")...executed\n")
  else
    write_nl("log","runsystem(" .. cmd .. ")...failed " .. (msg or ") .. ")\n")
  end
end)
```

(End definition for l3kernel.shellescape. This function is documented on page 98.)

luadef

An internal function for defining control sequences form Lua which behave like primitives. This acts as a wrapper around token.set_lua which accepts a function instead of an index into the functions table.

```lua
local luacmd do
  local set_lua = token.set_lua
  local undefined_cs = command_id'undefined_cs'
  if not context and not luatexbase then require'ltluatex' end
  if luatexbase then
    local new_luafunction = luatexbase.new_luafunction
    local functions = lua.get_functions_table()
    function luacmd(name, func, ...)
      local id
      local tok = token_create(name)
      if tok.command == undefined_cs then
        id = new_luafunction(name)
        set_lua(name, id, ...)
      else
        id = tok.index or tok.mode
        functions[id] = func
      end
      elseif context then
        local register = context.functions.register
        local functions = context.functions.known
        function luacmd(name, func, ...)
          local tok = token_create(name)
          if tok.command == undefined_cs then
            token.set_lua(name, register(func), ...)
          else
            functions[tok.index or tok.mode] = func
          end
        end
      end
  end
end
```

(End definition for luadef.)
50.4 Preserving iniTeX Lua data for runs

The Lua state is not dumped when a format is written, therefore any Lua variables filled doing format building need to be restored in order to be accessible during normal runs.

We provide some kernel-internal helpers for this. They will only be available if luatexbase is available. This is not a big restriction though, because ConTeXt (which does not use luatexbase) does not load expl3 in the format.

```latex
local register_luadata, get_luadata

if luatexbase then
  local register = token_create'@expl@luadata@bytecode'.index
  if status.ini_version then
    register_luadata
  end
  register_luadata

register_luadata register_luadata is only available during format generation. It accept a string which uniquely identifies the data object and has to be provided to retrieve it later. Additionally it accepts a function which is called in the pre_dump callback and which has to return a string that evaluates to a valid Lua object to be preserved.

```latex

```
local luadata, luadata_order = {}, {}

function register_luadata(name, func)
  if luadata[name] then
    error(format("LaTeX error: data name %q already in use", name))
  end
  luadata[name] = func
  luadata_order[#luadata_order + 1] = func and name
end

(End definition for register_luadata. This function is documented on page ??.)

The actual work is done in pre_dump. The luadata_order is used to ensure that the order is consistent over multiple runs.

```latex

```

```
luatexbase.add_to_callback("pre_dump", function()
  if next(luadata) then
    local str = "return {
      for i=1, #luadata_order do
        local name = luadata_order[i]
        str = format("%s[%q]=%s,, str, name, luadata[name]()})
      end
      lua.bytecode[register] = assert(load(str .. ")"))
    end
    end, "Itx.luadata")
  else
    get_luadata

get_luadata get_luadata is only available if data should be restored. It accept the identifier which was used when the data object was registered and returns the associated object. Every object can only be retrieved once.

```latex

```

```
local luadata = lua.bytecode[register]
if luadata then
  lua.bytecode[register] = nil
  luadata = luadata()
end
function get_luadata(name)

```
if not luadata then return end
local data = luadata[name]
luadata[name] = nil
return data
end
end
end

(End definition for get_luadata. This function is documented on page ??.)

⟨/lua⟩

⟨/package⟩
Chapter 51

l3legacy Implementation

A friendly wrapper.

\prg_new_conditional:Npn \legacy_if:n \prg_return_true:
\prg_return_false:
\fi:

(End definition for \legacy_if:nTF. This function is documented on page 99.)

A more elaborate wrapper.

\cs_new_protected:Npm \legacy_if_set:nn \#1\#2
\cs_new_protected:Npm \legacy_if_gset:nn \#1\#2
(End definition for \ legacy_if_set:nn and \ legacy_if_gset:nn. These functions are documented on page 99.)

11787 ⟨/package⟩
Chapter 52

\texttt{l3tl} implementation

A token list variable is a \TeX{} macro that holds tokens. By using the \texttt{\unexpanded} \\texttt{\edef} primitive inside a \TeX{} \texttt{\edef} it is possible to store any tokens, including \#\texttt{, in this way.}

### 52.1 Functions

\texttt{\__kernel_tl_set:Nx}  \texttt{\__kernel_tl_gset:Nx}

These two are supplied to get better performance for macros which would otherwise use \texttt{\tl_set:Nx} or \texttt{\tl_gset:Nx} internally.

\texttt{\cs_new_eq:NN \__kernel_tl_set:Nx \cs_set_nopar:Npx}
\texttt{\cs_new_eq:NN \__kernel_tl_gset:Nx \cs_gset_nopar:Npx}

(End definition for \texttt{\__kernel_tl_set:Nx} and \texttt{\__kernel_tl_gset:Nx}.)

\texttt{\tl_new:N}  \texttt{\tl_new:c}

Creating new token list variables is a case of checking for an existing definition and doing the definition.

\texttt{\cs_new_protected:Npm \tl_new:N \#1}
\texttt{\__kernel_chk_if_free_cs:N \#1}
\texttt{\cs_gset_eq:NN \#1 \c_empty_tl}
\texttt{\cs_generate_variant:Nn \tl_new:N { c }}

(End definition for \texttt{\tl_new:N}. This function is documented on page 101.)

\texttt{\tl_const:Nn}  \texttt{\tl_const:Nx}  \texttt{\tl_const:cn}  \texttt{\tl_const:cx}

Constants are also easy to generate. They use \texttt{\cs_gset_nopar:Npx} instead of \texttt{\__kernel_tl_gset:Nx} so that the correct scope checking is applied if \texttt{l3debug} is used.

\texttt{\cs_new_protected:Npm \tl_const:Nn \#1}\#2
\texttt{\__kernel_chk_if_free_cs:N \#1}
\texttt{\cs_gset_nopar:Npx \#1 { \__kernel_exp_not:w \#2 } }
\texttt{\cs_generate_variant:Nn \tl_const:Nn \#1}\#2
\texttt{\cs_new_protected:Npm \tl_const:Nx \#1}\#2
\texttt{\__kernel_chk_if_free_cs:N \#1}
\texttt{\cs_gset_nopar:Npx \#1 \#2}
Clearing a token list variable means setting it to an empty value. Error checking is sorted out by the parent function.

\cs_new_protected:Npn \tl_clear:N #1
\{ \tex_let:D #1 = \empty_tl \}
\cs_new_protected:Npn \tl_gclear:N #1
\{ \tex_global:D \tex_let:D #1 = \empty_tl \}
\cs_generate_variant:Nn \tl_clear:N { c }
\cs_generate_variant:Nn \tl_gclear:N { c }

(End definition for \tl_clear:N and \tl_gclear:N. These functions are documented on page 102.)

Clearing a token list variable means setting it to an empty value. Error checking is sorted out by the parent function.

\cs_new_protected:Npn \tl_clear_new:N #1
\{ \tl_if_exist:NTF #1 \{ \tl_clear:N #1 \} \{ \tl_new:N #1 \} \}
\cs_new_protected:Npn \tl_gclear_new:N #1
\{ \tl_if_exist:NTF #1 \{ \tl_gclear:N #1 \} \{ \tl_new:N #1 \} \}
\cs_generate_variant:Nn \tl_clear_new:N { c }\cs_generate_variant:Nn \tl_gclear_new:N { c }\cs_generate_variant:Nn \tl_clear:N { c }
\cs_generate_variant:Nn \tl_gclear:N { c }

(End definition for \tl_clear_new:N and \tl_gclear_new:N. These functions are documented on page 102.)

For setting token list variables equal to each other. To allow for patching, the arguments have to be explicit. In addition this ensures that a braced second argument will not cause problems.

\cs_new_protected:Npn \tl_set_eq:NN \{ \__kernel_tl_set:Nx \}
\cs_new_protected:Npn \tl_gset_eq:NN \{ \__kernel_tl_gset:Nx \}
\cs_generate_variant:Nn \tl_set_eq:NN { cN, Nc, cc }\cs_generate_variant:Nn \tl_gset_eq:NN { cN, Nc, cc }

(End definition for \tl_set_eq:NN and \tl_gset_eq:NN. These functions are documented on page 102.)

Concatenating token lists is easy. When checking is turned on, all three arguments must be checked: a token list \#2 or \#3 equal to \scan_stop: would lead to problems later on.

\cs_new_protected:Npn \tl_concat:NNN \{ \__kernel_tl_set:Nx \}
\cs_new_protected:Npn \tl_gconcat:NNN \{ \__kernel_tl_gset:Nx \}
\cs_generate_variant:Nn \tl_concat:NNN { cC }\cs_generate_variant:Nn \tl_gconcat:NNN { cC }

(End definition for \tl_concat:NNN and \tl_gconcat:NNN. These functions are documented on page 102.)
52.2 Constant token lists
\c_empty_tl
Never full. We need to define that constant before using \tl_new:N.
\c_novalue_tl
A special marker: as we don’t have \char_generate:nn yet, has to be created the old-fashioned way.
\c_space_tl
A space as a token list (as opposed to as a character).
52.3 Adding to token list variables
\tl_set:Nn
By using \exp_not:n token list variables can contain # tokens, which makes the token list registers provided by \TeX more or less redundant. The \tl_set:No version is done by hand as it is used quite a lot.
\tl_put_left:Nn Adding to the left is done directly to gain a little performance.
\tl_put_left:NV \cs_new_protected:Npm \tl_put_left:Nn #1#2
\tl_put_left:No \cs_new_protected:Npm \tl_put_left:Nx #1#2
\tl_put_left:Nx \cs_new_protected:Npm \tl_put_left:Nv #1#2
\tl_put_left:cn \cs_new_protected:Npm \tl_put_left:Nf #1#2
\tl_put_left:cV \cs_new_protected:Npm \tl_put_left:no #1#2
\tl_put_left:co \cs_new_protected:Npm \tl_put_left:co #1#2
\tl_put_left:cx \cs_new_protected:Npm \tl_put_left:cx #1#2

(End definition for \tl_set:Nn and \tl_gset:Nn. These functions are documented on page 102.)
\cs_generate_variant:Nn \tl_put_left:NV { c }
\cs_generate_variant:Nn \tl_put_left:No { c }
\cs_generate_variant:Nn \tl_put_left:Nx { c }
\cs_generate_variant:Nn \tl_gput_left:Nn { c }
\cs_generate_variant:Nn \tl_gput_left:NV { c }
\cs_generate_variant:Nn \tl_gput_left:No { c }
\cs_generate_variant:Nn \tl_gput_left:Nx { c }

(End definition for \tl_put_left:Nn and \tl_gput_left:Nn. These functions are documented on page 102.)

\tl_put_right:Nn
\tl_put_right:NV
\tl_put_right:No
\tl_put_right:Nx
\tl_put_right:cn
\tl_put_right:cV
\tl_put_right:co
\tl_put_right:cx
\tl_gput_right:Nn
\tl_gput_right:NV
\tl_gput_right:No
\tl_gput_right:Nx
\tl_gput_right:cn
\tl_gput_right:cV
\tl_gput_right:co
\tl_gput_right:cx

The same on the right.

\cs_new_protected:Npm \tl_put_right:Nn #1#2
\cs_new_protected:Npm \tl_put_right:No #1#2
\cs_new_protected:Npm \tl_put_right:Nx #1#2
\cs_new_protected:Npm \tl_gput_right:Nn #1#2
\cs_new_protected:Npm \tl_gput_right:No #1#2
\cs_new_protected:Npm \tl_gput_right:Nx #1#2
\cs_generate_variant:Nn \tl_put_right:Nn { c }
\cs_generate_variant:Nn \tl_put_right:NV { c }
\cs_generate_variant:Nn \tl_put_right:No { c }
\cs_generate_variant:Nn \tl_put_right:Nx { c }
\cs_generate_variant:Nn \tl_gput_right:Nn { c }
\cs_generate_variant:Nn \tl_gput_right:No { c }
\cs_generate_variant:Nn \tl_gput_right:Nx { c }
\cs_generate_variant:Nn \tl_gput_right:cx { c }
\cs_generate_variant:Nn \tl_gput_right:co { c }
\cs_generate_variant:Nn \tl_gput_right:cn { c }
\cs_generate_variant:Nn \tl_gput_right:cV { c }
\cs_generate_variant:Nn \tl_gput_right:No { c }
\cs_generate_variant:Nn \tl_gput_right:Nx { c }
\cs_generate_variant:Nn \tl_gput_right:cx { c }

665
52.4 Internal quarks and quark-query functions

Internal quarks.
\q__tl_nil 
\q__tl_mark 
\q__tl_stop

Internal recursion quarks.
\q__tl_recursion_tail
\q__tl_recursion_stop

Functions to query recursion quarks.
\__tl_if_recursion_tail_break:nN
\__tl_if_recursion_tail_stop_p:n
\__tl_if_recursion_tail_stop:n

52.5 Reassigning token list category codes

The rescanning code needs a special token list containing the same character (chosen here to be a colon) with two different category codes: it cannot appear in the tokens being rescanned since all colons have the same category code.
\tl_const:Nx \c__tl_rescan_marker_tl { : \token_to_str:N : }

In a group, after some initial setup explained below and the user setup \#3 (followed by \scan_stop: to be safe), there is a call to \__tl_set_rescan:nNN. This shared auxiliary defined later distinguishes single-line and multi-line “files”. In the simplest case of multi-line files, it calls (with the same arguments) \__tl_set_rescan_multi:nNN, whose code is included here to help understand the approach. This function rescans its argument \#1, closes the group, and performs the assignment.

One difficulty when rescanning is that \scantokens treats the argument as a file, and without the correct settings a \TeX error occurs:

! File ended while scanning definition of ...

A related minor issue is a warning due to opening a group before the \scantokens and closing it inside that temporary file; we avoid that by setting \tracingnesting. The standard solution to the “File ended” error is to grab the rescanned tokens as a delimited argument of an auxiliary, here \__tl_rescan:NNw, that performs the assignment, then let \TeX “execute” the end of file marker. As usual in delimited arguments we use \prg_do_-nothing: to avoid stripping an outer set braces: this is removed by using o-expanding assignments. The delimiter cannot appear within the rescanned token list because it contains twice the same character, with different catcodes.
For \texttt{\tl_rescan:nn} we cannot simply call \texttt{\_\tl_set_rescan:NNnn \prg_do_nothing: \use:n} because that would leave the end-of-file marker after the result of rescanning. If that rescanned result is code that looks further in the input stream for arguments, it would break.

For multi-line files the only subtlety is that \texttt{\newlinechar} should be equal to \texttt{\endlinechar} because \texttt{\newlinechar} characters become new lines and then become \texttt{\endlinechar} characters when writing to an abstract file and reading back. This equality is ensured by setting \texttt{\newlinechar} equal to \texttt{\endlinechar}. Prior to this, \texttt{\endlinechar} is set to $-1$ if it was 32 (in particular true after \texttt{\ExplSyntaxOn}) to avoid unreasonable line-breaks at every space for instance in error messages triggered by the user setup. Another side effect of reading back from the file is that spaces (catcode 10) are ignored at the beginning of lines, and spaces and tabs (character code 32 and 9) are ignored at the end of lines.

The two \texttt{\if_false: ... \fi:} are there to prevent alignment tabs to cause a change of tabular cell while rescanning. We put the “opening” one after \texttt{\group_begin:} so that if one accidentally \texttt{f-expands} \texttt{\tl_set_rescan:Nnn} braces remain balanced. This is essential in e-type arguments when \texttt{\expanded} is not available.
The function \_\_tl_set_rescan:nNN calls \_\_tl_set_rescan_multinNN or \_\_tl_set_rescan_single:nNN \{ \} depending on whether its argument is a single-line fragment of code/data or is made of multiple lines by testing for the presence of a newlinechar character. If newlinechar is out of range, the argument is assumed to be a single line.

For a single line, no \endlinechar should be added, so it is set to −1, and spaces should not be removed. Trailing spaces and tabs are a difficult matter, as \TeX removes these at a very low level. The only way to preserve them is to rescan not the argument but the argument followed by a character with a reasonable category code. Here, 11 (letter) and 12 (other) are accepted, as these are convenient, suitable for delimiting an argument, and it is very unlikely that none of the ASCII characters are in one of these categories. To avoid selecting one particular character to put at the end, whose category code may have been modified, there is a loop through characters from ‘ (ASCII 39) to ~ (ASCII 127). The choice of starting point was made because this is the start of a very long range of characters whose standard category is letter or other, thus minimizing the number of steps needed by the loop (most often just a single one). If no valid character is found (very rare), fall-back on \_\_tl_set_rescan_multinNN.

Otherwise, once a valid character is found (let us use ‘ in this explanation) run some code very similar to \_\_tl_set_rescan_multinNN but with ‘ added at both ends of the input. Of course, we need to define the auxiliary \_\_tl_set_rescan_single:nNNw on the fly to remove the additional ‘ that is just before :: (by which we mean \c\_\_tl_rescan_marker_tl). Note that the argument must be delimited by ‘ with the current catcode; this is done thanks to \char_generate:n. Yet another issue is that the rescanned token list may contain a comment character, in which case the ‘ we expected is not there. We fix this as follows: rather than just :: we set \everyeof to ::{\(\text{code1}\)} ‘::{\(\text{code2}\)} \_\_tl_stop. The auxiliary \_\_tl_set_rescan_single:nNNw runs the o-expanding assignment, expanding either \(\text{code1}\) or \(\text{code2}\) before the main argument \#3. In the typical case without comment character, \(\text{code1}\) is expanded, removing the leading ‘. In the rarer case with comment character, \(\text{code2}\) is expanded, calling \_\_tl_set_rescan_single_aux:w, which removes the trailing ::{\(\text{code1}\)} and the leading ‘.
52.6 Modifying token list variables

All of the replace functions call `\__tl_replace:NnNnnn` with appropriate arguments. The first two arguments are explained later. The next controls whether the replacement function calls itself (`\__tl_replace_next:w`) or stops (`\__tl_replace_wrap:w`) after
To implement the actual replacement auxiliary \texttt{\tl_replace} we need a \texttt{\tl_replace} auxiliary with the following properties:

- all occurrences of the \texttt{pattern} \#6 in \texttt{token list} \texttt{delimiter} belong to the \texttt{token list} and have no overlap with the \texttt{delimiter},
- the first occurrence of the \texttt{delimiter} in \texttt{token list} \texttt{delimiter} is the trailing \texttt{delimiter}.

We first find the building blocks for the \texttt{delimiter}, namely two tokens \texttt{A} and \texttt{B} such that \texttt{A} does not appear in \#6 and \#6 is not \texttt{B} (this condition is trivial if \#6 has more than one token). Then we consider the delimiters \texttt{A} and \texttt{A} \texttt{B} \texttt{A} \texttt{B} \texttt{A} \texttt{B}, for \( n \geq 1 \), where \texttt{A} \texttt{B} \texttt{B} \texttt{B} \texttt{B} denotes \texttt{n} copies of \texttt{A}, and we choose as our \texttt{delimiter} the first one which is not in the \texttt{token list}.

Every delimiter in the set obeys the first condition: \#6 does not contain \texttt{A} hence cannot be overlapping with the \texttt{token list} and the \texttt{delimiter}, and it cannot be within the \texttt{delimiter} since it would have to be in one of the two \texttt{B} hence be equal to this single token (or empty, but this is an error case filtered separately). Given the particular form of these delimiters, for which no prefix is also a suffix, the second condition is actually a consequence of the weaker condition that the \texttt{delimiter} choose does not appear in the \texttt{token list}. Additionally, the set of delimiters is such that a \texttt{token list} \texttt{n} tokens can contain at most \( O(n^{3/2}) \) of them, hence we find a \texttt{delimiter} with at most \( O(n^{3/2}) \) tokens in a time at most \( O(n^{3/2}) \). Bear in mind that these upper bounds are reached only in very contrived scenarios: we include the \texttt{A} in the list of delimiters to try, so that the \texttt{delimiter} is simply \texttt{q__tl_mark} in the most common situation where neither the \texttt{token list} nor the \texttt{pattern} contains \texttt{q__tl_mark}.

Let us now ahead, optimizing for this most common case. First, two special cases: an empty \texttt{pattern} \#6 is an error, and if \#1 is absent from both the \texttt{token list} \#5 and the \texttt{pattern} \#6 then we can use it as the \texttt{delimiter} through \texttt{\tl_replace} auxiliary \texttt{#1}. Otherwise, we end up calling \texttt{\tl_replace} repeatedly with the first two arguments \texttt{q__tl_mark} \texttt{?}, \texttt{?}, \texttt{?}, \texttt{?}, \texttt{?}, and so on, until \#6 does not contain the control sequence \#1, which we take as our \texttt{A}. The argument \#2 only serves to collect \texttt{characters} for \#1. Note that the order of the tests means that the
first two are done every time, which is wasteful (for instance, we repeatedly test for the
emptyness of #6). However, this is rare enough not to matter. Finally, choose (B) to be
\q_{\text{tl\_nil}} or \q_{\text{tl\_stop}} such that it is not equal to #6.

The \_\_tl_replace_auxi:NnnNNnN auxiliary receives \{⟨A⟩\} and \{(A)^n⟨B⟩\} as its
arguments, initially with \( n = 1 \). If “⟨A⟩ ⟨A⟩⟨B⟩ ⟨A⟩⟨B⟩” is in the ⟨token list⟩ then
increase \( n \) and try again. Once it is not anymore in the ⟨token list⟩ we take it as our
⟨delimiter⟩ and pass this to the auxii auxiliary.

\begin{verbatim}
\cs_new_protected:Npn \_\_tl_replace:NnNNNnn #1#2#3#4#5#6#7 #8
{ \tl_if_empty:nTF {#6} { \msg_error:nnx { kernel } { empty-search-pattern } { \tl_to_str:n {#7} } } { \tl_if_in:onTF { #5 #6 } {#1} { \tl_if_in:nnTF {#6} {#1} { \exp_args:Nc \_\_tl_replace:NnNNNnn {#2} {#2?} } { \__tl_quark_if_nil:nTF {#6} { \_\_tl_replace_auxi:NnnNNNnn #5 {#1} { #1 \q_{\text{tl\_stop}} } } { \_\_tl_replace_auxi:NnnNNNnn #5 {#1} { #1 \q_{\text{tl\_nil}} } } } \__tl_replace_auxii:nNNNnn {#1} \_\_\_tl_replace_auxii:nNNNnn #3#4#5 {#6} {#7} } \_\_\_tl_replace_auxii:nNNNnn #1\#2\#3 \_\_\_tl_replace_auxii:nNNNnn #1\#2\#3 \_\_\_tl_replace_next:w \_\_\_tl_replace_wrap:w \_\_\_tl_replace_wrap:w } \_\_\_tl_replace_wrap:w \_\_\_tl_replace_wrap:w \_\_\_tl_replace_wrap:w \_\_\_tl_replace_wrap:w
\end{verbatim}

The auxiliary \_\_tl_replace_auxii:nNNNnn receives the following arguments:

\{(delimiter)\} \langle function \rangle \langle assignment \rangle
\langle tl \rangle \langle \langle pattern \rangle \rangle \langle \langle replacement \rangle \rangle

All of its work is done between \group_align_safe Begin: and \group_align_safe End: to avoid issues in alignments. It does the actual replacement within \#3 \#4 \{\ldots\}, an
x-expanding \langle assignment \rangle \#3 to the \langle tl \rangle \var \#4. The auxiliary \_\_tl_replace_next:w is
called, followed by the \langle token list ⟩, some tokens including the \langle delimiter ⟩ \#1, followed
by the \langle pattern ⟩ \#5. This auxiliary finds an argument delimited by \#5 (the presence of a
trailing \#5 avoids runaway arguments) and calls \_\_tl_replace_wrap:w to test whether
this \#5 is found within the \langle token list ⟩ or is the trailing one.

If on the one hand it is found within the \langle token list ⟩, then \#1 cannot contain the
\langle delimiter ⟩ \#1 that we worked so hard to obtain, thus \_\_tl_replace_wrap:w gets \#1 as its own argument \#1, and protects it against the x-expanding assignment.
It also finds \exp_not:n as \#2 and does nothing to it, thus letting through \exp_not:n
\langle \langle replacement \rangle \rangle into the assignment. Note that \_\_tl_replace_next:w and \_\_tl_replace_wrap:w are always called followed by two empty brace groups. These are safe
because no delimiter can match them. They prevent losing braces when grabbing delimited arguments, but require the use of \texttt{\exp_not:o} and \texttt{\use_none:nn}, rather than simply \texttt{\exp_not:n}. Afterwards, \_\_tl_replace_next:w is called to repeat the replacement, or \_\_tl_replace_wrap:w if we only want a single replacement. In this second case, \#1 is the (remaining tokens) in the (token list) and \#2 is some (ending code) which ends the assignment and removes the trailing tokens \#5 using some \texttt{\if_false:} \{ \texttt{\fi:} \} trickery because \#5 may contain any delimiter.

If on the other hand the argument \#1 of \_\_tl_replace_next:w is delimited by the trailing (pattern) \#5, then \#1 is “{ } { } (token list) (delimiter) (ending code)”, hence \_\_tl_replace_wrap:w finds “{ } { } (token list)” as \#1 and the (ending code) as \#2. It leaves the (token list) into the assignment and unbraces the (ending code) which removes what remains (essentially the (delimiter) and (replacement)).

\begin{verbatim}
\cs_new_protected:Npn \_\_tl_replace_auxii:nNNNnn #1#2#3#4#5#6
{ \__kernel_exp_not:w \exp_after:wN { \use_none:nn ##1 } ##2 }
\cs_set:Npx \_\_tl_replace_next:w ##1 #5
{ \exp_not:N \_\_tl_replace_wrap:w ##1 \exp_not:n { #1
\exp_not:n { \exp_not:n {#6} } \exp_not:n { #2 { } } } }
#3 #4
{ \exp_after:wN \_\_tl_replace_next_aux:w
#4 #1
{ \if_false: { \fi: }
\exp_after:wN \use_none:n \exp_after:wN { \if_false: } \fi:
} #5
\__kernel_exp_not:w \exp_after:wN \use_none:nn \_\_tl_replace_next:w { } { }
\__kernel_exp_not:w \_\_tl_replace_next:w { } { }
\__kernel_exp_not:w \_\_tl_replace_next:w { } { }
\__kernel_exp_not:w \_\_tl_replace_next:w { } { }
(End definition for \_\_tl_replace:NnNNNnn and others.)
\cs_new_protected:Npn \_\_tl_replace_next_aux:w \_\_tl_replace_next:w { } { }
(End definition for \_\_tl_replace:Nn and \_\_tl_greplace:Nn. These functions are documented on page 115.)
\end{verbatim}
Removal is just a special case of replacement.

\begin{verbatim}
\cs_new_protected:Npn \tl_remove_all:Nn #1#2
{ \tl_replace_all:Nnn #1 {#2} { } }
\cs_new_protected:Npn \tl_gremove_all:Nn #1#2
{ \tl_greplace_all:Nnn #1 {#2} { } }
\cs_generate_variant:Nn \tl_remove_all:Nn { c }
\cs_generate_variant:Nn \tl_gremove_all:Nn { c }
\end{verbatim}

(End definition for $\tl_remove_all:Nn$ and $\tl_gremove_all:Nn$. These functions are documented on page 116.)

### 52.7 Token list conditionals

These functions check whether the token list in the argument is empty and execute the proper code from their argument(s).

\begin{verbatim}
\prg_new_conditional:Npnn \tl_if_empty:N #1 { p , T , F , TF }
{ \if_meaning:w #1 \c_empty_tl \prg_return_true: \else: \prg_return_false: \fi: }
\prg_generate_conditional_variant:Nnn \tl_if_empty:N
{ c } { p , T , F , TF }
\end{verbatim}

(End definition for $\tl_if_empty:NTF$. This function is documented on page 103.)

The $\if:w$ triggers the expansion of $\tl_to_str:n$ which converts the argument to a string: this is empty if and only if the argument is. Then $\if:w \scan_stop: \ldots \scan_stop:$ is true if and only if the string $\ldots$ is empty. It could be tempting to use $\if:w \scan_stop: #1 \scan_stop:$ directly. But this fails on a token list expanding to anything starting with $\scan_stop:$ leaving everything that follows in the input stream.

\begin{verbatim}
\prg_new_conditional:Npnn \tl_if_empty:n #1 { p , TF , T , F }
{ \if:w \scan_stop: \tl_to_str:n {#1} \scan_stop: \prg_return_true: \else: \prg_return_false: \fi: }
\prg_generate_conditional_variant:Nnn \tl_if_empty:n
{ V } { p , TF , T , F }
\end{verbatim}

(End definition for $\tl_if_empty:nTF$. This function is documented on page 103.)

The auxiliary function $\__tl_if_empty_if:o$ is for use in various token list conditionals which reduce to testing if a given token list is empty after applying a simple function to it. The test for emptiness is based on $\tl_if_empty:nTF$, but the expansion is hard-coded for efficiency, as this auxiliary function is used in several places. We don’t put $\prg_return_true$: and so on in the definition of the auxiliary, because that would prevent an optimization applied to conditionals that end with this code. Also the
\_\_if\_empty\_if:o is expanded once in \_\_if\_empty:oTF for efficiency as well (and to reduce code doubling).

\cs_new:Npn \_\_if\_empty\_if:o \#1
{\if:w \scan_stop: \_\_kernel_tl_to_str:w \exp_after:wN \#1 \scan_stop: \}
\exp_args:Nno \use:n
{ \prg_new_conditional:Npnn \_\_if\_empty\_if:o \#1 { p , TF , T , F } }
{ \prg_return_true: \else: \prg_return_false: \fi: }

(End definition for \_\_if\_empty:oTF and \_\_if\_empty\_if:o. This function is documented on page 103.)

\_\_if\_blank_p:n
\_\_if\_blank_p:V
\_\_if\_blank_p:o
\_\_if\_blank:p:TF
\_\_if\_blank:VTF
\_\_if\_blank:oTF
\_\_if\_blank:p:NNw

\_\_if\_eq:NN
\_\_if\_eq:Nc
\_\_if\_eq:cN
\_\_if\_eq:cc

\prg_new_conditional:Npnn \_\_if\_eq:NN \#1#2 { p , T , F , TF }  
{ \if_meaning:w \#1 \#2 \prg_return_true: \else: \prg_return_false: \fi: }  
\prg_generate_conditional_variant:Nnn \_\_if\_eq:NN { Nc , c , cc } { p , TF , T , F }  

(End definition for \_\_if\_eq:NNTF. This function is documented on page 103.)

Temporary storage.
\l__tl_internal_a_tl
\l__tl_internal_b_tl
A simple store and compare routine.

\prg_new_protected_conditional:Npnn \tl_if_eq:Nn #1#2 { T , F , TF }
{ \exp_after:wN \group_begin: \tl_set:Nn \l__tl_internal_b_tl {#2} \group_end: \if_meaning:w #1 \l__tl_internal_b_tl \prg_return_true: \else: \prg_return_false: \fi: \}

\prg_generate_conditional_variant:Nnn \tl_if_eq:Nn { c } { TF , T , F }

A simple store and compare routine.

\prg_new_protected_conditional:Npnn \tl_if_eq:nn #1#2 { T , F , TF }
{ \group_begin: \tl_set:Nn \l__tl_internal_a_tl {#1} \tl_set:Nn \l__tl_internal_b_tl {#2} \exp_after:wN \group_end: \if_meaning:w \l__tl_internal_a_tl \l__tl_internal_b_tl \prg_return_true: \else: \prg_return_false: \fi: \}

\prg_generate_conditional_variant:Nnn \tl_if_eq:nn { c } { TF , T , F }

A simple store and compare routine.

\cs_new_protected:Npn \tl_if_in:NnT { \exp_args:No \tl_if_in:nnT }
\cs_new_protected:Npn \tl_if_in:NnF { \exp_args:No \tl_if_in:nnF }
\cs_new_protected:Npn \tl_if_in:NnTF { \exp_args:No \tl_if_in:nnTF }
\prg_generate_conditional_variant:Nnn \tl_if_in:Nn { c } { T , F , TF }

See \tl_if_in:nnTF for further comments. Here we simply expand the token list variable and pass it to \tl_if_in:nnTF.

\cs_new_protected:Npm \tl_if_in:NnT { \exp_args:No \tl_if_in:nnT }
\cs_new_protected:Npm \tl_if_in:NnF { \exp_args:No \tl_if_in:nnF }
\cs_new_protected:Npm \tl_if_in:Nnn \tl_if_in:nnTF \{ \exp_args:No \tl_if_in:nnTF \}
\prg_generate_conditional_variant:Nnn \tl_if_in:Nn \tl_if_in:nn
{ c } { T , F , TF }

Once more, the test relies on the emptiness test for robustness. The function \_\tl_\_tmp:w removes tokens until the first occurrence of #2. If this does not appear in #1, then the final #2 is removed, leaving an empty token list. Otherwise some tokens remain, and the test is false. See \tl_if_empty:nTF for details on the emptiness test.

Treating correctly cases like \tl_if_in:nnTF \{ a state \} \{ states \}, where \#1\#2 contains \#2 before the end, requires special care. To cater for this case, we insert {}{} between the two token lists. This marker may not appear in \#2 because of \TeX\ limitations on what can delimit a parameter, hence we are safe. Using two brace groups makes the
test work also for empty arguments. The `\if_false:` constructions are a faster way to do `\group_align_safe_begin:` and `\group_align_safe_end:`. The `\scan_stop:` ensures that f-expanding `\tl_if_in:nnTF` does not lead to unbalanced braces.

```
\prg_new_protected_conditional:Npnn \tl_if_in:nn #1#2 { T , F , TF }
{ \scan_stop: \if_false: { \fi: \cs_set:Npn \__tl_tmp:w ##1 #2 { } \tl_if_empty:oTF { \__tl_tmp:w #1 } \if_false: \fi: } \if_false: } \fi: 
\prg_generate_conditional_variant:Nnn \tl_if_in:nn { V , o , no } { T , F , TF }
```

(End definition for \tl_if_in:nnTF. This function is documented on page 104.)

```
\tl_if_novalue_p:n \tl_if_novalue:nTF \__tl_if_novalue:w Tests whether ##1 matches -NoValue- exactly (with suitable catcodes): this is similar to `\quark_if_nil:nTF`. The first argument of `\__tl_if_novalue:w` is empty if and only if ##1 starts with -NoValue-, while the second argument is empty if ##1 is exactly -NoValue- or if it has a question mark just following -NoValue-. In this second case, however, the material after the first ?! remains and makes the emptiness test return false.
```

```
\cs_set_protected:Npn \__tl_if_novalue:w ##1 #1 ##2 ? ##3 ? ! { ##1 ##2 }
```

(End definition for \tl_if_novalue:nTF and \__tl_if_novalue:w. This function is documented on page 104.)

```
\tl_if_single_p:N \tl_if_single:nTF Expand the token list and feed it to \tl_if_single:nTF.
```

```
\cs_new:Npn \tl_if_single_p:n { \exp_args:No \tl_if_single_p:n }
\cs_new:Npn \tl_if_single:NT { \exp_args:No \tl_if_single:nT }
\cs_new:Npn \tl_if_single:NF { \exp_args:No \tl_if_single:nF }
\cs_new:Npn \tl_if_single:NTF { \exp_args:No \tl_if_single:nTF }
```

(End definition for \tl_if_single:nTF. This function is documented on page 104.)

```
\tl_if_single_p:n \tl_if_single:nTF \__tl_if_single:nnw This test is similar to \tl_if_empty:nTF. Expanding `\use_none:nn #1 ??` once yields an empty result if #1 is blank, a single ? if #1 has a single item, and otherwise yields some tokens ending with ?? Then, `\__kernel_tl_to_str:w` makes sure there are no odd category codes. An earlier version would compare the result to a single ? using string comparison, but the Lua call is slow in LuaTeX. Instead, `\__tl_if_single:nnw` picks
```

676
the second token in front of it. If \#1 is empty, this token is the trailing $?$. If \#1 has a single item, the token is $\text{\_\_kernel_tl_to_str:}$ and the $\text{\_\_kernel_tl_to_str:}$ test yields $\text{true}$. Otherwise, it is one of the characters resulting from $\text{\_\_kernel_tl_to_str:}$ and the $\text{\_\_kernel_tl_to_str:}$ test yields $\text{false}$. Note that $\text{\_\_kernel_tl_to_str:}$ and $\text{\_\_kernel_tl_to_str:}$ are primitives that take care of expansion.

\begin{verbatim}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\end{verbatim}

There are four cases: empty token list, token list starting with a normal token, with a brace group, or with a space token. If the token list starts with a normal token, remove it and check for emptiness. For the next case, an empty token list is not a single token. Finally, we have a non-empty token list starting with a space or a brace group. Applying \text{f-expansion} yields an empty result if and only if the token list is a single space.

\begin{verbatim}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\end{verbatim}

The aim here is to allow the case statement to be evaluated using a known number of expansion steps (two), and without needing to use an explicit “end of recursion” marker. That is achieved by using the test input as the final case, as this is always true. The trick is then to tidy up the output such that the appropriate case code plus either the $\text{true}$ or $\text{false}$ branch code is inserted.

\begin{verbatim}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\text{\_\_kernel_tl_to_str:}
\end{verbatim}
To tidy up the recursion, there are two outcomes. If there was a hit to one of the cases searched for, then \texttt{#1} is the code to insert, \texttt{#2} is the next case to check on and \texttt{#3} is all of the rest of the cases code. That means that \texttt{#4} is the true branch code, and \texttt{#5} tidies up the spare \texttt{\_\_tl_mark} and the false branch. On the other hand, if none of the cases matched then we arrive here using the “termination” case of comparing the search with itself. That means that \texttt{#1} is empty, \texttt{#2} is the first \texttt{\_\_tl_mark} and so \texttt{#4} is the false code (the true code is mopped up by \texttt{#3}).

\begin{verbatim}
\cs_new:Npn \tl_case:NnTF \__tl_case:NnF \tl_case:NnTF
\cs_generate_variant:Nn \tl_case:Nn { \c }
\prg_generate_conditional_variant:Nnn \tl_case:Nn { \c } { \texttt{TF} }
\end{verbatim}

(End definition for \texttt{tl_case:NnTF} and others. This function is documented on page 105.)

### 52.8 Mapping over token lists

Expandable loop macro for token lists. We use the internal scan mark \texttt{\_\_tl_stop} (defined later), which is not allowed to show up in the token list \texttt{#1} since it is internal to \texttt{l3tl}. This allows us a very fast test of whether some \texttt{(item)} is the end-marker \texttt{\_\_tl_stop}, namely call \texttt{\_\_tl_use_none_delimit_by_s_stop:w \_\_tl_case_end:nw \texttt{(item)} \texttt{(function)} \texttt{\_\_tl_stop}}, which calls \texttt{(function)} if the \texttt{(item)} is the end-marker. To speed up the loop even more, only test one out of eight items, and once we hit one of the eight end-markers, go more slowly through the last few items of the list using \texttt{\_\_tl_map_function_end:w}.

\begin{verbatim}
\__tl_map_function:nN \_\_tl_map_function:NN \_\_tl_map_function:cN
\__tl_map_function:Nn \__tl_map_function:Nnn \texttt{\_\_tl_map_function_end:w \_\_tl_use_none_delimit_by_s_stop:w}
\end{verbatim}

\texttt{\_\_tl_map_function:nN} #1 #2 #3 \texttt{\_\_tl_map_function:NN} #1 #2 #3
\texttt{\_\_tl_map_function:cN} #1 #2 #3 \texttt{\_\_tl_map_function:Nn} #1 #2 #3
\texttt{\_\_tl_map_function:Nnn} #1 #2 #3 \texttt{\_\_tl_map_function_end:w} #1 #2 #3
\texttt{\_\_tl_use_none_delimit_by_s_stop:w} #1 #2 #3

678
The inline functions are straightforward by now. We use a little trick with the counter \g__kernel_prg_map_int to make them nestable. We can also make use of \_\_\_tl_map_function:NNnnnnnn from before.

\cs_new_protected:Npn \tl_map_inline:nn #1#2
\int_gincr:N \g__kernel_prg_map_int
\cs_gset_protected:cpn
{ \_\_\_tl_map_ \int_use:N \g__kernel_prg_map_int :w } ##1 {#2}
\exp_args:Nc \_\_\_tl_map_function:NNnnnnnn
#1
\s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop
\prg_break_point:Nn \tl_map_break:
{ \int_gdecr:N \g__kernel_prg_map_int }

(End definition for \_\_\_tl_map_function:nn and others. These functions are documented on page 109.)

\cs_new_protected:Npn \tl_map_inline:Nn
\cs_gset_protected:cpn
\exp_args:Nc \_\_\_tl_map_function:NNnnnnnn
{ \_\_\_tl_map_ \int_use:N \g__kernel_prg_map_int :w } ##1 {#2}
\exp_args:No \tl_map_inline:nn
\cs_generate_variant:Nn \tl_map_inline:Nn { c }

(End definition for \_\_\_tl_map_function:nn and \_\_\_tl_map_function:Nn. These functions are documented on page 109.)

Much like the function mapping.

\cs_new:Npn \tl_map_tokens:nn #1#2
\_\_\_tl_token:nnnnnnnnn {#2} #1
\_\_\_tl_token:nnnnnnnnn
\_\_\_tl_token:nnnnnnnnn
\prg_break_point:Nn \tl_map_break:
{ }

(End definition for \_\_\_tl_token:nnnnnnnnn and \_\_\_tl_token:nnnnnnnnn. These functions are documented on page 109.)
\tl_map_variable:nNn
\tl_map_variable:NNN
\tl_map_variable:cNn
\__tl_map_variable:Nnn
\cs_new_protected:Npn \tl_map_variable:nNn #1#2#3
\begin{verbatim}
{ \tl_map_tokens:nn {#1} { \__tl_map_variable:Nnn #2 {#3} } }
\end{verbatim}
\cs_new_protected:Npn \tl_map_variable:NNn
{ \exp_args:No \tl_map_variable:nNn }
\cs_generate_variant:Nn \tl_map_variable:NNn { c }
\begin{verbatim}
\end{verbatim}
(End definition for \tl_map_tokens:nn and others. These functions are documented on page 110.)
\tl_map_break:
\tl_map_break:n
The break statements use the general \prg_map_break:Nn.
\begin{verbatim}
\end{verbatim}
(End definition for \tl_map_break: and \tl_map_break:n. These functions are documented on page 110.)

52.9 Using token lists
\tl_to_str:n
\tl_to_str:V
Another name for a primitive: defined in l3basics.
\begin{verbatim}
\end{verbatim}
(End definition for \tl_to_str:n. This function is documented on page 106.)
\tl_to_str:N
\tl_to_str:c
These functions return the replacement text of a token list as a string.
\begin{verbatim}
\end{verbatim}
(End definition for \tl_to_str:N. This function is documented on page 106.)
Token lists which are simply not defined give a clear \texttt{TeX} error here. No such luck for ones equal to \texttt{scan_stop}: so instead a test is made and if there is an issue an error is forced.

\begin{verbatim}
\cs_new:Npn \tl_use:N #1
\tl_if_exist:NTF #1 {#1}
\msg_expandable_error:nnn { kernel } { bad-variable } {#1}
\end{verbatim}

(End definition for \texttt{\tl_use:N}. This function is documented on page \pageref{page107}.)

\section{Working with the contents of token lists}

\begin{verbatim}
\cs_new:Npn \tl_count:n #1
\int_eval:n { 0 \tl_map_function:nN {#1} \__tl_count:n }
\cs_new:Npn \tl_count:N #1
\int_eval:n { 0 \tl_map_function:NN #1 \__tl_count:n }
\cs_new:Npn \__tl_count:n #1 { + 1 }
\cs_generate_variant:Nn \tl_count:n { V , o }
\cs_generate_variant:Nn \tl_count:N { c }
\end{verbatim}

(End definition for \texttt{\tl_count:n}, \texttt{\tl_count:N}, and \texttt{\__tl_count:n}. These functions are documented on page \pageref{page107}.)

\begin{verbatim}
\cs_new:Npn \tl_count_tokens:n #1
\int_eval:n
\tl_map_function:NN {#1} \__tl_act_count_normal:n
\__tl_act_count_group:nn
\__tl_act_count_space:n
\end{verbatim}

The token count is computed through an \texttt{\int_eval:n} construction. Each \texttt{1+} is output to the \emph{left}, into the integer expression, and the sum is ended by the \texttt{\exp_end:} inserted by \texttt{\__tl_act_end:wn} (which is technically implemented as \texttt{\c_zero_int}). Somewhat a hack!

\begin{verbatim}
\cs_new:Npn \tl_count_tokens:n #1
\int_eval:n
\tl_map_function:NN {#1} \__tl_act_count_normal:n
\__tl_act_count_group:nn
\__tl_act_count_space:n
\end{verbatim}
Reversal of a token list is done by taking one item at a time and putting it after \s__-\tl_stop.

Trimming spaces from around the input is deferred to an internal function whose first argument is the token list to trim, augmented by an initial \__tl_trim_mark:, and whose second argument is a ⟨continuation⟩, which receives as a braced argument \__tl_trim_mark: ⟨trimmed token list⟩. The control sequence \__tl_trim_mark: expands to nothing in a single expansion. In the case at hand, we take \__kernel_exp_not:w \exp_after:wN as our continuation, so that space trimming behaves correctly within an x-type expansion.
end of the token list: then ##1 is the token list and ##3 is \__tl_trim_spaces_auxii:w. This hands the relevant tokens to the loop \__tl_trim_spaces_auxiii:w, responsible for trimming trailing spaces. The end is reached when \s__tl_nil matches the one present in the definition of \tl_trim_spaces:n. Then \__tl_trim_spaces_auxiv:w puts the token list into a group, with a lingering \__tl_trim_mark: at the start (which will expand to nothing in one step of expansion), and feeds this to the (continuation).

\cs_set_protected:Npn \__tl_tmp:w #1
\cs_new:Npn \__tl_trim_spaces:nn ##1
\__tl_trim_spaces_auxi:w ##1 \s__tl_nil
\__tl_trim_mark: #1 { }
\__tl_trim_spaces_auxi:w \__tl_trim_spaces_auxiii:w
\__tl_nil #1 \s__tl_nil
\__tl_trim_spaces_auxiv:w
\s__tl_stop
\cs_new:Npn \__tl_trim_spaces_auxi:w ##1 \__tl_trim_mark: #1 ##2 \__tl_trim_mark: ##3
\__tl_trim_spaces_auxi:w \__tl_trim_mark: ##2
\__tl_trim_mark: #1 {##1}
\__tl_trim_spaces_auxiii:w ##1 #1 \s__tl_nil ##2
\__tl_trim_spaces_auxiii:w #1
\cs_new:Npn \__tl_trim_spaces_auxii:w
\__tl_trim_spaces_auxi:w \__tl_trim_mark: \__tl_trim_mark: ##1
\__tl_trim_spaces_auxiii:w
\__tl_nil #1
\cs_new:Npn \__tl_trim_spaces_auxiii:w ##1 #1 \s__tl_nil ##2
\__tl_trim_spaces_auxii:w
\__tl_nil #1 \s__tl_nil
\__tl_trim_spaces_auxiii:w
\cs_new:Npn \__tl_trim_spaces_auxiv:w #1 \s__tl_nil ##2 \s__tl_stop #3
\cs_new:Npn \__tl_trim_spaces_auxiv:w { ##3 { ##1 } }
\cs_new:Npn \__tl_trim_spaces_auxiv:w {}
\__tl_tmp:w { ~ }

(End definition for \tl_trim_spaces:n and others. These functions are documented on page 108.)

\tl_sort:Nn
\tl_sort:cn
\tl_gsort:Nn
\tl_gsort:cn
\tl_sort:nN

Implemented in l3sort.

(End definition for \tl_sort:Nn, \tl_gsort:Nn, and \tl_sort:nN. These functions are documented on page 115.)
52.11 The first token from a token list

Finding the head of a token list expandably always strips braces, which is fine as this is consistent with for example mapping over a list. The empty brace groups in \texttt{\tl_head:n} ensure that a blank argument gives an empty result. The result is returned within the \texttt{\unexpanded} primitive. The approach here is to use \texttt{\if\texttt{false} : to allow us to use } as the closing delimiter: this is the only safe choice, as any other token would not be able to parse it’s own code. If the \texttt{\expanded} primitive is available it is used to get a fast and safe code variant in which we don’t have to ensure that the left-most token is an internal to not break in an f-type expansion. If \texttt{\expanded} isn’t available, using a marker, we can see if what we are grabbing is exactly the marker, or there is anything else to deal with. If there is, there is a loop. If not, tidy up and leave the item in the output stream. More detail in \url{http://tex.stackexchange.com/a/70168}.

\begin{verbatim}
\cs_if_exist:NTF \tex_expanded:D {
  \cs_new:Npn \tl_head:n #1 {
    \__kernel_exp_not:w \tex_expanded:D { { \if\texttt{false} : { \fi: \__tl_head_auxi:nw #1 { } } } }
  }
\cs_new:Npn \__tl_head_auxi:nw #1#2 \s__tl_stop {
  \exp_after:wN \__tl_head_auxii:n \exp_after:wN { { \if\texttt{false} : { \fi: } } }
}
}
\cs_new:Npn \tl_head:w #1#2 \q_stop {#1}
\cs_new:Npn \__tl_tl_head:w #1#2 \s__tl_stop {#1}
\cs_new:Npn \tl_head:N { \exp_args:No \tl_head:n }
\cs_generate_variant:Nn \tl_head:n { V , v , f }
\end{verbatim}

To correctly leave the tail of a token list, it’s important not to absorb any of the tail part
as an argument. For example, the simple definition

\cs_new:Npn \tl_tail:n #1 { \tl_tail:w #1 \q_stop }
\cs_new:Npn \tl_tail:w #1#2 \q_stop

would give the wrong result for \tl_tail:n { a { bc } } (the braces would be stripped). Thus the only safe way to proceed is to first check that there is an item to grab (i.e. that the argument is not blank) and assuming there is to dispose of the first item. As with \tl_head:n, the result is protected from further expansion by \unexpanded. While we could optimise the test here, this would leave some tokens “banned” in the input, which we do not have with this definition.

\exp_args:Nno \use:n { \cs_new:Npn \tl_tail:n #1 }
\exp_after:wN \__kernel_exp_not:w
\tl_if_blank:nTF {#1}
{ { } }
{ \exp_after:wN { \use_none:n #1 } }
\cs_generate_variant:Nn \tl_tail:n { V , v , f }
\cs_new:Npn \tl_tail:N { \exp_args:No \tl_tail:n }

(End definition for \tl_head:N and others. These functions are documented on page 111.)

Accessing the first token of a token list is tricky in three cases: when it has category code 1 (begin-group token), when it is an explicit space, with category code 10 and character code 32, or when the token list is empty (obviously).

Forgetting temporarily about this issue we would use the following test in \tl_if_head_eq_charcode:nN. Here, \tl_head:w yields the first token of the token list, then passed to \exp_not:N.

\if_charcode:w
\exp_after:wN \exp_not:N \tl_head:w #1 \q_nil \q_stop
\exp_not:N #2
\prg_return_true:
\else:
\prg_return_false:
\fi:

The two first special cases are detected by testing if the token list starts with an N-type token (the extra ? sends empty token lists to the true branch of this test). In those cases, the first token is a character, and since we only care about its character code, we can use \str_head:n to access it (this works even if it is a space character). An empty argument results in \tl_head:w leaving two token: ^ and \__tl_if_head_eq_empty_arg:w which will result in the \if_charcode:w test being false and remove \exp_not:N and #2.

\prg_new_conditional:Npnn \tl_if_head_eq_charcode:nN \tl_if_head_is_N_type:nTF \__tl_if_head_exp_not:w \__tl_if_head_eq_empty_arg:w
\__tl_if_head_eq_meaning_p:nN
\__tl_if_head_eq_meaning:nN
\__tl_if_head_eq_charcode_p:nN
\__tl_if_head_eq_catcode_p:nN
\__tl_if_head_eq_catcode_p:oN
\__tl_if_head_exp_not:w
\__tl_if_head_eq_empty_arg:w
\__tl_if_head_eq_meaning_p:nN
\__tl_if_head_eq_meaning:nN
\__tl_if_head_eq_catcode_p:nN
\__tl_if_head_eq_catcode_p:oN
\__tl_if_head_eq_catcode_p:nN
\__tl_if_head_eq_catcode_p:oN
\str_head:n
\exp_not:N
\prg_return_true:
\else:
\prg_return_false:
\fi:

\prg_generate_conditional_variant:Nnn \tl_if_head_eq_charcode:nN \tl_if_head_is_N_type:nTF \__tl_if_head_exp_not:w \__tl_if_head_eq_empty_arg:w
For `\tl_if_head_eq_catcode:nN`, again we detect special cases with a `\tl_if_head_is_N_type:n`. Then we need to test if the first token is a begin-group token or an explicit space token, and produce the relevant token, either `\c_group_begin_token` or `\c_space_token`. Again, for an empty argument, a hack is used, removing the token given by the user and leaving two tokens in the input stream which will make the `\if_catcode:w` test return false.

```
\prg_new_conditional:Npn \tl_if_head_eq_catcode:nN #1 #2 { p , T , F , TF }
\prg_generate_conditional_variant:Nnn \tl_if_head_eq_catcode:nN { o } { p , TF , T , F }
```

For `\tl_if_head_eq_meaning:nN`, again, detect special cases. In the normal case, use `\tl_head:w` with no `\exp_not:N` this time, since `\if_meaning:w` causes no expansion. With an empty argument, the test is `true`, and `\use_none:nnn` removes `#2` and `\prg_return_true:` and `\else:` (it is safe this way here as in this case `\prg_new_conditional:Npn` didn’t optimize these two away). In the special cases, we know that the first token is a character, hence `\if_charcode:w` and `\if_catcode:w` together are enough. We combine them in some order, hopefully faster than the reverse. Tests are not nested because the arguments may contain unmatched primitive conditionals.

```
\prg_new_conditional:Npn \tl_if_head_eq_meaning:nN #1#2 { p , T , F , TF }
\cs_new:Npn \__tl_if_head_eq_meaning_normal:nN { o } { p , TF , T , F }
```

686
Both \tl_if_head_eq_charcode:nN and \tl_if_head_eq_catcode:nN will need to get the first token of their argument and apply \exp_not:N to it. \__tl_head_exp_not:w does exactly that.

If the argument of \tl_if_head_eq_charcode:nN and \tl_if_head_eq_catcode:nN was empty \__tl_if_head_eq_empty_arg:w will be left in the input stream. This macro has to remove \exp_not:N and the following token from the input stream to make sure no unbalanced if-construct is created and leave tokens there which make the two tests return false.

(End definition for \tl_if_head_eq_meaning:nNTF and others. These functions are documented on page 105.)

A token list can be empty, can start with an explicit space character (catcode 10 and charcode 32), can start with a begin-group token (catcode 1), or start with an N-type argument. In the first two cases, the line involving \__tl_if_head_is_N_type_auxi:w produces f (and otherwise nothing). In the third case (begin-group token), the lines involving \token_to_str:N produce a single closing brace. The category code test is thus true exactly in the fourth case, which is what we want. One cannot optimize by moving one of the \scan_stop: to the beginning: if #1 contains primitive conditionals, all of its occurrences must be dealt with before the \if:w tries to skip the true branch of the conditional.
Pass the first token of \#1 through \token_to_str:N, then check for the brace balance. The extra \? caters for an empty argument. This could be made faster, but we need all brace tricks to happen in one step of expansion, keeping the token list brace balanced at all times.

The auxiliary’s argument is all that is before the first explicit space in \prg_do_nothing:#1?. If that is a single \prg_do_nothing: the test yields true. Otherwise, that is more than one token, and the test yields false. The work is done within braces (with an \if_false: \fi: \if_true: \else: \fi: \prg_return_true: \prg_return_false: construction) both to hide potential alignment tab characters from \TeX{} in a table, and to allow for removing what remains of the token list after its first space. The use of \if:w ensures that the result of a single step of expansion directly yields a balanced token list (no trailing closing brace).

(End definition for \tl_if_head_is_group:nTF and others. This function is documented on page 106.)
52.12 Token by token changes

The \texttt{\_\_tl\_act\_...} functions may be applied to any token list. Hence, we use a private quark, to allow any token, even quarks, in the token list. Only \texttt{\_\_tl\_act\_stop} may not appear in the token lists manipulated by \texttt{\_\_tl\_act\_NNNn} functions.

\begin{verbatim}
\exp_args:Nnx \use:n { \cs_new:Npn \__tl_act_loop:w #1 \s__tl_act_stop }
\exp_not:o { \__tl_act_if_head_is_space:nTF {#1} }
\exp_not:N \__tl_act_space:wwNNN
\exp_not:N \__tl_act_normal:NwNNN
\exp_not:N \__tl_act_group:nwNNN
\exp_not:N \__tl_act_end:wn
\exp_not:o { \tl_if_head_is_group:nTF {#1} }
\exp_not:N \__tl_act_group:nwNNN
\exp_not:N \__tl_act_normal:NwNNN
\exp_after:wN \use_none:n \exp_after:wN { \if_false: } \fi:
\}
\end{verbatim}

(End definition for \texttt{\_\_tl\_act\_stop}.)

To help control the expansion, \texttt{\_\_tl\_act\_NNNn} should always be preceeded by \texttt{\exp:w} and ends by producing \texttt{\exp_end:} once the result has been obtained. This way no internal token of it can be accidentally end up in the input stream. Because \texttt{\_\_tl\_act\_stop} can’t appear without braces around it in the argument \#1 of \texttt{\_\_tl\_act\_loop:w}, we can use this marker to set up a fast test for leading spaces.

\begin{verbatim}
\cs_set_protected:Npn \__tl_tmp:w #1 { ~ } \end{verbatim}

(We expand the definition \texttt{\_\_tl\_act\_if\_head\_is\_space:nTF} when setting up \texttt{\_\_tl\_act\_loop:w}, so we can then undefine the auxiliary.) In the loop, we check how the token list begins and act accordingly. In the “group” case, we may have reached \texttt{\_\_tl\_act\_stop}, the end of the list. Then leave \texttt{\exp_end:} and the result in the input stream, to terminate the expansion of \texttt{\exp:w}. Otherwise, apply the relevant function to the “arguments”, \#3 and to the head of the token list. Then repeat the loop. The scheme is the same if the token list starts with an N-type or with a space, making sure that \texttt{\_\_tl\_act\_space:wwNNN} gobbles the space.
\exp_not:n {#1} \s__tl_act_stop
\cs_undefine:N \__tl_act_if_head_is_space:nTF
\cs_new:Npn \__tl_act_normal:NwNNN #1 #2 \s__tl_act_stop #3
{ #3 #1 \__tl_act_loop:w #2 \s__tl_act_stop #3
#3 }
\cs_new:Npn \__tl_use_none_delimit_by_s_act_stop:w #1 \s__tl_act_stop { }
\cs_new:Npn \__tl_act_end:wn #1 \__tl_act_result:n #2
{ \group_align_safe_end: \exp_end: #2 }
\cs_new:Npn \__tl_act_group:nwNNN #1 #2 \s__tl_act_stop #3#4#5
{ \__tl_use_none_delimit_by_s_act_stop:w #1 \__tl_act_end:wn \s__tl_act_stop
#5 {#1} \__tl_act_loop:w #2 \s__tl_act_stop #3 #4 #5
}
\exp_last_unbraced:NNo
\cs_new:Npn \__tl_act_space:wwNNN \c_space_tl #1 \s__tl_act_stop #2#3
{ #3 \__tl_act_loop:w #1 \s__tl_act_stop #2 #3 }
\__tl_act:NNNn loops over tokens, groups, and spaces in #4. \s_@@_act_stop serves as the end of token list marker, the ? after it avoids losing outer braces. The result is stored as an argument for the dummy function \__tl_act_result:n.
\cs_new:Npn \__tl_act:NNNn #1#2#3#4
{ \group_align_safe_begin:
\__tl_act_loop:w #4 { \s__tl_act_stop } ? \s__tl_act_stop #1 #3 #2
\__tl_act_result:n { }
}
Typically, the output is done to the right of what was already output, using \__tl_act_output:n, but for the \__tl_act_reverse functions, it should be done to the left.
\cs_new:Npn \__tl_act_output:n #1 #2 \__tl_act_result:n #3
{ #2 \__tl_act_result:n { #3 #1 } }
\cs_new:Npn \__tl_act_reverse_output:n #1 #2 \__tl_act_result:n #3
{ #2 \__tl_act_result:n { #1 #3 } }
(End definition for \__tl_act:NNNn and others.)
\tl_reverse:n \tl_reverse:o \tl_reverse:V \__tl_reverse_normal:nN \__tl_reverse_group_preserve:mm \__tl_reverse_space:n
The goal here is to reverse without losing spaces nor braces. This is done using the general internal function \__tl_act:NNNn. Spaces and “normal” tokens are output on the left of the current output. Grouped tokens are output to the left but without any reversal within the group.
\begin{verbatim}
   \exp:w
   \__tl_act:NNNn
   \__tl_reverse_normal:N
   \__tl_reverse_group_preserve:n
   \__tl_reverse_space:
   {#1}
   }
  }
 \cs_generate_variant:Nn \tl_reverse:n { o , V }
 \cs_new:Npn \__tl_reverse_normal:N { \__tl_act_reverse_output:n }
 \cs_new:Npn \__tl_reverse_group_preserve:n #1 { \__tl_act_reverse_output:n { {#1} } }
 \cs_new:Npn \__tl_reverse_space: { \__tl_act_reverse_output:n { ~ } }
 \end{verbatim}

(End definition for \tl_reverse:n and others. This function is documented on page 107.)

\textbf{\tl_reverse:N} This reverses the list, leaving \exp_stop_f: in front, which stops the f-expansion.

\textbf{\tl_greverse:N} This reverses the list, leaving \exp_stop_f: in front, which stops the f-expansion.

\textbf{52.13 Using a single item}

The idea here is to find the offset of the item from the left, then use a loop to grab the correct item. If the resulting offset is too large, then \__tl_if_recursion_tail_break:nN terminates the loop, and returns nothing at all.

\begin{verbatim}
 \cs_new:Npn \tl_item:nn #1#2
 { \exp_args:Nf \__tl_item:nn { \exp_args:Nf \__tl_item_aux:nn { \int_eval:n {#2} } {#1} } #1
 \q__tl_recursion_tail
 \prg_break_point:
 }
 \cs_new:Npn \__tl_item_aux:nn #1#2
 { \int_compare:nNnTF {#1} < 0
   { \int_eval:n { \tl_count:n {#2} + 1 + #1 } }
   {#1}
 }
 \cs_new:Npn \__tl_item:nn #1#2
 { \__tl_if_recursion_tail_break:nN {#2} \prg_break:
 \int_compare:nNnTF {#1} = 1
 \int_eval:n { \tl_count:n {#2} + 1 + #1 } }
 \end{verbatim}

(End definition for \tl_reverse:N and \tl_greverse:N. These functions are documented on page 107.)
\exp_args:Nf \__tl_item:nn { \int_eval:n { #1 - 1 } }
\end{definition_for \tl_item:nn and others. These functions are documented on page 112.}

Importantly \tl_item:nn only evaluates its argument once.
\end{definition_for \tl_item:nn and \tl_item:N. These functions are documented on page 113.}

To avoid checking for the end of the token list at every step, start by counting the number \l of items and “normalizing” the bounds, namely clamping them to the interval \([0, l]\) and dealing with negative indices. More precisely, \__tl_range_items:nnNn receives the number of items to skip at the beginning of the token list, the index of the last item to keep, a function which is either \__tl_range:w or the token list itself. If nothing should be kept, leave \{\}: this stops the f-expansion of \tl_head:f and that function produces an empty result. Otherwise, repeatedly call \__tl_range_skip:w to delete \#1 items from the input stream (the extra brace group avoids an off-by-one shift). For the braced version \__tl_range_braced:w sets up \__tl_range_collect_braced:w which stores items one by one in an argument after the semicolon. Depending on the first token of the tail, either just move it (if it is a space) or also decrement the number of items left to find. Eventually, the result is a brace group followed by the rest of the token list, and \tl_head:f cleans up and gives the result in \exp_not:n.
This function converts an \textit{index} argument into an explicit position in the token list (a result of 0 denoting “out of bounds”). Expects two explicit integer arguments: the \textit{index} \#1 and the string count \#2. If \#1 is negative, replace it by \#1 + \#2 + 1, then limit to the range \([0, \#2]\).

\begin{verbatim}
\cs_new:Npn \__tl_range_normalize:nn \#1\#2
{
  \int_eval:n {
    \if_int_compare:w \#1 < \c_zero_int
    \if_int_compare:w \#1 < -\#2 \exp_stop_f:
      0
    \else:
      \#1 + \#2 + 1
    \fi:
  \else:
    \if_int_compare:w \#1 < \#2 \exp_stop_f:
      \#1
    \else:
      \#2
    \fi:
  \fi:
}
\end{verbatim}

52.14 Viewing token lists

Showing token list variables is done after checking that the variable is defined (see \__kernel_register_show:N).

\begin{verbatim}
\tl_show:N
\tl_show:c
\tl_log:N
\tl_log:c
\__tl_show:NN
\__kernel_chk_defined:NT
\end{verbatim}
The token list \_tl_internal_a_tl containing the result of all these manipulations is displayed to the terminal using \tex_showtokens:D and an odd \exp_after:wN which expand the closing brace to improve the output slightly. The calls to \_kernel_iow_with:Nn ensure that the \newlinechar is set to 10 so that the \iownewline: inserted by the line-wrapping code are correctly recognized by \TeX, and that \errorcontextlines is −1 to avoid printing irrelevant context.

\cs_new_protected:Npn \tl_show:n \l__tl_internal_a_tl \s__tl_stop {#2}

(End definition for \tl_show:n, \__tl_show:n, and \__tl_show:w. This function is documented on page 108.)

\tl_log:n Logging is much easier, simply line-wrap. The >- and trailing period is there to match the output of \tl_show:n. 

\cs_new_protected:Npn \tl_log:n \l__tl_internal_a_tl \s__tl_stop {#2}

(End definition for \tl_show:n, \__tl_show:n, and \__tl_show:w. This function is documented on page 108.)

\__kernel_chk_tl_type:NnnT Helper for checking that \#1 has the correct internal structure to be of a certain type. Make sure that it is defined and that it is a token list, namely a macro with no \long nor \protected prefix. Then compare \#1 to an attempt at reconstructing a valid structure of the given type using \#2 (see implementation of \seq_show:N for instance). If that is successful run the requested code \#4.

\cs_new_protected:Npn \__kernel_chk_tl_type:NnnT \#1\#2\#3\#4

(End definition for \tl_log:n. This function is documented on page 109.)
52.15 Internal scan marks
\s__tl_nil Internal scan marks. These are defined here at the end because the code for \scan_new:N depends on some \l3tl functions.
\s__tl_mark
\s__tl_stop
\scan_new:N \s__tl_nil
\scan_new:N \s__tl_mark
\scan_new:N \s__tl_stop

(End definition for \s__tl_nil, \s__tl_mark, and \s__tl_stop.)

52.16 Scratch token lists
\g_tmpa_tl Global temporary token list variables. They are supposed to be set and used immediately, with no delay between the definition and the use because you can’t count on other macros not to redefine them from under you.
\g_tmpb_tl
\l_tmpa_tl
\l_tmpb_tl
\scan_new:N \l_tmpa_tl
\scan_new:N \l_tmpb_tl

(End definition for \g_tmpa_tl and \g_tmpb_tl. These variables are documented on page 118.)
\l_tmpa_tl
\l_tmpb_tl
\l_new:N \l_tmpa_tl
\l_new:N \l_tmpb_tl

(End definition for \l_tmpa_tl and \l_tmpb_tl. These variables are documented on page 117.)

We finally clean up a temporary control sequence that we have used at various points to set up some definitions.
\cs_undefine:N \__tl_tmp:w
Chapter 53

\texttt{l3str implementation}

53.1 Internal auxiliaries

\texttt{\_\_str\_mark}
\texttt{\_\_str\_stop}

Internal scan marks.

\texttt{\_\_str\_use\_none\_delimit\_by\_s\_stop\_w #1 \_\_str\_stop}
\texttt{\_\_str\_use\_i\_delimit\_by\_s\_stop\_nw #1 #2 \_\_str\_stop}

(End definition for \_\_str\_mark and \_\_str\_stop.)

\texttt{\_\_str\_recursion\_tail}
\texttt{\_\_str\_recursion\_stop}

Internal recursion quarks.

\texttt{\_\_kernel\_quark\_new\_test:N \_\_str\_if\_recursion\_tail\_break:NN}
\texttt{\_\_kernel\_quark\_new\_test:N \_\_str\_if\_recursion\_tail\_stop\_do:Nn}

(End definition for \_\_str\_if\_recursion\_tail\_break:NN and \_\_str\_if\_recursion\_tail\_stop\_do:Nn.)
53.2 Creating and setting string variables

```
\str_new:N
\str_new:c
\str_use:N
\str_use:c
\str_clear:N
\str_clear:c
\str_gclear:N
\str_gclear:c
\str_clear_new:N
\str_clear_new:c
\str_gclear_new:N
\str_gclear_new:c
\str_set_eq:NN
\str_set_eq:cN
\str_set_eq:Nc
\str_set_eq:cc
\str_gset_eq:NN
\str_gset_eq:cN
\str_gset_eq:Nc
\str_gset_eq:cc
\str_concat:NNN
\str_concat:ccc
\str_gconcat:NNN
\str_gconcat:ccc
```

A string is simply a token list. The full mapping system isn’t set up yet so do things by hand.

```
\cs_set_protected:Npn \__str_tmp:n #1
  \tl_if_blank:nF {#1}
  \cs_new_eq:cc { str_ #1 :N } { tl_ #1 :N }
  \exp_args:Nc \cs_generate_variant:Nn { str_ #1 :N } { c }
\__str_tmp:n
\__str_tmp:n
\__str_tmp:n
\__str_tmp:n
\__str_tmp:n
```

```
\cs_set_protected:Npn \__str_tmp:n #1
  \tl_if_blank:nF {#1}
  \cs_new_protected:cpx { str_ #1 :Nn } #1##2
  \exp_not:c { tl_ #1 :Nx } #1
  \exp_not:N \tl_to_str:n {##2}
\__str_tmp:n
```

```
\group_end:
```

(End definition for \str_new:N and others. These functions are documented on page 120.)

```
\str_set:Nn
\str_set:NV
\str_set:Nx
\str_set:cn
\str_set:cV
\str_set:cV
\str_gset:Nn
\str_gset:NV
\str_gset:Nx
\str_gset:cn
\str_gset:cV
\str_gset:cx
\str_const:Nn
\str_const:NV
\str_const:Nx
\str_const:cn
\str_const:cV
\str_const:cx
\str_put_left:Nn
\str_put_left:NV
\str_put_left:Nx
\str_put_left:cn
\str_put_left:cV
\str_put_left:cx
```

 Simply convert the token list inputs to (strings).

```
\group_begin:
\str_set:NN
\str_set:NV
\str_set:Nx
\str_set:cn
\str_set:cx
\str_set:cV
\str_set:cV
\str_get:NN
\str_get:NV
\str_get:Nx
\str_get:cn
\str_get:cV
\str_get:cx
\str_const:NN
\str_const:NV
\str_const:Nx
\str_const:cn
\str_const:cV
\str_const:cx
\str_put_left:NN
\str_put_left:NV
\str_put_left:Nx
\str_put_left:cn
\str_put_left:cV
\str_put_left:cx
```

```
\group_end:
```

(End definition for \str_set:NN and others. These functions are documented on page 120.)

```
\group_begin:
```

698
53.3 Modifying string variables

Start by applying \texttt{\tl_to_str:n} to convert the old and new token lists to strings, and also apply \texttt{\tl_to_str:N} to avoid any issues if we are fed a token list variable. Then the code is a much simplified version of the token list code because neither the delimiter nor the replacement can contain macro parameters or braces. The delimiter \texttt{\s__str-mark} cannot appear in the string to edit so it is used in all cases. Some x-expansion is unnecessary. There is no need to avoid losing braces nor to protect against expansion. The ending code is much simplified and does not need to hide in braces.
53.4 String comparisons

More copy-paste!

Simply rely on \_\_str_if_eq:nn, which expands to -1, 0 or 1. The ee version is created directly because it is more efficient.

(End definition for \_\_str_if_eq:nn.)
\_\_str_if_eq:nn \{ \exp_not:n \{#1\} \} \{ \exp_not:n \{#3\} \}
#2 \c_zero_int
\prg_return_true: \else: \prg_return_false: \fi:
\prg_new_conditional:Npnn \str_compare:eNe #1#2#3 \{ p , T , F , TF \}
{ \if_int_compare:w \__str_if_eq:nn \exp_not:n \{#1\} \{\exp_not:n \{#3\} \} #2 \c_zero_int
\prg_return_true: \else: \prg_return_false: \fi:
\prg_generate_conditional_variant:Nnn \str_if_eq:nn
\prg_new_conditional:Npnn \str_if_eq:ee #1#2 \{ p , T , F , TF \}
{ \if:w 0 \__str_if_eq:nn \{#1\} \{#2\} 
\prg_return_true: \else: \prg_return_false: \fi:
\prg_generate_conditional_variant:Nnn \str_if_eq:nn
\prg_new_conditional:Npnn \str_if_in:Nn \{ T , F , TF \}
{ \use:x \{ \tl_if_in:nnTF \{ \tl_to_str:N \{#1\} \} \{ \tl_to_str:n \{#2\} \} \{ \prg_return_true: \} \{ \prg_return_false: \} \fi:
\prg_generate_conditional_variant:Nnn \str_if_in:Nn
\prg_new_protected_conditional:Npnn \str_if_in:Nn \{ T , F , TF \}
{ \use:x \{ \tl_if_in:nnTF \{ \tl_to_str:N \{#1\} \} \{ \tl_to_str:n \{#2\} \} \{ \prg_return_true: \} \{ \prg_return_false: \} \fi:
\prg_generate_conditional_variant:Nnn \str_if_in:Nn
\prg_new_protected_conditional:Npnn \str_if_in:Nn \{ T , F , TF \}
{ \use:x \{ \tl_if_in:nnTF \{ \tl_to_str:N \{#1\} \} \{ \tl_to_str:n \{#2\} \} \{ \prg_return_true: \} \{ \prg_return_false: \} \fi:
\end{document}
\prg_new_protected_conditional:Npnn \str_if_in:nn #1#2 \{ T , F , TF \}
\{ \use:x \}
\{ \tl_if_in:nnTF \{ \tl_to_str:n \{#1\} \} \{ \tl_to_str:n \{#2\} \} \{ \prg_return_true: \} \{ \prg_return_false: \} \}

(End definition for \str_if_in:NnTF and \str_if_in:nnTF. These functions are documented on page 122.)

\str_case:nn Much the same as \tl_case:nnTF here: just a change in the internal comparison.
\str_case:Vn
\str_case:nV
\str_case:on
\str_case:nv
\str_case:nn
\str_case:VV
\str_case:nnT
\str_case:nnF
\str_case:nnTF
\str_case:nnT
\str_case:nnF
\str_case:nnTF
\str_case:nnT
\str_case:nnF
\str_case_e:nn
\str_case:e:nn
\str_case:e:nnT
\str_case:e:nnF
\str_case_e:nnT
\str_case_e:nnF
\str_case_e:nnTF
\str_case_e:nnT
\str_case_e:nnF
\__str_case:nnTF
\__str_case:nw
\__str_case_e:nn
\cs_new:Npn \str_case:nn #1#2
\cs_new:Npn \str_case:nnT #1#2#3
\cs_new:Npn \str_case:nnF #1#2
\cs_new:Npn \str_case:nnTF #1#2
\cs_new:Npn \__str_case:nnTF #1#2#4#5
\cs_generate_variant:Nn \str_case:nn { V , o , nV , nv }
\prg_generate_conditional_variant:Nnn \str_case:nn { V , o , nV , nv }
53.5 Mapping over strings

The inline and variable mappings are similar to the usual token list mappings but start out by turning the argument to an “other string”. Doing the same for the expandable function mapping would require \_\_kernel_str_to_other:n, quadratic in the string length. To deal with spaces in that case, \_\_str_map_function:w replaces the following space by a braced space and a further call to itself. These are received by \_\_str_map_function:nn, which passes the space to \#1 and calls \_\_str_map_function:w to deal with the next space. The space before the braced space allows to optimize the \q__str_recursion_tail test. Of course we need to include a trailing space (the question mark is needed to avoid losing the space when \TeX tokenizes the line). At the cost of about three more auxiliaries this code could get a 9 times speed up by testing only every 9-th character for whether it is \q__str_recursion_tail (also by converting 9 spaces at a time in the \str_map_function:nN case).

For the map_variable functions we use a string assignment to store each character because spaces are made catcode 12 before the loop.
\str_map_tokens:NN \str_map_tokens:cn \str_map_tokens:nn

Uses an auxiliary of \str_map_function:NN.

\cs_new:Npn \str_map_tokens:nn #1#2
\exp_args:Nno \use:nn
{ \__str_map_function:w \__str_map_function:nn {#2} }
{ \__kernel_tl_to_str:w {#1} }
\q__str_recursion_tail ? ~
\prg_break_point:Nn \str_map_break: { }
}
\cs_new:Npn \str_map_tokens:Nn { \exp_args:No \str_map_tokens:nn }
\cs_generate_variant:Nn \str_map_tokens:Nn { c }

(End definition for \str_map_tokens:nn and others. These functions are documented on page 123.)

\__kernel_str_to_other:n \__str_to_other_loop:w \__str_to_other_end:w

53.6 Accessing specific characters in a string

First apply \tl_to_str:n, then replace all spaces by “other” spaces, 8 at a time, storing the converted part of the string between the \s__str_mark and \s__str_stop markers. The end is detected when \__str_to_other_loop:w finds one of the trailing A, distinguished from any contents of the initial token list by their category. Then \__str_to_other_end:w is called, and finds the result between \s__str_mark and the first A (well, there is also the need to remove a space).

\cs_new:Npn \__kernel_str_to_other:n #1
\exp_after:wN \__str_to_other_loop:w
\tl_to_str:n (#1) \* A \* A \* A \* A \* A \* A \* A \* A \* \s__str_mark \s__str_stop
\group_begin:
\tex_lccode:D ‘\* = ‘ %
\tex_lccode:D ‘A = ‘A %
\tex_lowercase:D
{ \group_end:
\cs_new:Npn \__str_to_other_loop:w #1 \* #2 \* #3 \* #4 \* #5 \* #6 \* #7 \* #8 \* #9 \s__str_stop
{ \if_meaning:w A #8
\__str_to_other_end:w
\fi:
\__str_to_other_loop:w
#9 \* #1 \* #2 \* #3 \* #4 \* #5 \* #6 \* #7 \* #8 \* \s__str_stop
}
\cs_new:Npn \__str_to_other_end:w \fi: #1 \s__str_mark #2 \* A \* #3 \s__str_stop
{ \fi: #2 }
}

(End definition for \__kernel_str_to_other:n, \__str_to_other_loop:w, and \__str_to_other_end:w.)
The difference with \_kernel_str_to_other:n is that the converted part is left in the input stream, making these commands only restricted-expandable.

\cs_new:Npn \_kernel_str_to_other_fast:n #1
\exp_after:wN \_str_to_other_fast_loop:w \tl_to_str:n {#1} -

\group_begin:
\tex_lccode:D '{ = '{ \%
\tex_lccode:D 'A = '{A \%
\tex_lowercase:D
{\group_end:
\cs_new:Npn \_str_to_other_fast_loop:w
#1 * #2 * #3 * #4 * #5 * #6 * #7 * #8 * #9 *
\if_meaning:w A #9
\_str_to_other_fast_end:w
\fi:
#1 * #2 * #3 * #4 * #5 * #6 * #7 * #8 * #9
\_str_to_other_fast_loop:w *
\}
\cs_new:Npn \_str_to_other_fast_end:w
#1 * A #2 \s__str_stop {#1}
\}

(End definition for \_kernel_str_to_other_fast:n, \_kernel_str_to_other_fast_loop:w, and \_str_to_other_fast_end:w.)

\str_item:Nn \str_item:cn \str_item:nn \str_item_ignore_spaces:nn
\_str_item:nn \_str_item:w
The \str_item:nn hands its argument with spaces escaped to \_str_item:nn, and makes sure to turn the result back into a proper string (with category code 10 spaces) eventually. The \str_item_ignore_spaces:nn function does not escape spaces, which are thus ignored by \_str_item:nn since everything else is done with undelimited arguments. Evaluate the ⟨index⟩ argument #2 and count characters in the string, passing those two numbers to \_str_item:w for further analysis. If the ⟨index⟩ is negative, shift it by the ⟨count⟩ to know the how many character to discard, and if that is still negative give an empty result. If the ⟨index⟩ is larger than the ⟨count⟩, give an empty result, and otherwise discard ⟨index⟩ – 1 characters before returning the following one. The shift by –1 is obtained by inserting an empty brace group before the string in that case: that brace group also covers the case where the ⟨index⟩ is zero.

\cs_new:Npn \str_item:Nn { \exp_args:No \str_item:nn }
\cs_generate_variant:Nn \str_item:Nn { c }
\cs_new:Npn \str_item:nn #1#2
\exp_args:Nf \tl_to_str:n
{ \exp_args:Nf \_str_item:nn
{ \__kernel_str_to_other:n {#1} } {#2}
}
\}
\cs_new:Npn \str_item_ignore_spaces:nn #1
{ \exp_args:No \_str_item:nn
{ \tl_to_str:n {#1} } }
\cs_new:Npn \_str_item:nn #1#2
{
\__str_item:w \__str_item:w
\int_value:w \int_eval:n \#2 \exp_after:wN \int_value:w \__str_count:n \#1 ;
\#1 \s__str_stop
\cs_new:Npn \__str_item:w #1; #2;
{
\int_compare:nNnTF {#1} < 0
{
\int_compare:nNnTF {#1} < {-#2}
{
\__str_use_none_delimit_by_s_stop:w
\exp_after:wN \__str_use_i_delimit_by_s_stop:nw
\exp:w \exp_after:wN \__str_skip_exp_end:w
\int_value:w \int_eval:n \{ \#1 + \#2 \} ;
\}
\}
\}
\int_compare:nNnTF {#1} > {#2}
{
\__str_use_none_delimit_by_s_stop:w
\exp_after:wN \__str_use_i_delimit_by_s_stop:nw
\exp:w \__str_skip_exp_end:w \#1 ; \{ \}
\}
\}
\int_compare:nNnTF {#1} < {#2}
{
\__str_use_none_delimit_by_s_stop:w
\exp_after:wN \__str_use_i_delimit_by_s_stop:nw
\exp:w \exp_after:wN \__str_skip_exp_end:w
\int_value:w \int_eval:n \{ \#1 + \#2 \} ;
\}
\}
\}

(End definition for \str_item:Nn and others. These functions are documented on page 126.)

\__str_skip_exp_end:w
\__str_skip_loop:wNNNNNNNN
\__str_skip_end:w
\__str_skip_end:NNNNNNNN

Removes max(#1,0) characters from the input stream, and then leaves \exp_end:. This should be expanded using \exp:w. We remove characters 8 at a time until there are at most 8 to remove. Then we do a dirty trick: the \if_case:w construction leaves between 0 and 8 times the \or: control sequence, and those \or: become arguments of \__str_skip_end:NNNNNNNN. If the number of characters to remove is 6, say, then there are two \or: left, and the 8 arguments of \__str_skip_end:NNNNNNNN are the two \or:, and 6 characters from the input stream, exactly what we wanted to remove. Then close the \if_case:w conditional with \fi:, and stop the initial expansion with \exp_end: (see places where \__str_skip_exp_end:w is called).
\cs_new:Npn \__str_skip_exp_end:w \#1;
{
\if_int_compare:w \#1 > 8 \exp_stop_f:
\exp_after:wN \__str_skip_loop:wNNNNNNNN
\else:
\exp_after:wN \__str_skip_end:w
\int_value:w \int_eval:w
\fi:
\#1 ;
\}
\cs_new:Npn \__str_skip_loop:wNNNNNNNN \#1; \#2#3#4#5#6#7#8#9
{
\exp_after:wN \__str_skip_exp_end:w
\int_value:w \int_eval:n \{ \#1 - 8 \} ;
\}
\cs_new:Npn \__str_skip_end:w \#1 ;

707
Sanitize the string. Then evaluate the arguments. At this stage we also decrement the \langle start index\rangle, since our goal is to know how many characters should be removed. Then limit the range to be non-negative and at most the length of the string (this avoids needing to check for the end of the string when grabbing characters), shifting negative numbers by the appropriate amount. Afterwards, skip characters, then keep some more, and finally drop the end of the string.

This function converts an \langle index\rangle argument into an explicit position in the string (a result of 0 denoting “out of bounds”). Expects two explicit integer arguments: the \langle index\rangle \#1 and the string count \#2. If \#1 is negative, replace it by \#1 + \#2 + 1, then limit to the range [0, \#2].
\begin{verbatim}
\__str_collect_delimit_by_q_stop:w #1;
\__str_collect_loop:wn #1 ; { } }
\cs_new:Npn \__str_collect_loop:wn #1 ;
\exp_after:wN \__str_collect_loop:wnNNNNNNN
\int_value:w \int_eval:n { #1 - 7 } ; { #2 #3#4#5#6#7#8#9 }
\cs_new:Npn \__str_collect_end:wn #1 ;
\exp_after:wN \__str_collect_end:nnnnnnnnw
\if_case:w \if_int_compare:w #1 > \c_zero_int
\else:
\or: \or: \or: \or: \or: \fi:
\fi:
\cs_new:Npn \__str_collect_end:nnnnnnnnw #1#2#3#4#5#6#7#8 #9 \s__str_stop
\{ #1#2#3#4#5#6#7#8 }
\end{verbatim}

Collects \texttt{max(#1,0)} characters, and removes everything else until \texttt{\s__str_stop}. This is somewhat similar to \texttt{\__str_skip_exp_end:w}, but accepts integer expression arguments. This time we can only grab 7 characters at a time. At the end, we use an \texttt{\if_case:w} trick again, so that the 8 first arguments of \texttt{\__str_collect_end:nnnnnnnnw} are some \texttt{\or:}; followed by an \texttt{\fi:}, followed by \texttt{#1} characters from the input stream. Simply leaving this in the input stream closes the conditional properly and the \texttt{\or:} disappear.

\begin{verbatim}
{ \int_eval:n
  { \if_int_compare:w #1 < \c_zero_int
    \if_int_compare:w #1 < \c_zero_int
      \fi:
    \fi:
  \else:
    #1 + #2 + 1
    \fi:
  \else:
  \fi:
}

(End definition for \_\_str_range_normalize:nn.)
\end{verbatim}

\begin{verbatim}
\exp_after:wN \__str_collect_loop:wn
\exp_after:wN \__str_collect_end:wn
\exp_after:wN \__str_collect_end:nnnnnnnnw
\end{verbatim}

(End definition for \_\_str_collect_delimit_by_q_stop:w and others.)
53.7 Counting characters

To speed up this function, we grab and discard 9 space-delimited arguments in each iteration of the loop. The loop stops when the last argument is one of the trailing \textlangle \text{number} \textrangle, and that \langle \text{number} \rangle is added to the sum of 9 that precedes, to adjust the result.

\begin{verbatim}
\cs_new:Npn \str_count_spaces:N { \exp_args:No \str_count_spaces:n \}
\cs_generate_variant:Nn \str_count_spaces:N { c, n }
\cs_new:Npn \str_count_spaces:n #1 #2 #3 #4 #5 #6 #7 #8 #9 ~ \__str_stop
\{ \int_eval:n \{ \exp_after:wN \__str_count_spaces_loop:w \tl_to_str:n {#1} ~ \}
\}
\cs_new:Npn \__str_count_spaces_loop:w #1~#2~#3~#4~#5~#6~#7~#8~#9~
\{ \if_meaning:w X #9 \__str_use_i_delimit_by_s_stop:nw \fi: \}
\end{verbatim}

(End definition for \texttt{\str_count_spaces:N}, \texttt{\str_count_spaces:n}, and \texttt{\__str_count_spaces_loop:w}. These functions are documented on page 125.)

To count characters in a string we could first escape all spaces using \texttt{\__kernel_str_to_other:n}, then pass the result to \texttt{\tl_count:n}. However, the escaping step would be quadratic in the number of characters in the string, and we can do better. Namely, sum the number of spaces (\texttt{\str_count_spaces:n}) and the result of \texttt{\tl_count:n}, which ignores spaces. Since strings tend to be longer than token lists, we use specialized functions to count characters ignoring spaces. Namely, loop, grabbing 9 non-space characters at each step, and end as soon as we reach one of the 9 trailing items. The internal function \texttt{\__str_count:n}, used in \texttt{\str_item:nn} and \texttt{\str_range:nnn}, is similar to \texttt{\str_count_ignore_spaces:n} but expects its argument to already be a string or a string with spaces escaped.

\begin{verbatim}
\cs_new:Npn \str_count:N { \exp_args:No \str_count:n \}
\cs_generate_variant:Nn \str_count:N { c, n }
\cs_new:Npn \str_count:n #1 \{ \__str_count_aux:n \begin{verbatim}
\{ \if_meaning:w X \textlangle \text{number} \textrangle \__str_use_i_delimit_by_s_stop:nw \fi: \}
\end{verbatim}
\}
\cs_new:Npn \__str_count:n #1 \{ \__str_count_aux:n \begin{verbatim}
\{ \if_meaning:w X \textlangle \text{number} \textrangle \__str_use_i_delimit_by_s_stop:nw \fi: \}
\}
\end{verbatim}
\end{verbatim}

710
The first character in a string

The \texttt{\_ignore_spaces} variant applies \texttt{\tl_to_str:n} then grabs the first item, thus skipping spaces. As usual, \texttt{\str_head:N} expands its argument and hands it to \texttt{\str_head:n}. To circumvent the fact that \TeX{} skips spaces when grabbing undelimited macro parameters, \texttt{\_str_head:w} takes an argument delimited by a space. If \texttt{#1} starts with a non-space character, \texttt{\_str_use_i_delimit_by_s_stop:nw} leaves that in the input stream. On the other hand, if \texttt{#1} starts with a space, the \texttt{\_str_head:w} takes an empty argument, and the single (initially braced) space in the definition of \texttt{\_str_head:w} makes its way to the output. Finally, for an empty argument, the (braced) empty brace group in the definition of \texttt{\str_head:n} gives an empty result after passing through \texttt{\_str_use_i_delimit_by_s_stop:nw}.

\begin{verbatim}
\cs_new:Npn \str_head:N  { \exp_args:No \str_head:n }
\cs_generate_variant:Nn \str_head:N { c }
\cs_new:Npn \str_head:n #1
{ \exp_after:wN \__str_head:w
  \tl_to_str:n {#1} { { } } ~ \s__str_stop }
\cs_new:Npn \__str_head:w #1 ~ %
{ \__str_use_i_delimit_by_s_stop:nw #1 { ~ } }
\cs_new:Npn \str_head:n #1
{ \exp_after:wN \__str_use_i_delimit_by_s_stop:nw
  \str_count:n {#1} #1 ~ %
  \tl_to_str:n {#1} { { } } ~ \s__str_stop }
\end{verbatim}

(End definition for \texttt{\str_count:N} and others. These functions are documented on page 125.)
Getting the tail is a little bit more convoluted than the head of a string. We hit the front
of the string with \reverse_if:N \if_charcode:w \scan_stop:. This removes the first
character, and necessarily makes the test true, since the character cannot match \scan_-
stop:. The auxiliary function then inserts the required \fi: to close the conditional, and
leaves the tail of the string in the input stream. The details are such that an empty string
has an empty tail (this requires in particular that the end-marker X be unexpandable and
not a control sequence). The \_ignore_spaces is rather simpler: after converting the
input to a string, \_str_tail_auxii:w removes one undelimited argument and leaves
everything else until an end-marker \s__str_mark. One can check that an empty (or
blank) string yields an empty tail.

\str_tail:N
\str_tail:c
\str_tail:n
\str_tail_ignore_spaces:n
\_str_tail_auxi:w
\_str_tail_auxii:w

Case changing for programmatic reasons is done by first detokenizing input then doing
a simple loop that only has to worry about spaces and everything else. The output is
detokenized to allow data sharing with text-based case changing.

\str_foldcase:n
\str_foldcase:V
\str_lowercase:n
\str_lowercase:f
\str_uppercase:n
\str_uppercase:f
\_str_change_case:nn
\_str_change_case_aux:nn
\_str_change_case_result:n
\_str_change_case_output:nw
\_str_change_case_output:fw
\_str_change_case_end:nw
\_str_change_case_loop:nw
\_str_change_case_space:n
\_str_change_case_char:nW

53.9 String manipulation
\begin{verbatim}
\__str_change_case_loop:nw {#2} #1 \__str_recursion_tail \q__str_recursion_stop
\__str_change_case_result:n { }
\}
\cs_new:Npn \__str_change_case_output:nw #1#2 \__str_change_case_result:n #3
{ #2 \__str_change_case_result:n { #3 #1 } }
\cs_generate_variant:Nn \__str_change_case_output:nw { f }
\cs_new:Npn \__str_change_case_end:wn #1 \__str_change_case_result:n #2
{ \tl_to_str:n {#2} }
\cs_new:Npn \__str_change_case_loop:nw #1#2 \q__str_recursion_stop
{\tl_if_head_is_space:nTF {#2}
{ \__str_change_case_space:n }
{ \__str_change_case_char:nN }
{#1} #2 \q__str_recursion_stop
}\exp_last_unbraced:NNNNo
\cs_new:Npn \__str_change_case_space:n #1 \c_space_tl
{ \__str_change_case_output:nw { ~ }
\__str_change_case_loop:nw {#1} }
\cs_new:Npn \__str_change_case_char:nN #1#2
{ \__str_if_recursion_tail_stop_do:Nn #2
{ \__str_change_case_end:wn }
\__str_change_case_output:fw
{ \use:c { char_str_#1 case:N } #2 }
\__str_change_case_loop:nw {#1} }
\end{verbatim}

(End definition for \texttt{\str_foldcase:n} and others. These functions are documented on page 130.)

\texttt{\c_ampersand_str} \texttt{\c_atsign_str} \texttt{\c_backslash_str} \texttt{\c_left_brace_str} \texttt{\c_right_brace_str} \texttt{\c_circumflex_str} \texttt{\c_colon_str} \texttt{\c_dollar_str} \texttt{\c_hash_str} \texttt{\c_percent_str} \texttt{\c_tilde_str} \texttt{\c_underscore_str} \texttt{\c_zero_str}

For all of those strings, use \texttt{\cs_to_str:N} to get characters with the correct category code without worries

\texttt{\c_ampersand_str} \texttt{\c_atsign_str} \texttt{\c_backslash_str} \texttt{\c_left_brace_str} \texttt{\c_right_brace_str} \texttt{\c_circumflex_str} \texttt{\c_colon_str} \texttt{\c_dollar_str} \texttt{\c_hash_str} \texttt{\c_percent_str} \texttt{\c_tilde_str} \texttt{\c_underscore_str} \texttt{\c_zero_str}

(End definition for \texttt{\c_ampersand_str} and others. These variables are documented on page 131.)

\texttt{\l_tmpa_str} \texttt{\l_tmpb_str} \texttt{\g_tmpa_str} \texttt{\g_tmpb_str}

Scratch strings.

\texttt{\l_tmpa_str} \texttt{\l_tmpb_str} \texttt{\g_tmpa_str} \texttt{\g_tmpb_str}
53.10 Viewing strings

\str_show:n  Displays a string on the terminal.
\str_show:N
\str_show:c
\str_log:n
\str_log:N
\str_log:c

\cs_new_eq:NN \str_show:n \tl_show:n
\cs_new_protected:Npn \str_show:N #1
{ \__kernel_chk_tl_type:NnnT #1 { str } { \tl_to_str:N #1 } { \tl_show:N #1 } }
\cs_generate_variant:Nn \str_show:N { c }
\cs_new_eq:NN \str_log:n \tl_log:n
\cs_new_protected:Npn \str_log:N #1
{ \__kernel_chk_tl_type:NnnT #1 { str } { \tl_to_str:N #1 } { \tl_log:N #1 } }
\cs_generate_variant:Nn \str_log:N { c }

(End definition for \\textbackslash tmpl\_str and others. These functions are documented on page 190.)
Chapter 54

l3str-convert implementation

54.1 Helpers

54.1.1 Variables and constants

\_\_str\_tmp:w
Internal scratch space for some functions.
\l\_str\_internal\_tl
\cs_new_protected:Npn \__str\_tmp:w { }
\tl_new:N \l__str\_internal\_tl
(End definition for \_\_str\_tmp:w and \l__str\_internal\_tl.)

\g\_str\_result\_tl
The \g__str_result_tl variable is used to hold the result of various internal string operations (mostly conversions) which are typically performed in a group. The variable is global so that it remains defined outside the group, to be assigned to a user-provided variable.
\tl_new:N \g__str_result_tl
(End definition for \g__str_result_tl.)

\c\_str\_replacement\_char\_int
When converting, invalid bytes are replaced by the Unicode replacement character "FFFD.
\int_const:Nn \c__str_replacement_char_int { "FFFD }
(End definition for \c__str_replacement_char_int.)

\c\_str\_max\_byte\_int
The maximal byte number.
\int_const:Nn \c__str_max_byte_int { 255 }
(End definition for \c__str_max_byte_int.)

\s\_str
Internal scan marks.
\scan_new:N \s__str
(End definition for \s__str.)
\__str_nil Internal quarks.
\quark_new:N \__str_nil

(End definition for \__str_nil.)

\g__str_alias_prop To avoid needing one file per encoding/escaping alias, we keep track of those in a property list.
\prop_new:N \g__str_alias_prop
\prop_gput:Nnn \g__str_alias_prop { latin1 } { iso88591 }
\prop_gput:Nnn \g__str_alias_prop { latin2 } { iso88592 }
\prop_gput:Nnn \g__str_alias_prop { latin3 } { iso88593 }
\prop_gput:Nnn \g__str_alias_prop { latin4 } { iso88594 }
\prop_gput:Nnn \g__str_alias_prop { latin5 } { iso88599 }
\prop_gput:Nnn \g__str_alias_prop { latin6 } { iso885910 }
\prop_gput:Nnn \g__str_alias_prop { latin7 } { iso885913 }
\prop_gput:Nnn \g__str_alias_prop { latin8 } { iso885914 }
\prop_gput:Nnn \g__str_alias_prop { latin9 } { iso885915 }
\prop_gput:Nnn \g__str_alias_prop { latin10 } { iso885916 }
\prop_gput:Nnn \g__str_alias_prop { utf16le } { utf16 }
\prop_gput:Nnn \g__str_alias_prop { utf16be } { utf16 }
\prop_gput:Nnn \g__str_alias_prop { utf32le } { utf32 }
\prop_gput:Nnn \g__str_alias_prop { utf32be } { utf32 }
\prop_gput:Nnn \g__str_alias_prop { hexadecimal } { hex }
\bool_lazy_any:nTF
\sys_if_engine_luatex_p:
\sys_if_engine_xetex_p:
\prop_gput:Nnn \g__str_alias_prop { default } { }
\prop_gput:Nnn \g__str_alias_prop { default } { utf8 }

(End definition for \g__str_alias_prop.)

\g__str_error_bool In conversion functions with a built-in conditional, errors are not reported directly to the user, but the information is collected in this boolean, used at the end to decide on which branch of the conditional to take.
\bool_new:N \g__str_error_bool

(End definition for \g__str_error_bool.)

\str_byte Conversions from one ⟨encoding⟩/(escaping) pair to another are done within x-expanding assignments. Errors are signalled by raising the relevant flag.
\flag_new:n { str_byte }
\flag_new:n { str_error }

(End definition for \str_byte and \str_error. These variables are documented on page ??.)
54.2 String conditionals

\_\_str\_if\_contains\_char:nnTF (\langle token list\rangle) (char)

Expects the \langle token list\rangle to be an \langle other string\rangle: the caller is responsible for ensuring that no (too-)special catcodes remain. Loop over the characters of the string, comparing character codes. The loop is broken if character codes match. Otherwise we return “false”.

\prg\new\conditional:Npnn \_\_str\_if\_contains\_char:nn #1#2 { T , TF }

{ \exp\after:wN \_\_str\_if\_contains\_char\_aux:nn \exp\after:wN \{#1\} \{#2\}
\prg\break:n \{ ? \fi: \}
\prg\break\point:
\prg\return\false:
}

\cs\new\Npn \_\_str\_if\_contains\_char\_aux:nn #1#2
{ \_\_str\_if\_contains\_char\_auxi:nN \{#2\} \{#1\} }
\prg\new\conditional:Npnn \_\_str\_if\_contains\_char:nn \_\_str\_if\_contains\_char:nn #1#2 { TF }
{ \_\_str\_if\_contains\_char\_auxi:nN \{#2\} \{#1\} }
\prg\break\point:
\prg\return\false:
\cs\new\Npn \_\_str\_if\_contains\_char\_auxi:nN #1#2
{ \if\char\code:w #1 #2
\exp\after:wN \_\_str\_if\_contains\_char\_true:
\fi:
\_\_str\_if\_contains\_char\_auxi:nN \{#1\}
}
\cs\new\Npn \_\_str\_if\_contains\_char\_true:
{ \prg\break:n \{ \prg\return\true: \use\none:n \} }
(End definition for \_\_str\_if\_contains\_char:nnT and others.)

\_\_str\_octal\_use:NTF (token) \{\langle true code\rangle\} \{\langle false code\rangle\}

If the \langle token\rangle is an octal digit, it is left in the input stream, followed by the \langle true code\rangle. Otherwise, the \langle false code\rangle is left in the input stream.

\TeX\ hackers\ note: This function will fail if the escape character is an octal digit. We are thus careful to set the escape character to a known value before using it. \TeX\ dutifully detects octal digits for us: if \#1 is an octal digit, then the right-hand side of the comparison is ’\#1, greater than 1. Otherwise, the right-hand side stops as ’1, and the conditional takes the false branch.

\prg\new\conditional:Npnn \_\_str\_octal\_use:N \#1 { TF }
{ \if\int\compare:w \#1 < ’1 \token\to\str:N \#1 \exp\stop\f:
\#1 \prg\return\true:
\else:
\prg\return\false:
\fi:
}
(End definition for \_\_str\_octal\_use:NTF.)
\_\_str\_hexadecimal\_use:NTF \TeX{} detects uppercase hexadecimal digits for us (see \_\_str\_octal\_use:NTF), but not the lowercase letters, which we need to detect and replace by their uppercase counterpart.

\begin{verbatim}
\prg_new_conditional:Npnn \__str\_hexadecimal\_use:N #1 { TF }
{ \if_int_compare:w 1 < "1 \token_to_str:N #1 \exp_stop_f:
  #1 \prg_return_true:
\else:
  \if_case:w \int_eval:n { \exp_after:wN ' \token_to_str:N #1 - 'a }
    A \or: B \or: C \or: D \or: E \or: F \else:
      \prg_return_false:
    \exp_after:wN \use_none:n
  \fi:
  \prg_return_true:
\fi:
}(End definition for \_\_str\_hexadecimal\_use:NTF.)
\end{verbatim}

54.3 Conversions

54.3.1 Producing one byte or character

For each integer \( N \) in the range \([0, 255]\), we create a constant token list which holds three character tokens with category code other: the character with character code \( N \), followed by the representation of \( N \) as two hexadecimal digits. The value \(-1\) is given a default token list which ensures that later functions give an empty result for the input \(-1\).

\texttt{\_\_str\_byte\_0\_tl}
\texttt{\_\_str\_byte\_1\_tl}
\texttt{\_\_str\_byte\_255\_tl}
\texttt{\_\_str\_byte\_-1\_tl}

\texttt{\_\_kernel\_tl\_set:Nx \l__str\_internal\_tl { \tl_to_str:n { 0123456789ABCDEF } }}
\texttt{\tl_map_inline:Nn \l__str\_internal\_tl}{\tl_const:cx \{ \_\_str\_byte\_ \int_eval:n {"#1##1}_tl \}{ \char\_generate:nn { "#1##1} { 12 } \#1 \#1 \}}

\texttt{(End definition for \_\_str\_byte\_0\_tl and others.)}

\texttt{\_\_str\_output\_byte:n}
\texttt{\_\_str\_output\_byte:w}
\texttt{\_\_str\_output\_hexadecimal:n}
\texttt{\_\_str\_output\_end:}

Those functions must be used carefully: feeding them a value outside the range \([-1, 255]\] will attempt to use the undefined token list variable \_\_str\_byte\_(number)\_tl. Assuming that the argument is in the right range, we expand the corresponding token list, and pick either the byte (first token) or the hexadecimal representations (second and third tokens). The value \(-1\) produces an empty result in both cases.
\cs_new:Npn \__str_output_byte:n #1
\{ \__str_output_byte:w #1 \__str_output_end: \}
\cs_new:Npn \__str_output_byte:w
\{ \exp_after:wN \exp_after:wN \exp_after:wN \use_i:nnn \cs:w c__str_byte_ \int_eval:w \}
\cs_new:Npn \__str_output_hexadecimal:n #1
\{ \exp_after:wN \exp_after:wN \exp_after:wN \use_none:n \cs:w c__str_byte_ \int_eval:n {#1} _tl \cs_end: \}
\cs_new:Npn \__str_output_end:
\{ \scan_stop: _tl \cs_end: \}
\__str_output_byte_pair_be:n \__str_output_byte_pair_le:n \__str_output_byte_pair:nnN
\cs_new:Npn \__str_output_byte_pair_be:n #1
\{ \exp_args:Nf \__str_output_byte_pair:nnN { \int_div_truncate:nn { #1 } { "100 } } {#1} \use:nn \}
\cs_new:Npn \__str_output_byte_pair_le:n #1
\{ \exp_args:Nf \__str_output_byte_pair:nnN { \int_div_truncate:nn { #1 } { "100 } } {#1} \use_ii_i:nn \}
\cs_new:Npn \__str_output_byte_pair:nnN #1#2#3
\{ #3 { \__str_output_byte:n { #1 } } { \__str_output_byte:n { #2 - #1 * "100 } } \}
\__str_convert_gmap:N \__str_convert_gmap_loop:NN
\cs_new_protected:Npn \__str_convert_gmap:N #1
\{ \__kernel_tl_gset:Nx \g__str_result_tl \exp_after:wN \__str_convert_gmap_loop:NN \exp_after:wN #1 \g__str_result_tl { ? \prg_break: } \prg_break_point: \}

(End definition for \__str_output_byte:n and others.)

\__str_output_byte_pair_be:n \__str_output_byte_pair_le:n \__str_output_byte_pair:nnN
\cs_new:Npn \__str_output_byte_pair_be:n #1
\{ \exp_args:Nf \__str_output_byte_pair:nnN { \int_div_truncate:nn { #1 } { "100 } } {#1} \use:nn \}
\cs_new:Npn \__str_output_byte_pair_le:n #1
\{ \exp_args:Nf \__str_output_byte_pair:nnN { \int_div_truncate:nn { #1 } { "100 } } {#1} \use_ii_i:nn \}
\cs_new:Npn \__str_output_byte_pair:nnN #1#2#3
\{ #3 { \__str_output_byte:n { #1 } } { \__str_output_byte:n { #2 - #1 * "100 } } \}
(End definition for \__str_output_byte_pair_be:n, \__str_output_byte_pair_le:n, and \__str_output_byte_pair:nnN)

54.3.2 Mapping functions for conversions

This maps the function #1 over all characters in \g__str_result_tl, which should be a byte string in most cases, sometimes a native string.
\cs_new_protected:Npn \__str_convert_gmap:N #1
\{ \__kernel_tl_gset:Nx \g__str_result_tl \exp_after:wN \__str_convert_gmap_loop:NN \exp_after:wN #1 \g__str_result_tl { ? \prg_break: } \prg_break_point: \}

719
This maps the function \#1 over all character codes in \texttt{g\_str\_result\_tl}, which must be in the internal representation.

54.3.3 Error-reporting during conversion

When converting using the function \texttt{str\_set\_convert:Nnnn}, errors should be reported to the user after each step in the conversion. Errors are signalled by raising some flag (typically \texttt{@@\_error}), so here we test that flag: if it is raised, give the user an error, otherwise remove the arguments. On the other hand, in the conditional functions \texttt{str\_set\_convert:NnnnTF}, errors should be suppressed. This is done by changing \texttt{\_str\_if\_flag\_error:nxx} into \texttt{\_str\_if\_flag\_no\_error:nxx} locally.

At the end of each conversion step, we raise all relevant errors as one error message, built on the fly. The height of each flag indicates how many times a given error was encountered. This function prints \#2 followed by the number of occurrences of an error if it occurred, nothing otherwise.
54.3.4 Framework for conversions

Most functions in this module expect to be working with “native” strings. Strings can also be stored as bytes, in one of many encodings, for instance UTF8. The bytes themselves can be expressed in various ways in terms of TeX tokens, for instance as pairs of hexadecimal digits. The questions of going from arbitrary Unicode code points to bytes, and from bytes to tokens are mostly independent.

Conversions are done in four steps:

- “unescape” produces a string of bytes;
- “decode” takes in a string of bytes, and converts it to a list of Unicode characters in an internal representation, with items of the form
  \[
  \langle \text{bytes} \rangle \s__str \langle \text{Unicode code point} \rangle \s__str
  \]
  where we have collected the \(\langle \text{bytes} \rangle\) which combined to form this particular Unicode character, and the \(\langle \text{Unicode code point} \rangle\) is in the range \([0,10FFFF]\).
- “encode” encodes the internal list of code points as a byte string in the new encoding;
- “escape” escapes bytes as requested.

The process is modified in case one of the encoding is empty (or the conversion function has been set equal to the empty encoding because it was not found): then the unescape or escape step is ignored, and the decode or encode steps work on tokens instead of bytes. Otherwise, each step must ensure that it passes a correct byte string or internal string to the next step.

The input string is stored in \(\g__str\) _\_result_t\_l, then we: unescape and decode; encode and escape; exit the group and store the result in the user’s variable. The various conversion functions all act on \(\g__str\) _\_result_t\_l. Errors are silenced for the conditional functions by redefining \(\_\_str\) _\_if_flag_error:nnx locally.
The task of \_\_str_convert:wwwnn is to split \langle encoding\rangle/\langle escaping\rangle pairs into their components, #1 and #2. Calls to \_\_str_convert:nnn ensure that the corresponding conversion functions are defined. The third auxiliary does the main work.

- #1 is the encoding conversion function;
- #2 is the escaping function;
- #3 is the escaping name for use in an error message;
- #4 is \prg_do_nothing: for unescaping/decoding, and \use_ii_i:nn for encoding/escaping;
- #5 is the default encoding function (either “decode” or “encode”), for which there should be no escaping.

Let us ignore the native encoding for a second. In the unescaping/decoding phase, we want to do #2#1 in this order, and in the encoding/escaping phase, the order should be reversed: #4#2#1 does exactly that. If one of the encodings is the default (native), then the escaping should be ignored, with an error if any was given, and only the encoding, #1, should be performed.
The arguments of \cs_new_protected:Nnn \__str_convert:nnn are: enc or esc, used to build filenames, the type of the conversion (unescape, decode, encode, escape), and the encoding or escaping name. If the function is already defined, no need to do anything. Otherwise, filter out all non-alphanumerics in the name, and lowercase it. Feed that, and the same three arguments, to \cs_new_protected:Nnn \__str_convert:nnnn. The task is then to make sure that the conversion function \#3_\#1 corresponding to the type \#3 and filtered name \#1 is defined, then set our initial conversion function \#3_\#4 equal to that.

How do we get the \#3_\#1 conversion to be defined if it isn’t? Two main cases.

First, if \#1 is a key in \g__str_alias_prop, then the value \l__str_internal_tl tells us what file to load. Loading is skipped if the file was already read, i.e., if the conversion command based on \l__str_internal_tl already exists. Otherwise, try to load the file; if that fails, there is an error, use the default empty name instead.

Second, \#1 may be absent from the property list. The \cs_if_exist:cF test is automatically false, and we search for a file defining the encoding or escaping \#1 (this should allow third-party .def files). If the file is not found, there is an error, use the default empty name instead.

In all cases, the conversion based on \l__str_internal_tl is defined, so we can set the \#3_\#1 function equal to that. In some cases (e.g., utf16be), the \#3_\#1 function is actually defined within the file we just loaded, and it is different from the \l__str_internal_tl-based function: we mustn’t clobber that different definition.

(End definition for \cs_new_protected:Nnn \__str_convert:nnn and \cs_new_protected:Nnn \__str_convert:nnnn.)
\{\group_begin:
\cctab_select:N \c_code_cctab
\file_input:n { l3str-#2- \l__str_internal_tl .def }
\group_end:
\}
\tl_clear:N \l__str_internal_tl
\msg_error:nnxx { str } { unknown-#2 } {#4} {#1}
\}
\cs_if_exist:cF { __str_convert_#3_#1: }
{ \cs_gset_eq:cc { __str_convert_#3_#1: }
\file_input:n { l3str-#2- \l__str_internal_tl .def }
\msg_error:nnxx { str } { unknown-#2 } {#4} {#1}
\}
\cs_gset_eq:cc { __str_convert_#3_#4: } { __str_convert_#3_#1: }
\}
\end_definition_for __str_convert:nnn and __str_convert:nnnn.
\}
\cs_new:Npn \__str_convert_lowercase_alphanum:n #1
{ \exp_after:wN \__str_convert_lowercase_alphanum_loop:N \tl_to_str:n {#1} { ? \prg_break: }
\prg_break_point:
\}
\cs_new:Npn \__str_convert_lowercase_alphanum_loop:N #1
{ \use_none:n #1
\if_int_compare:w '#1 > 'Z \exp_stop_f:
\if_int_compare:w '#1 > 'z \exp_stop_f: \else:
\if_int_compare:w '#1 < 'a \exp_stop_f: \else:
#1
\fi:
\fi:
\else:
\if_int_compare:w '#1 < 'A \exp_stop_f:
\if_int_compare:w 1 < #1 \exp_stop_f: \else:
#1
\fi:
\fi:
\else:
\__str_output_byte:n { '#1 + 'a - 'A }
\fi:
\fi:
\__str_convert_lowercase_alphanum_loop:N
\}
\end_definition_for __str_convert_lowercase_alphanum:n and __str_convert_lowercase_alphanum_loop:N.
\}
\end_definition_for __str_convert_lowercase_alphanum:n and __str_convert_lowercase_alphanum_loop:N.

This function keeps only letters and digits, with upper case letters converted to lower case.

\{\cs_new:Npn \__str_convert_lowercase_alphanum:n #1
{ \exp_after:wN \__str_convert_lowercase_alphanum_loop:N \tl_to_str:n {#1} { ? \prg_break: }
\prg_break_point:
\}
\cs_new:Npn \__str_convert_lowercase_alphanum_loop:N #1
{ \use_none:n #1
\if_int_compare:w '#1 > 'Z \exp_stop_f:
\if_int_compare:w '#1 > 'z \exp_stop_f: \else:
\if_int_compare:w '#1 < 'a \exp_stop_f: \else:
#1
\fi:
\fi:
\else:
\if_int_compare:w '#1 < 'A \exp_stop_f:
\if_int_compare:w 1 < #1 \exp_stop_f: \else:
#1
\fi:
\fi:
\else:
\__str_output_byte:n { '#1 + 'a - 'A }
\fi:
\fi:
\__str_convert_lowercase_alphanum_loop:N
\}
\end_definition_for __str_convert_lowercase_alphanum:n and __str_convert_lowercase_alphanum_loop:N.

(End definition for __str_convert_lowercase_alphanum:n and __str_convert_lowercase_alphanum_loop:N.)
54.3.5 Byte unescape and escape

Strings of bytes may need to be stored in auxiliary files in safe “escaping” formats. Each such escaping is only loaded as needed. By default, on input any non-byte is filtered out, while the output simply consists in letting bytes through.

In the case of 8-bit engines, every character is a byte. For Unicode-aware engines, test the character code; non-bytes cause us to raise the flag \texttt{str\_byte}. Spaces have already been given the correct category code when this function is called.

The simplest unescaping method removes non-bytes from \texttt{g\_str\_result\_tl}.
The simplest form of escape leaves the bytes from the previous step of the conversion unchanged.

\cs_new_protected:Npn \__str_convert_escape_: { }
\cs_new_eq:NN \__str_convert_escape_bytes: \__str_convert_escape_:

(End definition for \__str_convert_escape_: and \__str_convert_escape_bytes:)

54.3.6 Native strings

Convert each character to its character code, one at a time.

\cs_new_protected:Npn \__str_convert_decode_: { \__str_convert_gmap:N \__str_decode_native_char:N }
\cs_new:Npn \__str_decode_native_char:N #1 { #1 \s__str \int_value:w '#1 \s__str }

(End definition for \__str_convert_decode_: and \__str_decode_native_char:N)

The conversion from an internal string to native character tokens basically maps \char_generate:nn through the code-points, but in non-Unicode-aware engines we use a fallback character ? rather than nothing when given a character code outside [0, 255]. We detect the presence of bad characters using a flag and only produce a single error after the x-expanding assignment.

\bool_lazy_any:nTF
\sys_if_engine_luatex_p:
\sys_if_engine_xetex_p:
\cs_new_protected:Npn \__str_convert_encode_: { \__str_convert_gmap_internal:N \__str_encode_native_char:n }
\cs_new:Npn \__str_encode_native_char:n #1 { \char_generate:nn {#1} {12} }

\cs_new_protected:Npn \__str_convert_encode_: { \__str_decode_native_char:n #1 }
\msg_new:nnnn { str } { native-overflow } { Character~code~too~large~for~this~engine. }

726
This engine only supports 8-bit characters:
valid character codes are in the range [0, 255].
To manipulate arbitrary Unicode, use LuaTeX or XeTeX.

(End definition for \_\_str_convert_encode_: and \_\_str_encode_native_char:n.)

54.3.7 clist
\_\_str_convert_decode_clist:
\_\_str_decode_clist_char:n

Convert each integer to the internal form. We first turn \g__str_result_tl into a clist variable, as this avoids problems with leading or trailing commas.

\cs_new_protected:Npn \_\_str_convert_decode_clist:

\clist_gset:No \g__str_result_tl \g__str_result_tl
\__kernel_tl_gset:Nx \g__str_result_tl
\exp_args:No \clist_map_function:nN
\g__str_result_tl \_\_str_decode_clist_char:n
\cs_new:Npn \_\_str_decode_clist_char:n #1
{ #1 \s__str \int_eval:n { #1 } \s__str }

(End definition for \_\_str_convert_decode_clist: and \_\_str_decode_clist_char:n.)

\_\_str_convert_encode_clist:
\_\_str_encode_clist_char:n

Convert the internal list of character codes to a comma-list of character codes. The first line produces a comma-list with a leading comma, removed in the next step (this also works in the empty case, since \tl_tail:N does not trigger an error in this case).

\cs_new_protected:Npn \_\_str_convert_encode_clist:
\__str_convert_gmap_internal:N \_\_str_encode_clist_char:n
\__kernel_tl_gset:Nx \g__str_result_tl \{ \tl_tail:N \g__str_result_tl \}
\cs_new:Npn \_\_str_encode_clist_char:n #1
{ #1 \s__str \int_eval:n { #1 } \s__str }

(End definition for \_\_str_convert_encode_clist: and \_\_str_encode_clist_char:n.)

54.3.8 8-bit encodings

It is not clear in what situations 8-bit encodings are used, hence it is not clear what should be optimized. The current approach is reasonably efficient to convert long strings, and it scales well when using many different encodings.

The data needed to support a given 8-bit encoding is stored in a file that consists of a single function call

\_\_str_declare_eight_bit_encoding:nnnn {\langle name\rangle} {\langle modulo\rangle} {\langle mapping\rangle} {\langle missing\rangle}

This declares the encoding \langle name\rangle to map bytes to Unicode characters according to the \langle mapping\rangle, and map those bytes which are not mentioned in the \langle mapping\rangle either to the replacement character (if they appear in \langle missing\rangle), or to themselves. The \langle mapping\rangle argument is a token list of pairs \langle \langle byte\rangle \langle Unicode\rangle \rangle expressed in uppercase hexadecimal notation. The \langle missing\rangle argument is a token list of \langle \langle byte\rangle \rangle. Every \langle byte\rangle which does
not appear in the \textlangle mapping\textrangle nor the \textlangle missing\textrangle lists maps to itself in Unicode, so for instance the latin1 encoding has empty \textlangle mapping\textrangle and \textlangle missing\textrangle lists. The \textlangle modulo\textrangle is a (decimal) integer between 256 and 558 inclusive, modulo which all Unicode code points supported by the encodings must be different.

We use two integer arrays per encoding. When decoding we only use the \textit{decode} integer array, with entry \(n + 1\) (offset needed because integer array indices start at 1) equal to the Unicode code point that corresponds to the \(n\)-th byte in the encoding under consideration, or \(-1\) if the given byte is invalid in this encoding. When encoding we use both arrays: upon seeing a code point \(n\), we look up the entry \((1\text{ plus})\ n\) modulo some number \(M\) in the \textit{encode} array, which tells us the byte that might encode the given Unicode code point, then we check in the \textit{decode} array that indeed this byte encodes the Unicode code point we want. Here, \(M\) is an encoding-dependent integer between 256 and 558 (it turns out), chosen so that among the Unicode code points that can be validly represented in the given encoding, no pair of code points have the same value modulo \(M\).

Loop through both lists of bytes to fill in the \textit{decode} integer array, then fill the \textit{encode} array accordingly. For bytes that are invalid in the given encoding, store \(-1\) in the \textit{decode} array.

\begin{verbatim}
\cs_new_protected:Npn \__str_declare_eight_bit_encoding:nnnn #1
{ \tl_set:Nn \l__str_internal_tl {#1} \cs_new_protected:cpn { __str_convert_decode_#1: } { \__str_convert_decode_eight_bit:n {#1} } \cs_new_protected:cpn { __str_convert_encode_#1: } { \__str_convert_encode_eight_bit:n {#1} } \exp_args:Ncc \__str_declare_eight_bit_aux:NNnnn \cs_new_protected:Npn \__str_declare_eight_bit_aux:NNnnn #1#2#3#4#5
{ \intarray_new:Nn #1 { 256 } \int_step_inline:nnn { 0 } { 255 } { \intarray_gset:Nnn #1 { 1 + ##1 } {##1} } \__str_declare_eight_bit_loop:Nn #1
#4 { \s__str_stop \prg_break: } { } \prg_break_point: \__str_declare_eight_bit_loop:Nn #1
#5 { \s__str_stop \prg_break: } \prg_break_point: \intarray_new:Nn #2 {#3} \int_step_inline:nnn { 0 } { 255 } { \intarray_gset:Nn #1 { 1 + ##1 } { -1 } } \intarray_gset:Nn #2
{ 1 + \intarray_item:Nn #1 { 1 + ##1 } \int_mod:nn { \intarray_item:Nn #1 { 1 + ##1 } } { \intarray_count:N #2 } } \intarray_gset:Nn #1 { 1 + ##1 } } \intarray_gset:Nn #2
{##1} }
}
\end{verbatim}
\__str_declare_eight_bit_loop:Nnn #1#2#3
\intarray_gset:Nnn #1 \{ 1 + "#2 \} \{ "#3 \}
\__str_declare_eight_bit_loop:Nnn #1
\__str_declare_eight_bit_loop:Nn #1#2
\intarray_gset:Nnn #1 \{ 1 + "#2 \} \{ -1 \}
\__str_declare_eight_bit_loop:Nn #1

(End definition for \__str_declare_eight_bit_encoding:nnnn and others.)

The map from bytes to Unicode code points is in the decode array corresponding to the given encoding. Define \__str_tmp:w and pass it successively all bytes in the string. It produces an internal representation with suitable \s__str inserted, and the corresponding code point is obtained by looking it up in the integer array. If the entry is \texttt{-1} then issue a replacement character and raise the flag indicating that there was an error.

\begin{verbatim}
\cs_new_protected:Npn \__str_convert_decode_eight_bit:n \__str_decode_eight_bit_aux:n \__str_decode_eight_bit_aux:Nn
\end{verbatim}

It is not practical to make an integer array with indices in the full Unicode range, so we work modulo some number, which is simply the size of the encode integer array for the
given encoding. This gives us a candidate byte for representing a given Unicode code point. Of course taking the modulo leads to collisions so we check in the decode array that the byte we got is indeed correct. Otherwise the Unicode code point we started from is simply not representable in the given encoding.

```
\int_new:N \l__str_modulo_int
\cs_new_protected:Npn \__str_convert_encode_eight_bit:n #1
  { \cs_set:Npx \__str_tmp:w
    { \exp_not:N \__str_encode_eight_bit_aux:NNn
      \exp_not:c { g__str_encode_#1_intarray }
      \exp_not:c { g__str_decode_#1_intarray }
    }
    \flag_clear:n { str_error }
    \__str_convert_gmap_internal:N \__str_tmp:w
    \__str_if_flag_error:nnx { str_error } { encode-8-bit } {#1}
  }
\cs_new:Npn \__str_encode_eight_bit_aux:NNn #1#2#3
  { \exp_args:Nf \__str_encode_eight_bit_aux:nnN
    { \intarray_item:Nn #1 { 1 + \int_mod:nn {#3} { \intarray_count:N #1 } } }
    {#3}
    #2 }
\cs_new:Npn \__str_encode_eight_bit_aux:nnN #1#2#3
  { \int_compare:nNnTF { \intarray_item:Nn #3 { 1 + #1 } } = {#2}
    { \__str_output_byte:n {#1} }
    { \flag_raise:n { str_error } }
 }
```

(End definition for \__str_convert_encode_eight_bit:n, \__str_encode_eight_bit_aux:nnN, and \__str_encode_eight_bit_aux:NNn.)

## 54.4 Messages

General messages, and messages for the encodings and escapings loaded by default (“native”, and “bytes”).

```
\msg_new:nnn { str } { unknown-esc }
{ Escaping-scheme-‘#1’-(filtered:–‘#2’)-unknown. }
\msg_new:nnn { str } { unknown-enc }
{ Encoding-scheme-‘#1’-(filtered:–‘#2’)-unknown. }
\msg_new:nnnn { str } { native-escaping }
{ The-‘native’-encoding-scheme-does-not-support-any-escaping. }
\msg_new:nnn { str } { file-not-found }
```

730
Message used when the “bytes” unescaping fails because the string given to \str_set_convert:Nnnn contains a non-byte. This cannot happen for the -S-bit engines. Messages used for other escapings and encodings are defined in each definition file.

\bool_lazy_any:nT
{ sys_if_engine_luatex_p: }
{ sys_if_engine_xetex_p: }
{ \msg_new:nnnn { str } { non-byte } 
{ String-invalid-in-escaping-'#1':-it-may-only-contain-bytes. } 
{ Some-characters-in-the-string-you-asked-to-convert-are-not-
8-bit-characters.-Perhaps-the-string-is-a-'native'-Unicode-string?-If-it-is,-try-using\\ 
\\ \ioindent:n 
{ \iochar:N\\str_set_convert:Nnnn \ 
\ <str-var>-\{<string>-\}<-native>-\{<target-encoding>-\} } 
} 
}

Those messages are used when converting to and from 8-bit encodings.

\msg_new:nnnn { str } { decode-8-bit }
{ Invalid-string-in-encoding-'#1'. } 
{ LaTeX-came-across-a-byte-which-is-not-defined-to-represent-
any-character-in-the-encoding-'#1'. } 
\msg_new:nnnn { str } { encode-8-bit }
{ Unicode-string-cannot-be-converted-to-encoding-'#1'. } 
{ The-encoding-'#1'-only-contains-a-subset-of-all-Unicode-characters.-
LaTeX-was-asked-to-convert-a-string-to-that-encoding,-but-that-
string-contains-a-character-that-'#1'-does-not-support. } 

54.5 Escaping definitions

Several of those encodings are defined by the pdf file format. The following byte storage methods are defined:

- \texttt{bytes} (default), non-bytes are filtered out, and bytes are left untouched (this is defined by default);
- \texttt{hex} or \texttt{hexadecimal}, as per the pdf\TeX primitive \texttt{\pdfescapehex}
- \texttt{name}, as per the pdf\TeX primitive \texttt{\pdfescapename}
- \texttt{string}, as per the pdf\TeX primitive \texttt{\pdfescapestring}
- \texttt{url}, as per the percent encoding of urls.
54.5.1 Unescape methods

Take chars two by two, and interpret each pair as the hexadecimal code for a byte. Anything else than hexadecimal digits is ignored, raising the flag. A string which contains an odd number of hexadecimal digits gets 0 appended to it: this is equivalent to appending a 0 in all cases, and dropping it if it is alone.

```latex
\cs_new_protected:Npn \__str_convert_unescape_hex: \\
{ \\
\group_begin: \\
\flag_clear:n { str_error } \\
\int_set:Nn \tex_escapechar:D { 92 } \\
\__kernel_tl_gset:Nx \g__str_result_tl \\
\__str_output_byte:w " \exp_last_unbraced:Nf \__str_unescape_hex_auxi:N \\
\{ \tl_to_str:N \g__str_result_tl \} \\
0 \{ ? 0 \- 1 \prg_break: \} \\
\prg_break_point: \\
\__str_output_end: \\
\__str_if_flag_error:nnx { str_error } { unescape-hex } { } \\
\group_end: \\
} \\
\cs_new:Npn \__str_unescape_hex_auxi:N #1 \\
{ \\
\use_none:n #1 \\
\__str_hexadecimal_use:NTF #1 \\
\{ \__str_unescape_hex_auxii:N \} \\
\flag_raise:n { str_error } \\
\__str_unescape_hex_auxii:N \\
} \\
\cs_new:Npn \__str_unescape_hex_auxii:N #1 \\
{ \\
\use_none:n #1 \\
\__str_hexadecimal_use:NTF #1 \\
\__str_output_end: \\
\__str_output_byte:w " \__str_unescape_hex_auxi:N \\
\{ \tl_to_str:N \g__str_result_tl \} \\
\prg_break_point: \\
\__str_output_end: \\
\__str_output_byte:w " \__str_unescape_hex_auxi:N \\
\flag_raise:n { str_error } \\
\__str_unescape_hex_auxii:N \\
} \\
\msg_new:nnnn { str } { unescape-hex } \\
{ String invalid in escaping 'hex': only hexadecimal digits allowed. } \\
{ Some characters in the string you asked to convert are not hexadecimal digits (0-9, A-F, a-f) nor spaces. }
```

(End definition for \__str_convert_unescape_hex:, \__str_unescape_hex_auxi:N, and \__str_unescape_hex_auxii:N.)
The \_str_convert_unescape_name: function replaces each occurrence of # followed by two hexadecimal digits in \_g\_str_result_tl by the corresponding byte. The url function is identical, with escape character % instead of #. Thus we define the two together. The arguments of \_str_tmp:w are the character code of # or % in hexadecimal, the name of the main function to define, and the name of the auxiliary which performs the loop.

The looping auxiliary \#3 finds the next escape character, reads the following two characters, and tests them. The test \_str_hexadecimal_use:NTF leaves the upper-case digit in the input stream, hence we surround the test with \_str_output_byte:w " and \_str_output_end:. If both characters are hexadecimal digits, they should be removed before looping; this is done by \use_i:nnn. If one of the characters is not a hexadecimal digit, then feed '\#1 to \_str_output_byte:w to produce the escape character, raise the flag, and call the looping function followed by the two characters (remove \use_i:nnn).

\begin{verbatim}
cs_set_protected:8pn \_str_tmp:w \#1\#2\#3
{
cs_new_protected:cpn { \_str_convert_unescape_##2: }
{
group_begin:
\flag_clear:n { str_byte }
\flag_clear:n { str_error }
\int_set:Nn \tex_escapechar:D { 92 }
\_kernel_tl_gset:Nx \g\_str_result_tl
{
 \exp_after:wN \#3 \g\_str_result_tl
 #1 ? { ? \prg_break: }
 \prg_break_point:
 \__str_if_flag_error:nnx { str_byte } { non-byte } { \#2 }
 \__str_if_flag_error:nnx { str_error } { unescape-\#2 } { }
\group_end:

\cs_new:Npn \#3 ##1#1##2##3
{
 \_str_filter_bytes:n {##1}
 \use_none:n ##3
 \__str_output_byte:w "
 \__str_hexadecimal_use:NTF \#2
{
 \__str_hexadecimal_use:NTF \#3
{
 \flag_raise:n { str_error }
 * 0 + '\#1 \use_i:nn
 }
}
\__str_output_end:
\use_i:nnn \#3 \#2\#3
}
\end{verbatim}

733
The string escaping is somewhat similar to the name and url escapings, with escape character \. The first step is to convert all three line endings, \^J, \^M, and \^M\^J to the common \^J, as per the PDF specification. This step cannot raise the flag.

Then the following escape sequences are decoded.

\n Line feed (10)
\r Carriage return (13)
\t Horizontal tab (9)
\b Backspace (8)
\f Form feed (12)
\( Left parenthesis
\) Right parenthesis
\ Backslash

\ddd (backslash followed by 1 to 3 octal digits) Byte ddd (octal), subtracting 256 in case of overflow.

If followed by an end-of-line character, the backslash and the end-of-line are ignored. If followed by anything else, the backslash is ignored, raising the error flag.
\Kernel_tl_gset:Nx \g__str_result_tl
{
\exp_after:wN \__str_unescape_string_loop:wNNN
\g__str_result_tl \#1 ?? \prg_break_point:
\prg_break_point:
\__str_if_flag_error:nnx { str_byte } { non-byte } { string }
\__str_if_flag_error:nnx { str_error } { unescape-string } { }
\group_end:
}
\exp_args:No \__str_tmp:w { \c_backslash_str }
\exp_last_unbraced:NWNNn
\cs_new:Npn \__str_unescape_string_loop:wNNN \#1 \c_backslash_str \#2#3#4
{
\__str_filter_bytes:n {#1}
\use_none:n \#4
\__str_output_byte:w ' \__str_octal_use:NTF \#2
\__str_output_byte:w ' \__str_octal_use:NTF \#3
\__str_octal_use:NTF \#4
\__str_unescape_string_repeat:NNNNNN \#2 \#3 \#4
\if_int_compare:w \#2 > 3 \exp_stop_f:
- 256 \fi:
\__str_unescape_string_repeat:NNNNNN \#1 \#2 \#3 \#4
\__str_unescape_string_repeat:NNNNNN \#1 \#2 \#3 \#4
\str_case_e:nnF {#2}
{
{ \c_backslash_str } { 134 }
{ } { 50 }
{ } { 51 }
{ r } { 15 }
{ f } { 14 }
{ n } { 12 }
{ t } { 11 }
{ b } { 10 }
{ ^{\sim}J } { 0 - 1 }
}
\flag_raise:n { str_error }
-1 \use_i:nn
\use_i:nn \__str_unescape_string_loop:wNNN \#2 \#3 \#4
\__str_output_end:
54.5.2 Escape methods

Currently, none of the escape methods can lead to errors, assuming that their input is made out of bytes.

\_\_str\_convert\_escape\_hex:
\_\_str\_escape\_hex\_char:N

Loop and convert each byte to hexadecimal.

\_\_str\_convert\_escape\_name:
\_\_str\_escape\_name\_char:n
\_\_str\_if\_escape\_name:nTF
\c\_\_str\_escape\_name\_not\_str

For each byte, test whether it should be output as is, or be “hash-encoded”. Roughly, bytes outside the range \["2A, "7E\] are hash-encoded. We keep two lists of exceptions: characters in \c\_\_str\_escape\_name\_not\_str are not hash-encoded, and characters in the \c\_\_str\_escape\_name\_str are encoded.

\str\_const:Nn \c\_\_str\_escape\_name\_not\_str \{ ! "$ & ' \} %$
Any character below (and including) space, and any character above (and including) \texttt{del}, are converted to octal. One backslash is added before each parenthesis and backslash.

This function is similar to \texttt{\_\_str_convert_escape_name:}; escaping different characters.
54.6 Encoding definitions

The native encoding is automatically defined. Other encodings are loaded as needed. The following encodings are supported:

- **utf-8**;
- **utf-16**, big-, little-endian, or with byte order mark;
- **utf-32**, big-, little-endian, or with byte order mark;
- the iso 8859 code pages, numbered from 1 to 16, skipping the inexistent iso 8859-12.

54.6.1 utf-8 support

Loop through the internal string, and convert each character to its UTF-8 representation. The representation is built from the right-most (least significant) byte to the left-most (most significant) byte. Continuation bytes are in the range [128, 191], taking 64 different values, hence we roughly want to express the character code in base 64, shifting the first digit in the representation by some number depending on how many continuation bytes there are. In the range [0, 127], output the corresponding byte directly. In the range [128, 2047], output the remainder modulo 64, plus 128 as a continuation byte, then output the quotient (which is in the range [0, 31]), shifted by 192. In the next range, [2048, 65535], split the character code into residue and quotient modulo 64, output the residue as a first continuation byte, then repeat; this leaves us with a quotient in the range [0, 15], which we output shifted by 224. The last range, [65536, 1114111], follows the same pattern: once we realize that dividing twice by 64 leaves us with a number larger than 15, we repeat, producing a last continuation byte, and offset the quotient by 240 for the leading byte.

How is that implemented? \_\_str_encode_utf_vii_loop:wwnnw takes successive quotients as its first argument, the quotient from the previous step as its second argument (except in step 1), the bound for quotients that trigger one more step or not, and finally the offset used if this step should produce the leading byte. Leading bytes can be in
the ranges [0, 127], [192, 223], [224, 239], and [240, 247] (really, that last limit should be 244 because Unicode stops at the code point 1114111). At each step, if the quotient \#1 is less than the limit \#3 for that range, output the leading byte (\#1 shifted by \#4) and stop. Otherwise, we need one more step: use the quotient of \#1 by 64, and \#1 as arguments for the looping auxiliary, and output the continuation byte corresponding to the remainder \#2 − 64\#1 + 128. The bizarre construction \(-1 + 0 * \) removes the spurious initial continuation byte (better methods welcome).

\begin{verbatim}
\cs_new_protected:cpn { __str_convert_encode_utf8: }
{ __str_convert_gmap_internal:N __str_encode_utf_viii_char:n }
\cs_new:Npn __str_encode_utf_viii_char:n #1
{ __str_encode_utf_viii_loop:wwnnw #1 ; - 1 + 0 * ;
  { 128 } { 0 }
  { 32 } { 192 }
  { 16 } { 224 }
  { 8 } { 240 }
  \\s__str_stop
}
\cs_new:Npn __str_encode_utf_viii_loop:wwnnw #1; #2; #3#4 #5 \s__str_stop
{ \if_int_compare:w #1 < #3 \exp_stop_f:
  \__str_output_byte:n { #1 + #4 }
  \exp_after:wN \__str_usedelete_s_delimit_by_s_stop:w
  \fi:
  \exp_after:wN \__str_encode_utf_viii_loop:wwnnw
  \int_value:w \int_div_truncate:nn {#1} {64} ; #1 ;
  \int_value:w \int_div_truncate:nn {#2 - 64 * ( #1 - 2 ) } {64} ; #1 ;
  \int_value:w \int_div_truncate:nn {#5} {64} ; #1 ;
  \exp_after:wN \__str_output_byte:n { #2 - 64 * ( #1 - 2 ) }
}
\end{verbatim}

(End definition for \__str_convert_encode_utf8:, \__str_encode_utf_viii_char:n, and \__str_-
encode_utf_viii_loop:wwnnw.)

\l__str_missing_flag
\l__str_extra_flag
\l__str_overlong_flag
\l__str_overflow_flag

When decoding a string that is purportedly in the \texttt{utf-8} encoding, four different errors can occur, signalled by a specific flag for each (we define those flags using \texttt{\flag_clear-}
\texttt{new:n} rather than \texttt{\flag_new:n}, because they are shared with other encoding definition files).

- “Missing continuation byte”: a leading byte is not followed by the right number of continuation bytes.
- “Extra continuation byte”: a continuation byte appears where it was not expected, \textit{i.e.}, not after an appropriate leading byte.
- “Overlong”: a Unicode character is expressed using more bytes than necessary, for instance, \texttt{“C0”80} for the code point 0, instead of a single null byte.
- “Overflow”: this occurs when decoding produces Unicode code points greater than 1114111.

We only raise one \LaTeX error message, combining all the errors which occurred. In the short message, the leading comma must be removed to get a grammatically correct sentence. In the long text, first remind the user what a correct \texttt{utf-8} string should look like, then add error-specific information.
Invalid UTF-8 string:

```
\exp_last_unbraced:Nf \use_none:n
{\str_if_flag_times:nT {str_missing}{,missing-continuation-byte}
 \str_if_flag_times:nT {str_extra}{,extra-continuation-byte}
 \str_if_flag_times:nT {str_overlong}{,overlong-form}
 \str_if_flag_times:nT {str_overflow}{,code-point-too-large}}
```

In the UTF-8 encoding, each Unicode character consists in 1 to 4 bytes, with the following bit pattern:

```
\lowindent:n
{\Codepoint-<128:0xxxxxxx
 \Codepoint-<2048:110xxxxx\ 10xxxxxx
 \Codepoint-<65536:1110xxxx\ 10xxxxxx\ 10xxxxxx
 \Codepoint-<-1114112:11110xxx\ 10xxxxxx\ 10xxxxxx\ 10xxxxxx
}
```

Bytes of the form 10xxxxxx are called continuation bytes.

```
\str_if_flag_times:nT {str_missing}
{\\
 A leading byte (in the range \[192,255\]) was not followed by the appropriate number of continuation bytes.
}
```

```
\str_if_flag_times:nT {str_extra}
{\\
 LaTeX came across a continuation byte when it was not expected.
}
```

```
\str_if_flag_times:nT {str_overlong}
{\\
 Every Unicode code point must be expressed in the shortest possible form. For instance, \'0xC0\' \'0x83\' is not a valid representation for the code point 3.
}
```

```
\str_if_flag_times:nT {str_overflow}
{\\
 Unicode limits code points to the range \[0,1114111\].
}
```

(End definition for \l__str_missing_flag and others.)
Decoding is significantly harder than encoding. As before, lower some flags, which are tested at the end (in bulk, to trigger at most one \texttt{LaTeX} error, as explained above). We expect successive multi-byte sequences of the form \texttt{\{start byte\} \{continuation bytes\}}. The \texttt{\_start} auxiliary tests the first byte:

- \texttt{[0, \textasciitilde7F]}: the byte stands alone, and is converted to its own character code;
- \texttt{[\textasciitilde80, \textasciitildeBF]}: unexpected continuation byte, raise the appropriate flag, and convert that byte to the replacement character \texttt{\sechd};
- \texttt{[\textasciitildeC0, \textasciitildeFF]}: this byte should be followed by some continuation byte(s).

In the first two cases, \texttt{use\ none\ delimit\ by\ q\ stop:w} removes data that only the third case requires, namely the limits of ranges of Unicode characters which can be expressed with 1, 2, 3, or 4 bytes.

We can now concentrate on the multi-byte case and the \texttt{\_continuation} auxiliary. We expect \texttt{\#3} to be in the range \texttt{[\textasciitilde80, \textasciitildeBF]}. The test for this goes as follows: if the character code is less than \texttt{\~80}, we compare it to \texttt{\~C0}, yielding \texttt{false}; otherwise to \texttt{\~C0}, yielding \texttt{true} in the range \texttt{[\textasciitilde80, \textasciitildeBF]} and \texttt{false} otherwise. If we find that the byte is not a continuation range, stop the current slew of bytes, output the replacement character, and continue parsing with the \texttt{\_start} auxiliary, starting at the byte we just tested. Once we know that the byte is a continuation byte, leave it behind us in the input stream, compute what code point the bytes read so far would produce, and feed that number to the \texttt{\_aux} function.

The \texttt{\_aux} function tests whether we should look for more continuation bytes or not. If the number it receives as \texttt{\#1} is less than the maximum \texttt{\#4} for the current range, then we are done: check for an overlong representation by comparing \texttt{\#1} with the maximum \texttt{\#3} for the previous range. Otherwise, we call the \texttt{\_continuation} auxiliary again, after shifting the “current code point” by \texttt{\#4} (maximum from the range we just checked).

Two additional tests are needed: if we reach the end of the list of range maxima and we are still not done, then we are faced with an overflow. Clean up, and again insert the code point \texttt{\sechd} for the replacement character. Also, every time we read a byte, we need to check whether we reached the end of the string. In a correct UTF-8 string, this happens automatically when the \texttt{\_start} auxiliary leaves its first argument in the input stream: the end-marker begins with \texttt{\prg_break:}, which ends the loop. On the other hand, if the end is reached when looking for a continuation byte, the \texttt{\use\ none: \#3} construction removes the first token from the end-marker, and leaves the \texttt{\_end} auxiliary, which raises the appropriate error flag before ending the mapping.

```latex
\cs_new_protected:cpn { __str_convert_decode_utf8: } {
\begin{Verbatim}
\flag_clear:n { str_error }
\flag_clear:n { str_missing }
\flag_clear:n { str_extra }
\flag_clear:n { str_overlong }
\flag_clear:n { str_overflow }
\__kernel_tl_gset:Nx \g__str_result_tl \__str_decode_utf_viii_start:N \g__str_result_tl
\exp_after:wN \__str_decode_utf_viii_end: }
\prg_break_point:
\__str_if_flag_error:nnx { str_error } { utf8-decode } { }
```

741
54.6.2 utf-16 support

The definitions are done in a category code regime where the bytes 254 and 255 used by
the byte order mark have catcode 12.

When the endianness is not specified, it is big-endian by default, and we add a byte-order
mark. Convert characters one by one in a loop, with different behaviours depending on
the character code.

- [0, "D7FF]: converted to two bytes;
- ["D800", "DFFF"] are used as surrogates: they cannot be converted and are replaced
  by the replacement character;
- ["E000", "FFFF"]: converted to two bytes;
- ["10000", "10FFFF"]: converted to a pair of surrogates, each two bytes. The magic
  "D7C0 is "D800 − "10000)/400.

For the duration of this operation, \_\_str_tmp:w is defined as a function to convert a
number in the range [0, "FFFF"] to a pair of bytes (either big endian or little endian), by
feeding the quotient of the division of #1 by "100, followed by #1 to \_\_str_decode_utf_xvi
be:nn or its le analog: those compute the remainder, and output two bytes for the
quotient and remainder.
When encoding a Unicode string to UTF-16, only one error can occur: code points in the range \[U+D800, U+DFFF\], corresponding to surrogates, cannot be encoded. We use the all-purpose flag `\@error` to signal that error.

When decoding a Unicode string which is purportedly in UTF-16, three errors can occur: a missing trail surrogate, an unexpected trail surrogate, and a string containing an odd number of bytes.
In the UTF-16 encoding, each Unicode character is encoded as 2- or 4-byte encoding. 

- Code point in \[U+0000, U+D7FF\]: two bytes
- Code point in \[U+D800, U+DFFF\]: illegal
- Code point in \[U+E000, U+FFFF\]:
  - a lead surrogate and a trail surrogate
- Lead surrogates are pairs of bytes in the range \[0xD800, 0xDBFF\], and trail surrogates are in the range \[0xDC00, 0xDFFF\].

\flag_if_raised:nT \{ str_missing \}
{ \% \% A lead surrogate was not followed by a trail surrogate.
}
\flag_if_raised:nT \{ str_extra \}
{ \% \% LaTeX came across a trail surrogate when it was not expected.
}
\flag_if_raised:nT \{ str_end \}
{ \% \% The string contained an odd number of bytes. This is invalid:
  the basic code unit for UTF-16 is 16-bits (2-bytes).
}

(End definition for \l__str_missing_flag, \l__str_extra_flag, and \l__str_end_flag.)

As for UTF-8, decoding UTF-16 is harder than encoding it. If the endianness is unknown, check the first two bytes: if those are \"FE and FF in either order, remove them and use the corresponding endianness, otherwise assume big-endianess. The three endianness cases are based on a common auxiliary whose first argument is 1 for big-endian and 2 for little-endian, and whose second argument, delimited by the scan mark \s__str_stop, is expanded once (the string may be long; passing \g__str_result_tl as an argument before expansion is cheaper).

The \__str_decode_utf_xvi:Nw function defines \__str_tmp:w to take two arguments and return the character code of the first one if the string is big-endian, and the second one if the string is little-endian, then loops over the string using \__str_decode_utf_xvi_pair:NN described below.
Bytes are read two at a time. At this stage, \@@_tmp:w #1#2 expands to the character code of the most significant byte, and we distinguish cases depending on which range it lies in:

- \[\text{"D8, "DB}\] signals a lead surrogate, and the integer expression yields 1 (\texttt{-TEX rounds ties away from zero});
- \[\text{"DC, "DF}\] signals a trail surrogate, unexpected here, and the integer expression yields 2;
- any other value signals a code point in the Basic Multilingual Plane, which stands for itself, and the \texttt{-if_case:w} construction expands to nothing (cases other than 1 or 2), leaving the relevant material in the input stream, followed by another call to the _pair auxiliary.

The case of a lead surrogate is treated by the _quad auxiliary, whose arguments #1, #2, #4 and #5 are the four bytes. We expect the most significant byte of #4#5 to be in the range \[\text{"DC, "DF}\] (trail surrogate). The test is similar to the test used for continuation bytes
in the UTF-8 decoding functions. In the case where #4#5 is indeed a trail surrogate, leave
#1#2#4#5 \texttt{\texttt{s\_str (code point) \texttt{s\_str}}, and remove the pair #4#5 before looping with
\texttt{\_str\_decode\_utf\_xvi\_pair:NN}. Otherwise, of course, complain about the missing
surrogate.

The magic number "D7F7 is such that "D7F7×400 = "D800×400+"DC00−"10000.

Every time we read a pair of bytes, we test for the end-marker \q\texttt{\_str\_nil}. When
reaching the end, we additionally check that the string had an even length. Also, if the
end is reached when expecting a trail surrogate, we treat that as a missing surrogate.

\begin{verbatim}
\cs_new:Npn \_str_decode_utf_xvi_pair:NN \#1\#2
{\if_meaning:w \q\texttt{\_str\_nil} \#2
\_str_decode_utf_xvi_pair_end:Nw \#1
\fi:\if_case:w
\int_eval:n { ( \__str_tmp:w \#1\#2 - "D6 ) / 4 } \scan_stop:
\or: \exp_after:wN \_str_decode_utf_xvi_quad:NNwNN \#1\#2
\or: \exp_after:wN \_str_decode_utf_xvi_extra:NNw
\fi: \#1#2 \s\texttt{\_str}
\int_eval:n { "100 * \__str_tmp:w \#1\#2 + \__str_tmp:w \#2\#1 } \s\__str
\_str_decode_utf_xvi_pair:NN
}
\cs_new:Npn \_str_decode_utf_xvi_quad:NNwNN \#1#2 #3 \_str_decode_utf_xvi_pair:NN \#4#5
{\if_meaning:w \q\texttt{\_str\_nil} \#5
\_str_decode_utf_xvi_error:nNN { missing } \#1#2
\_str_decode_utf_xvi_pair_end:Nw \#4
\fi:\if_int_compare:w
\if_int_compare:w \__str_tmp:w \#4#5 < "DC \exp_stop_f:
0 = 1
\else:
\__str_tmp:w \#4#5 < "E0
\fi:
\exp_stop_f:
\if_int_compare:w \__str_tmp:w \#1#2 < "D7F7
\exp_stop_f:
\if_int_compare:w \__str_tmp:w \#1#2 + \__str_tmp:w \#2#1 - "D7F7
+ "100 * \__str_tmp:w \#4#5 + \__str_tmp:w \#5#4
\fi: \#1 \#2 \#4 \#5 \s\texttt{\_str}
\int_eval:n
{ ( "100 * \__str_tmp:w \#1#2 + \__str_tmp:w \#2#1 - "D7F7 ) * "400
+ "100 * \__str_tmp:w \#4#5 + \__str_tmp:w \#5#4
}
\s\texttt{\_str}
\exp_after:wN \use_i:nnn
\else:
\_str_decode_utf_xvi_error:nNN { missing } \#1#2
\fi:
\_str_decode_utf_xvi_pair:NN \#4#5
}
\cs_new:Npn \_str_decode_utf_xvi_pair_end:Nw \#1
{\if_meaning:w \q\texttt{\_str\_nil} \#1
\end{verbatim}
54.6.3 utf-32 support

The definitions are done in a category code regime where the bytes 0, 254 and 255 used by
the byte order mark have catcode “other”.

Convert each integer in the comma-list \g__str_result_tl to a sequence of four bytes. The functions for
big-endian and little-endian encodings are very similar, but
the \_str_output_byte:n instructions are reversed.

(Restore the original catcodes of bytes 254 and 255.)
\exp_args:Nf \__str_encode_utf_xxxii_le_aux:nn
  \{ \int_div_truncate:nn \{#1\} \{"100\} \{#1\} \}
\cs_new:Npn \__str_encode_utf_xxxii_le_aux:nn #1#2
  \{ \__str_output_byte:n { #2 - #1 * "100 } \__str_output_byte_pair_le:n {#1} \}
\^00
\end definition for \__str_convert_encode_utf32: and others.

\flag_clear_new:n { str_overflow }
\flag_clear_new:n { str_end }
\msg_new:nnnn { str } { utf32-decode }
  \{ \Invalid~UTF-32~string:
  \exp_last_unbraced:Nf \use_none:n
  \{ \__str_if_flag_times:nT { str_overflow } { ,~code~point~too~large }
  \__str_if_flag_times:nT { str_end } { ,~truncated~string } \}
\}
\end definition for \str_overflow and \str_end. These variables are documented on page ??

\__str_convert_decode_utf32: \__str_convert_decode_utf32be:
\__str_convert_decode_utf32le:
\__str_decode_utf_xxxii_bom:NNNN
\__str_decode_utf_xxxii:Nw
\__str_decode_utf_xxxii_loop:NNNN
\__str_decode_utf_xxxii_end:w

The structure is similar to UTF-16 decoding functions. If the endianness is not given, test the first 4 bytes of the string (possibly \_\_str_stop if the string is too short) for the presence of a byte-order mark. If there is a byte-order mark, use that endianness, and remove the 4 bytes, otherwise default to big-endian, and leave the 4 bytes in place. The \__str_decode_utf_xxxii:Nw auxiliary receives 1 or 2 as its first argument indicating endianness, and the string to convert as its second argument (expanded or not). It sets

\section*{str Overflow str End}

There can be no error when encoding in UTF-32. When decoding, the string may not have length 4n, or it may contain code points larger than "10FFFF. The latter case often happens if the encoding was in fact not UTF-32, because most arbitrary strings are not valid in UTF-32.

\flag_clear_new:n { str_overflow }
\flag_clear_new:n { str_end }
\msg_new:nnnn { str } { utf32-decode }
  \{ \Invalid~UTF-32~string:
  \exp_last_unbraced:Nf \use_none:n
  \{ \__str_if_flag_times:nT { str_overflow } { ,~code~point~too~large }
  \__str_if_flag_times:nT { str_end } { ,~truncated~string } \}
\}
\end definition for \str_overflow and \str_end. These variables are documented on page ??
\_str\_tmp:w to expand to the character code of either of its two arguments depending on endianness, then triggers the _loop auxiliary inside an x-expanding assignment to \g__str_result_tl.

The _loop auxiliary first checks for the end-of-string marker \s__str_stop, calling the _end auxiliary if appropriate. Otherwise, leave the ⟨4 bytes⟩ \s__str behind, then check that the code point is not overflowing: the leading byte must be 0, and the following byte at most 16.

In the ending code, we check that there remains no byte: there should be nothing left until the first \s__str_stop. Break the map.

\cs_new_protected:cpn { __str_convert_decode_utf32be: }
\cs_new_protected:cpn { __str_decode_utf_xxxii:Nw 1 \g__str_result_tl \s__str_stop }
\cs_new_protected:cpn { __str_convert_decode_utf32le: }
\cs_new_protected:cpn { __str_decode_utf_xxxii:Nw 2 \g__str_result_tl \s__str_stop }
\cs_new_protected:cpn { __str_convert_decode_utf32: }
\exp_after:wN \__str_decode_utf_xxxii_bom:NNNN \g__str_result_tl \s__str_stop \s__str_stop \s__str_stop \s__str_stop
\cs_new_protected:Npn \__str_decode_utf_xxxii_bom:NNNN #1#2#3#4
\str_if_eq:nnTF { #1#2#3#4 } { ^^ff ^^fe ^^00 ^^00 } { \__str_decode_utf_xxxii:Nw 2 }
\str_if_eq:nnTF { #1#2#3#4 } { ^^00 ^^00 ^^fe ^^ff } { \__str_decode_utf_xxxii:Nw 1 }
\__str_decode_utf_xxxii:Nw 1 #1#2#3#4
\cs_new_protected:Npn \__str_decode_utf_xxxii_loop:NNNN #1#2#3#4
\if_meaning:w \s__str_stop #4 \exp_after:wN \__str_decode_utf_xxxii_end:w \fi:
\if_int_compare:w \__str_tmp:w #1#4 > \c_zero_int \flag_raise:n { str_overflow } \flag_raise:n { str_error } \int_use:N \c__str_replacement_char_int \else:
54.7 PDF names and strings by expansion

To convert to PDF names by expansion, we work purely on UTF-8 input. The first step is to make a string with “other” spaces, after which we use a simple token-by-token approach. In Unicode engines, we break down everything before one-byte codepoints, but for 8-bit engines there is no need to worry. Actual escaping is covered by the same code as used in the non-expandable route.
\exp_args:N \__str_convert_pdfname_bytes_aux:n
{ \char_to_utfviii_bytes:n {'#1} }
}
\cs_new:Npn \__str_convert_pdfname_bytes_aux:n #1
{ \__str_convert_pdfname_bytes_aux:nnnn #1 #1}
\cs_new:Npx \__str_convert_pdfname_bytes_aux:nnnn #1#2#3#4
{\c_hash_str \exp_not:N \__str_output_hexadecimal:n {#1}
\exp_not:N \tl_if_blank:nF {#2}
\exp_not:N \tl_if_blank:nF {#3}
\exp_not:N \tl_if_blank:nF {#4}
\c_hash_str \exp_not:N \__str_output_hexadecimal:n {#3}
\exp_not:N \tl_if_blank:nF {#4}
\c_hash_str \exp_not:N \__str_output_hexadecimal:n {#4}
}
\cs_new_eq:NN \__str_convert_pdfname:n \__str_escape_name_char:n
(End definition for str_convert_pdfname:n and others. This function is documented on page 134.)

\package{54.7.1 iso 8859 support}

The iso-8859-1 encoding exactly matches with the 256 first Unicode characters. For other 8-bit encodings of the iso-8859 family, we keep track only of differences, and of unassigned bytes.

\__str_declare_eight_bit_encoding:nnnn { iso88591 } { 256 }
\__str_declare_eight_bit_encoding:nnnn { iso88592 } { 399 }

\__str_declare_eight_bit_encoding:nnnn { iso88591 } { 256 }
{ A1 } { 0104 }
{ A2 } { 02D8 }
{ A3 } { 0141 }
{ A5 } { 013D }
{ A6 } { 015A }
{ A9 } { 0160 }
{ AA } { 015E }
{ AB } { 0164 }
{ AC } { 0179 }
{ AE } { 017D }
{ AF } { 017B }
{ B1 } { 0105 }
{ B2 } { 02DB }
{ B3 } { 0142 }

752
| \{ B5 \} | \{ 013E \} |
| \{ B6 \} | \{ 015B \} |
| \{ B7 \} | \{ 02C7 \} |
| \{ B9 \} | \{ 0161 \} |
| \{ BA \} | \{ 015F \} |
| \{ BB \} | \{ 0165 \} |
| \{ BC \} | \{ 017A \} |
| \{ BD \} | \{ 02DD \} |
| \{ BE \} | \{ 017E \} |
| \{ BF \} | \{ 017C \} |
| \{ CO \} | \{ 0154 \} |
| \{ C3 \} | \{ 0102 \} |
| \{ C5 \} | \{ 0139 \} |
| \{ C6 \} | \{ 0106 \} |
| \{ C8 \} | \{ 010C \} |
| \{ CA \} | \{ 0118 \} |
| \{ CC \} | \{ 011A \} |
| \{ CF \} | \{ 010E \} |
| \{ D0 \} | \{ 0110 \} |
| \{ D1 \} | \{ 0143 \} |
| \{ D2 \} | \{ 0147 \} |
| \{ D5 \} | \{ 015D \} |
| \{ D8 \} | \{ 0158 \} |
| \{ D9 \} | \{ 016E \} |
| \{ DB \} | \{ 0170 \} |
| \{ DE \} | \{ 0162 \} |
| \{ E0 \} | \{ 0155 \} |
| \{ E3 \} | \{ 0103 \} |
| \{ E5 \} | \{ 013A \} |
| \{ E6 \} | \{ 0107 \} |
| \{ E8 \} | \{ 010D \} |
| \{ EA \} | \{ 0119 \} |
| \{ EC \} | \{ 011B \} |
| \{ EF \} | \{ 010F \} |
| \{ F0 \} | \{ 0111 \} |
| \{ F1 \} | \{ 0144 \} |
| \{ F2 \} | \{ 0148 \} |
| \{ F5 \} | \{ 0151 \} |
| \{ F8 \} | \{ 0159 \} |
| \{ F9 \} | \{ 016F \} |
| \{ FB \} | \{ 0171 \} |
| \{ FE \} | \{ 0163 \} |
| \{ FF \} | \{ 02D9 \} |

```python
__str_declare_eight_bit_encoding:nnnn { iso88593 } { 384 }
```
\__str\_declare\_eight\_bit\_encoding:nnnn { iso88594 } { 383 }
```
{ B9 }  { 0161 }  
{ BA }  { 0113 }  
{ BB }  { 0123 }  
{ BC }  { 0167 }  
{ BD }  { 014A }  
{ BE }  { 017E }  
{ BF }  { 014B }  
{ CO }  { 0100 }  
{ C7 }  { 012E }  
{ C8 }  { 010C }  
{ CA }  { 0118 }  
{ CC }  { 0116 }  
{ CF }  { 012A }  
{ D0 }  { 0110 }  
{ D1 }  { 0145 }  
{ D2 }  { 014C }  
{ D3 }  { 0136 }  
{ D9 }  { 0172 }  
{ DD }  { 0168 }  
{ DE }  { 016A }  
{ E0 }  { 0101 }  
{ E7 }  { 012F }  
{ E8 }  { 010D }  
{ EA }  { 0119 }  
{ EC }  { 0117 }  
{ EF }  { 012B }  
{ F0 }  { 0111 }  
{ F1 }  { 0146 }  
{ F2 }  { 014D }  
{ F3 }  { 0137 }  
{ F9 }  { 0173 }  
{ FD }  { 0169 }  
{ FE }  { 016B }  
{ FF }  { 02D9 }  
}  
{  
{  
}  
{/iso88594}  
{iso88595}  
\_strDeclareEightBitEncoding:nnnn { iso88595 } { 374 }  
{  
{ A1 }  { 0401 }  
{ A2 }  { 0402 }  
{ A3 }  { 0403 }  
{ A4 }  { 0404 }  
{ A5 }  { 0405 }  
{ A6 }  { 0406 }  
{ A7 }  { 0407 }  
{ A8 }  { 0408 }  
{ A9 }  { 0409 }  
{ AA }  { 040A }  
{ AB }  { 040B }  
{ AC }  { 040C }  
{ AE }  { 040E }  
```
15130 { AF } { 040F }
15131 { B0 } { 0410 }
15132 { B1 } { 0411 }
15133 { B2 } { 0412 }
15134 { B3 } { 0413 }
15135 { B4 } { 0414 }
15136 { B5 } { 0415 }
15137 { B6 } { 0416 }
15138 { B7 } { 0417 }
15139 { B8 } { 0418 }
15140 { B9 } { 0419 }
15141 { BA } { 041A }
15142 { BB } { 041B }
15143 { BC } { 041C }
15144 { BD } { 041D }
15145 { BE } { 041E }
15146 { BF } { 041F }
15147 { C0 } { 0420 }
15148 { C1 } { 0421 }
15149 { C2 } { 0422 }
15150 { C3 } { 0423 }
15151 { C4 } { 0424 }
15152 { C5 } { 0425 }
15153 { C6 } { 0426 }
15154 { C7 } { 0427 }
15155 { C8 } { 0428 }
15156 { C9 } { 0429 }
15157 { CA } { 042A }
15158 { CB } { 042B }
15159 { CC } { 042C }
15160 { CD } { 042D }
15161 { CE } { 042E }
15162 { CF } { 042F }
15163 { D0 } { 0430 }
15164 { D1 } { 0431 }
15165 { D2 } { 0432 }
15166 { D3 } { 0433 }
15167 { D4 } { 0434 }
15168 { D5 } { 0435 }
15169 { D6 } { 0436 }
15170 { D7 } { 0437 }
15171 { D8 } { 0438 }
15172 { D9 } { 0439 }
15173 { DA } { 043A }
15174 { DB } { 043B }
15175 { DC } { 043C }
15176 { DD } { 043D }
15177 { DE } { 043E }
15178 { DF } { 043F }
15179 { E0 } { 0440 }
15180 { E1 } { 0441 }
15181 { E2 } { 0442 }
15182 { E3 } { 0443 }
15183 { E4 } { 0444 }

756
{ E5 } { 0445 }
{ E6 } { 0446 }
{ E7 } { 0447 }
{ E8 } { 0448 }
{ E9 } { 0449 }
{ EA } { 044A }
{ EB } { 044B }
{ EC } { 044C }
{ ED } { 044D }
{ EE } { 044E }
{ EF } { 044F }
{ F0 } { 2116 }
{ F1 } { 0451 }
{ F2 } { 0452 }
{ F3 } { 0453 }
{ F4 } { 0454 }
{ F5 } { 0455 }
{ F6 } { 0456 }
{ F7 } { 0457 }
{ F8 } { 0458 }
{ F9 } { 0459 }
{ FA } { 045A }
{ FB } { 045B }
{ FC } { 045C }
{ FD } { 00A7 }
{ FE } { 045E }
{ FF } { 045F }

{ AC } { 060C }
{ BB } { 061B }
{ BF } { 061F }
{ C1 } { 0621 }
{ C2 } { 0622 }
{ C3 } { 0623 }
{ C4 } { 0624 }
{ C5 } { 0625 }
{ C6 } { 0626 }
{ C7 } { 0627 }
{ C8 } { 0628 }
{ C9 } { 0629 }
{ CA } { 062A }
{ CB } { 062B }
{ CC } { 062C }
{ CD } { 062D }
{ CE } { 062E }
{ CF } { 062F }
{ D0 } { 0630 }
{ D1 } { 0631 }
\__str_declare_eight_bit_encoding:nnnn { iso88597 } { 498 }

{ A1 } { 2018 }
{ A2 } { 2019 }
{ A4 } { 20AC }
{ A5 } { 20AF }
{ AA } { 037A }
{ AF } { 2015 }
{ B4 } { 0384 }
{ B5 } { 0385 }
{ B6 } { 0386 }
{ B8 } { 0388 }
{ B9 } { 0389 }
{ BA } { 038A }
{ BC } { 038C }
{ BE } { 038E }
{ BF } { 038F }
{ CO } { 0390 }
{ C1 } { 0391 }
{ C2 } { 0392 }
{ C3 } { 0393 }
{ C4 } { 0394 }
{ C5 } { 0395 }
{ C6 } { 0396 }
{ C7 } { 0397 }
{ C8 } { 0398 }
{ C9 } { 0399 }
{ CA } { 039A }
{ CB } { 039B }
{ CC } { 039C }
{ CD } { 039D }
{ CE } { 039E }
{ CF } { 039F }
{ D0 } { 03A0 }
{ D1 } { 03A1 }
{ D3 } { 03A3 }
{ D4 } { 03A4 }
{ D5 } { 03A5 }
{ D6 } { 03A6 }
{ D7 } { 03A7 }
{ D8 } { 03A8 }
{ D9 } { 03A9 }
{ DA } { 03AA }
\__str_declare_eight_bit_encoding:nnnn { \textit{iso88598} } { 308 }

{ AE } { D2 }

\langle iso88598 \rangle

\langle /iso88597 \rangle
\_\_str\_declare\_eight\_bit\_encoding:nnnn { iso885911 } { 369 }

{ A1 } { OE01 }
{ A2 } { OE02 }
{ A3 } { OE03 }
{ A4 } { OE04 }
{ A5 } { OE05 }
{ A6 } { OE06 }
{ A7 } { OE07 }
{ A8 } { OE08 }
{ A9 } { OE09 }
{ AA } { OE0A }
{ AB } { OE0B }
{ AC } { OE0C }
{ AD } { OE0D }
{ AE } { OE0E }
{ AF } { OE0F }
{ B0 } { OE10 }
{ B1 } { OE11 }
{ B2 } { OE12 }
{ B3 } { OE13 }
{ B4 } { OE14 }
{ B5 } { OE15 }
{ B6 } { OE16 }
{ B7 } { OE17 }
{ B8 } { OE18 }
{ B9 } { OE19 }
{ BA } { OE1A }
{ BB } { OE1B }
{ BC } { OE1C }
{ BD } { OE1D }
{ BE } { OE1E }
{ F9 } { 0E59 }
{ FA } { 0E5A }
{ FB } { 0E5B }

{ DB }
{ DC }
{ DD }
{ DE }

{/iso885911}

{iso885913}

\_\_str\_declare\_eight\_bit\_encoding:nnnn { iso885913 } { 399 }

{ A1 } { 201D }
{ A5 } { 201E }
{ A8 } { 00D8 }
{ AA } { 0156 }
{ AF } { 00C6 }
{ B4 } { 201C }
{ B8 } { 00F8 }
{ BA } { 0157 }
{ BF } { 00E6 }
{ CO } { 0104 }
{ C1 } { 012E }
{ C2 } { 0100 }
{ C3 } { 0106 }
{ C6 } { 0118 }
{ C7 } { 0112 }
{ C8 } { 010C }
{ CA } { 0179 }
{ CB } { 0116 }
{ CC } { 0122 }
{ CD } { 0136 }
{ CE } { 012A }
{ CF } { 0138 }
{ D0 } { 0160 }
{ D1 } { 0143 }
{ D2 } { 0145 }
{ D4 } { 014C }
{ D8 } { 0172 }
{ D9 } { 0141 }
{ DA } { 015A }
{ DB } { 016A }
{ DD } { 017B }
{ DE } { 017D }
{ EO } { 0105 }
{ E1 } { 012F }
{ E2 } { 0101 }
{ E3 } { 0107 }
{ E6 } { 0119 }
{ E7 } { 0113 }
{ E8 } { 010D }
{ EA } { 017A }

765
{ EB } { 0117 }
{ EC } { 0123 }
{ ED } { 0137 }
{ EE } { 012B }
{ EF } { 013C }
{ F0 } { 0161 }
{ F1 } { 0144 }
{ F2 } { 0146 }
{ F4 } { 014D }
{ F8 } { 0173 }
{ F9 } { 0142 }
{ FA } { 015B }
{ FB } { 016B }
{ FD } { 017C }
{ FE } { 017E }
{ FF } { 2019 }

{ iso885913 }
{ iso885914 }

\__str DECLARE_EIGHT_BIT_ENCODING:nnnn { iso885914 } { 529 }

{ A1 } { 1E02 }
{ A2 } { 1E03 }
{ A4 } { 010A }
{ A5 } { 010B }
{ A6 } { 1E0A }
{ A8 } { 1E80 }
{ AA } { 1E82 }
{ AB } { 1E8B }
{ AC } { 1E8F }
{ AF } { 0178 }
{ B0 } { 1E1E }
{ B1 } { 1E1F }
{ B2 } { 0120 }
{ B3 } { 0121 }
{ B4 } { 1E40 }
{ B5 } { 1E41 }
{ B7 } { 1E56 }
{ B8 } { 1E81 }
{ B9 } { 1E57 }
{ BA } { 1E83 }
{ BB } { 1E60 }
{ BC } { 1E93 }
{ BD } { 1E84 }
{ BE } { 1E85 }
{ BF } { 1E61 }
{ D0 } { 0174 }
{ D7 } { 1E6A }
{ DE } { 0176 }
{ F0 } { 0175 }
{ F7 } { 1E6B }
{ FE } { 0177 }

766
\__str_declare_eight_bit_encoding\nnnn { iso885915 } { 383 } { 383 }
\__str_declare_eight_bit_encoding\nnnn { iso885916 } { 558 } { 558 }
\__str_declare_eight_bit_encoding\nnnn { iso885914 } { 767 } { 767 }
15776  \{ \text{E3} \} \{ \text{0103} \}
15777  \{ \text{E5} \} \{ \text{0107} \}
15778  \{ \text{F0} \} \{ \text{0111} \}
15779  \{ \text{F1} \} \{ \text{0144} \}
15780  \{ \text{F5} \} \{ \text{0151} \}
15781  \{ \text{F7} \} \{ \text{015B} \}
15782  \{ \text{F8} \} \{ \text{0171} \}
15783  \{ \text{FD} \} \{ \text{0119} \}
15784  \{ \text{FE} \} \{ \text{021B} \}
15785  \}
15786  \}
15787  \}
15788  \langle iso885916 \rangle
Chapter 55

l3quark implementation

The following test files are used for this code: m3quark001.lvt.

55.1 Quarks

\quark_new:N Allocate a new quark.
\cs_new_protected:Npn \quark_new:N #1
{\__kernel_chk_if_free_cs:N #1\cs_gset_nopar:Npn #1 {#1}}

(End definition for \quark_new:N. This function is documented on page 137.)

\q_nil \q_mark \q_no_value \q_stop

Some “public” quarks. \q_stop is an “end of argument” marker, \q_nil is a empty value and \q_no_value marks an empty argument.

\quark_new:N \q_nil \quark_new:N \q_mark \quark_new:N \q_no_value \quark_new:N \q_stop

(End definition for \q_nil and others. These variables are documented on page 137.)

\q_recursion_tail \q_recursion_stop

Quarks for ending recursions. Only ever used there! \q_recursion_tail is appended to whatever list structure we are doing recursion on, meaning it is added as a proper list item with whatever list separator is in use. \q_recursion_stop is placed directly after the list.

\quark_new:N \q_recursion_tail \quark_new:N \q_recursion_stop

(End definition for \q_recursion_tail and \q_recursion_stop. These variables are documented on page 138.)

\s__quark
Private scan mark used in l3quark. We don’t have l3scan yet, so we declare the scan mark here and add it to the scan mark pool later.
\cs_new_eq:NN \s__quark \scan_stop:
Private quark use for some tests.

15803 \quark_new:N \q__quark_nil

When doing recursions, it is easy to spend a lot of time testing if the end marker has been found. To avoid this, a dedicated end marker is used each time a recursion is set up. Thus if the marker is found everything can be wrapped up and finished off. The simple case is when the test can guarantee that only a single token is being tested. In this case, there is just a dedicated copy of the standard quark test. Both a gobbling version and one inserting end code are provided.

15804 \cs_new:Npn \quark_if_recursion_tail_stop:N #1
15805 { \if_meaning:w \q_recursion_tail #1 \exp_after:wN \use_none_delimit_by_q_recursion_stop:w \fi: }
15806 \cs_new:Npn \quark_if_recursion_tail_stop_do:Nn #1
15807 { \if_meaning:w \q_recursion_tail #1 \exp_after:wN \use_i_delimit_by_q_recursion_stop:nw \else: \exp_after:wN \use_none:n \fi: }

See \quark_if_nil:nTF for the details. Expanding \__quark_if_recursion_tail:w once in front of the tokens chosen here gives an empty result if and only if #1 is exactly \q_recursion_tail.

15813 \cs_new:Npn \__quark_if_recursion_tail:w #1 \q_recursion_tail #2 ? #3 ?! { #1 #2 }
15814 \cs_generate_variant:Nn \quark_if_recursion_tail_stop:N { o } \cs_generate_variant:Nn \quark_if_recursion_tail_stop_do:Nn { o }

(End definition for \q__quark_nil.)

(End definition for \quark_if_recursion_tail_stop:N and \quark_if_recursion_tail_stop_do:Nn. These functions are documented on page 138.)
Analogues of the \texttt{\quark_if_recursion_tail_break:NN} functions. Break the mapping using \texttt{#2}.

\begin{verbatim}
\cs_new:Npn \quark_if_recursion_tail_break:NN #1#2
{ \if_meaning:w \q_recursion_tail #1 \exp_after:wN #2 \fi: }
\cs_new:Npn \quark_if_recursion_tail_break:nN #1#2
{ \tl_if_empty:oT { \__quark_if_recursion_tail:w { } #1 {} ?! \q_recursion_tail ??! } {#2} }
\end{verbatim}

(End definition for \texttt{\quark_if_recursion_tail_break:NN} and \texttt{\quark_if_recursion_tail_break:nN}. These functions are documented on page 139.)

\begin{verbatim}
\quark_if_nil_p:N \quark_if_nil:NTF \quark_if_no_value_p:N \quark_if_no_value_p:c \quark_if_no_value_p:N \quark_if_no_value:cTF
\end{verbatim}

Here we test if we found a special quark as the first argument. We better start with \texttt{\q_no_value} as the first argument since the whole thing may otherwise loop if \texttt{#1} is wrongly given a string like \texttt{aabc} instead of a single token.\footnote{It may still loop in special circumstances however!}

\begin{verbatim}
\prg_new_conditional:Npnn \quark_if_nil:N #1 { p, T , F , TF }
{ \if_meaning:w \q_nil #1 \prg_return_true: \else: \prg_return_false: \fi: }
\prg_new_conditional:Npnn \quark_if_no_value:N #1 { p, T , F , TF }
{ \if_meaning:w \q_no_value #1 \prg_return_true: \else: \prg_return_false: \fi: }
\prg_generate_conditional_variant:Nnn \quark_if_no_value:N { c } { p , T , F , TF }
\end{verbatim}

(End definition for \texttt{\quark_if_nil:NTF} and \texttt{\quark_if_no_value:NTF}. These functions are documented on page 137.)

\begin{verbatim}
\quark_if_nil_p:n \quark_if_nil_p:V \quark_if_nil_p:o \quark_if_nil:nTF \quark_if_nil:VTF \quark_if_nil:oTF \quark_if_no_value_p:n \quark_if_no_value:nTF \quark_if_no_value:oTF \__quark_if_nil:w \__quark_if_no_value:w \__quark_if_empty_if:o
\end{verbatim}

Let us explain \texttt{\quark_if_nil:nTF}. Expanding \texttt{\__quark_if_nil:w} once is safe thanks to the trailing \texttt{\q_nil ??!}. The result of expanding once is empty if and only if both delimited arguments \texttt{#1} and \texttt{#2} are empty and \texttt{#3} is delimited by the last tokens \texttt{?!}. Thanks to the leading \texttt{\{} the argument \texttt{#1} is empty if and only if the argument of \texttt{\quark_if_nil:n} starts with \texttt{\q_nil}. The argument \texttt{#2} is empty if and only if this \texttt{\q_nil} is followed immediately by \texttt{?} or by \texttt{\{}?, coming either from the trailing tokens in the definition of \texttt{\quark_if_nil:n}, or from its argument. In the first case, \texttt{\__quark_if_nil:w} is followed by \texttt{\{}\texttt{\q_nil} \texttt{?} \texttt{\q_nil} \texttt{??!}, hence \texttt{#3} is delimited by the final \texttt{?!}, and the test returns \texttt{true} as wanted. In the second case, the result is not empty since...
the first ?! in the definition of \quark_if_nil:n stop #3. The auxiliary here is the same as \_tl_if_empty_if:o, with the same comments applying.

\begin{verbatim}
\prg_new_conditional:Npnn \quark_if_nil:n #1 \{ p, T , F , TF \}
{ \__quark_if_empty_if:o
 { \__quark_if_nil:w {} #1 {} ? ! \q_nil ? ? ! }\prg_return_true:
\else:
 \prg_return_false:
\fi:
}
\cs_new:Npn \__quark_if_nil:w #1 \q_nil #2 ? #3 ? ! { #1 #2 }
\prg_new_conditional:Npnn \quark_if_no_value:n #1 \{ p, T , F , TF \}
{ \__quark_if_empty_if:o
 { \__quark_if_no_value:w {} #1 {} ? ! \q_no_value ? ? ! }\prg_return_true:
\else:
 \prg_return_false:
\fi:
\cs_new:Npn \__quark_if_no_value:w #1 \q_no_value #2 ? #3 ? ! { #1 #2 }
\prg_generate_conditional_variant:Nnn \quark_if_nil:n { V , o } { p , TF , T , F }
\cs_new:Npn \__quark_if_empty_if:o #1
{ \exp_after:wN \if_meaning:w \exp_after:wN \q_nil \__kernel_tl_to_str:w \exp_after:wN {#1} \q_nil }

(End definition for \quark_if_nil:nTF and others. These functions are documented on page 137.)
\end{verbatim}

\__kernel_quark_new_test:N

The function \_kernel_quark_new_test:N defines #1 in a similar way as \quark_if_recursion_tail... functions (as described below), using \q___\langle namespace\rangle_recursion_tail as the test quark and \q___\langle namespace\rangle_recursion_stop as the delimiter quark, where the \langle namespace\rangle is determined as the first \_delimited part in #1.

There are six possible function types which this function can define, and which is defined depends on the signature of the function being defined:

:n gives an analogue of \quark_if_recursion_tail_stop:n
:nn gives an analogue of \quark_if_recursion_tail_stop_do:nn
:nN gives an analogue of \quark_if_recursion_tail_break:nN
:N gives an analogue of \quark_if_recursion_tail_stop:N
:Nn gives an analogue of \quark_if_recursion_tail_stop_do:Nn
:NN gives an analogue of \quark_if_recursion_tail_break:NN

Any other signature causes an error, as does a function without signature.
Similar to \_kernel_quark_new_test:N, but defines quark branching conditionals like \quark_if_nil:nTF that test for the quark \q__⟨namespace⟩⟨name⟩. The ⟨namespace⟩ and ⟨name⟩ are determined from the conditional #1, which must take the rather rigid form \__⟨namespace⟩_quark_if_⟨name⟩:{arg spec}. There are only two cases for the ⟨arg spec⟩ here:

:n gives an analogue of \quark_if_nil:nTF

:N gives an analogue of \quark_if_nil:NTF

Any other signature causes an error, as does a function without signature. We use low-level emptiness tests as \l3tl is not available yet when these functions are used; thankfully we only care about whether strings are empty so a simple \if_meaning:w \q_nil ⟨string⟩ \q_nil suffices.

These macros implement the six possibilities mentioned above, passing the right arguments to \__quark_new_test_aux_do:nNNnnnnNNn, which defines some auxiliaries, and then to \__quark_new_test_define_tl:nNnnnnnNnn (N(n)) variants) or to \__quark_new_test_define_ifx:nNnnnnnNnn (N(n)) which define the main conditionals.

\cs_new_protected:Npn \__quark_new_test_n:Nnnn #1 #2 #3 #4
\__quark_new_test_nN:Nnnn #1 #2 #3 #4
\__quark_new_test_N:Nnnn #1 #2 #3 #4
\__quark_new_test_Nn:Nnnn #1 #2 #3 #4
\__quark_new_test_NN:Nnnn #1 #2 #3 #4
\__quark_new_test_NN:Nnnn #1 #2 #3 #4

(End definition for \__kernel_quark_new_test:N and others.)
\_quark_new_test_aux_do:nNNnnNNn makes the control sequence names which will be used by \_quark_test_define_aux:NNNNnnNNn, and then later by \_quark_new_test_define_tl:nNnNNn or \_quark_new_test_define_ifx:nNnNNn. The control sequences defined here are analogous to \_quark_if_recursion_tail:w and to \_use_(none|i)_delimit_by_q_recursion_stop:(|n)w.

The name is composed by the name-space and the name of the quarks. Suppose \_kernel_quark_new_test:N was used with:

```
\_kernel_quark_new_test:N \_test_quark_tail:w
```

then the first auxiliary will be \_test_quark_recursion_tail:w, and the second one will be \_test_use_none_delimit_by_q_recursion_stop:w.

Note that the actual quarks are not defined here. They should be defined separately using \quark_new:N.

```
\cs_new_protected:Npn \__quark_new_test_aux_do:nNNnnnnNNn #1 #2 #3 #4 #5
\exp_args:Ncc \_quark_test_define_aux:NNNNnnNNn { #1 _quark_recursion_tail:w }
\{ #1 _use_ #4 _delimit_by_q_recursion_stop: #5 w } #2 #3
\}
\cs_new_protected:Npn \_quark_test_define_aux:NNNNnnNNn #1 #2 #3 #4 #5 #6 #7
\cs_gset:Npn #1 ##1 #3 ##2 ? ##3 ?! { ##1 ##2 }
\cs_gset:Npn #2 ##1 #6 #4 {#5} #7 {##1} #1 #2 #3
\}
```

(End definition for \_quark_new_test_aux_do:nNNnnNNn and \_quark_test_define_aux:NNNNnnNNn.)

Finally, these two macros define the main conditional function using what’s been set up before.

```
\cs_new_protected:Npn \__quark_new_test_define_tl:nNnNNn #1 #2 #3 #4 #5 #6
\{ \cs_new:Npn #5 #1
\tl_if_empty:oTF { #2 {} ##1 {} ?! #4 ??! } {#3} {#6}
\}
```

```
\cs_new_protected:Npn \__quark_new_test_define_ifx:nNnNNn #1 #2 #3
\cs_new:Npn #5 #1
\if_meaning:w #4 ##1 \exp_after:wN #3 #6 \fi:
\}
```

```
\cs_new_protected:Npn \__quark_new_test_define_break_tl:nNNNNn {##1##2} #2 {##2}
\cs_new_protected:Npn \__quark_new_test_define_break_ifx:nNNNNn #1 #2 #3
\cs_new_protected:Npn \__quark_new_test_define_break_tl:nNNNNn {##1##2} #2 {##2}
\cs_new_protected:Npn \__quark_new_test_define_break_ifx:nNNNNn #1 #2 #3
```

775
These macros implement the two possibilities for branching quark conditionals, passing the right arguments to \__quark_new_conditional_aux_do:Nnnnn, which defines some auxiliaries and defines the main conditionals.

\__quark_new_conditional_n:Nnnn
\__quark_new_conditional_N:Nnnn

Similar to the previous macros, but branching conditionals only require one auxiliary, so we take a shortcut. In \__quark_new_conditional_define:NNNNn, \texttt{#4} is \use_i:nn to define the \texttt{n}-type function (which needs an auxiliary) and is \use_ii:nn to define the \texttt{N}-type function.

\__quark_module_name:N
\__quark_module_name:w
\__quark_module_name_loop:w
\__quark_module_name_end:w

\__quark_module_name:N takes a control sequence and returns its \langle module \rangle name, determined as the first non-empty non-single-character word, separated by \_ or ::. These rules give the correct result for public functions \langle module \rangle\_\_\_\_, private functions \__-\langle module \rangle\_\_\_\_, and variables such as \l_\langle module \rangle_\_\_. If no valid module is found (such as in \::n) we get the trailing marker \use_none:n \{}, which expands to nothing.
\cs_new:Npn \_quark_module_name_loop:w \#1 \#2
{
  \use_i_ii:nnn \if_meaning:w \prg_do_nothing:
    \#1 \prg_do_nothing: \prg_do_nothing:
    \exp_after:wN \_quark_module_name_loop:w
  \else:
    \_quark_module_name_end:w \#1
    \fi:
}
\cs_new:Npn \_quark_module_name_end:w \#1 \fi: \s__quark { \fi: \#1 }
\exp_after:wN \__quark_tmp:w \tl_to_str:n { : _ }

(End definition for \_quark_module_name:N and others.)

\_quark_quark_conditional_name:N \_quark_quark_conditional_name:w
\_quark_quark_conditional_name:N determines the quark name that the quark conditional function \#1 queries, as the part of the function name between _quark_if_ and the trailing :. Again we define it through \_quark_tmp:w, which receives : as \#1 and _quark_if_ as \#2. The auxiliary \_quark_quark_conditional_name:w returns the part between the first _quark_if_ and the next :, and we apply this auxiliary to the function name followed by : (in case the function name is lacking a signature), and _quark_if_; so that \_quark_quark_conditional_name:N returns an empty string if _quark_if_ is not present.

\cs_set:Npn \_quark_tmp:w \#1 \#2 \s__quark
{
  \cs_new:Npn \_quark_quark_conditional_name:N \#1
  {
    \exp_last_unbraced:Nf \_quark_quark_conditional_name:w
    { \cs_to_str:N \#1 } \#1 \#2 \s__quark
  }
}
\exp_after:wN \_quark_tmp:w \tl_to_str:n { : _quark_if_ } \s__quark
(End definition for \_quark_quark_conditional_name:N and \_quark_quark_conditional_name:w.)

55.2 Scan marks

\g__scan_marks_tl The list of all scan marks currently declared. No \l3tl yet, so define this by hand.
\cs_gset:Npn \g__scan_marks_tl { }
(End definition for \g__scan_marks_tl.)

\scan_new:N Check whether the variable is already a scan mark, then declare it to be equal to \scan_stop: globally.
\cs_new_protected:Npn \scan_new:N \#1
{ \tl_if_in:NnTF \g__scan_marks_tl { \#1 } }
We only declare one scan mark here, more can be defined by specific modules. Can’t use \scan_new:N yet because l3tl isn’t loaded, so define \s_stop by hand and add it to \g__scan_marks_tl. We also add \s__quark (declared earlier) to the pool here. Since it lives in a different namespace, a little \l3docstrip cheating is necessary.

\s_stop
\cs_new:Npn \use_none_delimit_by_s_stop:w #1 \s_stop { }

(End definition for \use_none_delimit_by_s_stop:w. This function is documented on page 140.)
Chapter 56

\textbf{l3seq implementation}

The following test files are used for this code: m3seq002,m3seq003.

\texttt{\package{\@=seq}}

A sequence is a control sequence whose top-level expansion is of the form "\_\_\_\_seq \_\_\_\_item:n \{\langle\text{item}\rangle\} \ldots \_\_\_\_item:n \{\langle\text{item}_n\rangle\}" , with a leading scan mark followed by \textit{n} items of the same form. An earlier implementation used the structure "\seq elt:w \{\langle\text{item}\rangle\} \seq elt_end: \ldots \seq elt:w \{\langle\text{item}_n\rangle\} \seq elt_end:". This allowed rapid searching using a delimited function, but was not suitable for items containing \{,\} and \# tokens, and also lead to the loss of surrounding braces around items

\_\_\_\_item:n \star

\_\_\_\_item:n \{\langle\text{item}\rangle\}

The internal token used to begin each sequence entry. If expanded outside of a mapping or manipulation function, an error is raised. The definition should always be set globally.

\_\_\_\_seq_push_item_def:n
\_\_\_\_seq_push_item_def:x

\_\_\_\_seq_pop_item_def:

\_\_\_\_seq_push_item_def:n \{\langle\text{code}\rangle\}

Saves the definition of \_\_\_\_item:n and redefines it to accept one parameter and expand to \langle\text{code}\rangle. This function should always be balanced by use of \_\_\_\_seq_pop_item_def:n.

\_\_\_\_seq_pop_item_def:

Restores the definition of \_\_\_\_item:n most recently saved by \_\_\_\_seq_push_item_def:n. This function should always be used in a balanced pair with \_\_\_\_seq_push_item_def:n.

\_\_\_\_seq

This private scan mark.

\_\_\_\_seq: \texttt{\scan_new:N \_\_\_\_seq}

(End definition for \_\_\_\_seq.)

\_\_\_\_seq_mark

Private scan marks.

\_\_\_\_seq: \texttt{\scan_new:N \_\_\_\_seq_mark}
\_\_\_\_seq: \texttt{\scan_new:N \_\_\_\_seq_stop}

(End definition for \_\_\_\_seq_mark and \_\_\_\_seq_stop.)
The delimiter is always defined, but when used incorrectly simply removes its argument and hits an undefined control sequence to raise an error.

\cs_new:Npn \__seq_item:n
\msg_expandable_error:nn { seq } { misused }
\use_none:n
(End definition for \__seq_item:n.)

\l__seq_internal_a_tl Scratch space for various internal uses.
(End definition for \l__seq_internal_a_tl and \l__seq_internal_b_tl.)

\__seq_tmp:w Scratch function for internal use.
(End definition for \__seq_tmp:w.)

\c_empty_seq A sequence with no item, following the structure mentioned above.
(End definition for \c_empty_seq. This variable is documented on page 152.)

56.1 Allocation and initialisation

\seq_new:N Sequences are initialized to \c_empty_seq.
(End definition for \seq_new:N. This function is documented on page 141.)

\seq_clear:N Clearing a sequence is similar to setting it equal to the empty one.
(End definition for \seq_clear:N and \seq_gclear:N. These functions are documented on page 141.)

Once again we copy code from the token list functions.
Copying a sequence is the same as copying the underlying token list.

Setting a sequence from a comma-separated list is done using a simple mapping.

Almost identical to \texttt{\seq_set_from_clist:Nn}. These functions are documented on page \pageref{page-142}.

(End definition for \texttt{\seq_clear_new:N} and \texttt{\seq_gclear_new:N}. These functions are documented on page \pageref{page-141}.)
When the separator is empty, everything is very simple, just map \_\_seq_wrap_item:n through the items of the last argument. For non-trivial separators, the goal is to split a given token list at the marker, strip spaces from each item, and remove one set of outer braces if after removing leading and trailing spaces the item is enclosed within braces. After \tl_replace_all:Nnn, the token list \l__seq_internal_a_tl is a repetition of the pattern \_\_seq_set_split_auxi:w \prg_do_nothing: (item with spaces) \_\_seq_set_split_end:. Then, x-expansion causes \_\_seq_set_split_auxi:w to trim spaces, and leaves its result as \_\_seq_set_split_auxii:w (trimmed item) \_\_seq_set_split_end:. This is then converted to the \l3seq internal structure by another x-expansion. In the first step, we insert \prg_do_nothing: to avoid losing braces too early: that would cause space trimming to act within those lost braces. The second step is solely there to strip braces which are outermost after space trimming.
When concatenating sequences, one must remove the leading $\s__seq$ of the second sequence. The result starts with $\s__seq$ (of the first sequence), which stops $f$-expansion.

```latex
\cs_new_protected:Npn \seq_concat:NNN \#1\#2\#3
\{ \tl_set:Nf \#1 { \exp_after:wN \use_i:nn \exp_after:wN \#2 \#3 } \}
\cs_new_protected:Npn \seq_gconcat:NNN \#1\#2\#3
\{ \tl_gset:Nf \#1 { \exp_after:wN \use_i:nn \exp_after:wN \#2 \#3 } \}
\cs_generate_variant:Nn \seq_concat:NNN { ccc }
\cs_generate_variant:Nn \seq_gconcat:NNN { ccc }
```

(End definition for $\seq_concat:NNN$ and $\seq_gconcat:NNN$. These functions are documented on page 142.)

When adding to the left of a sequence, remove $\s__seq$. This is done by $\__seq_put_left_aux:w$, which also stops $f$-expansion.

```latex
\cs_new_protected:Npn \seq_put_left:Nn \#1\#2
\{ \__kernel_tl_set:Nx \#1
\{ \exp_not:n { \s__seq \__seq_item:n {\#2} } \exp_not:f { \exp_after:wN \__seq_put_left_aux:w \#1 } \}
\}
\cs_new_protected:Npn \seq_gput_left:Nn \#1\#2
\{ \__kernel_tl_gset:Nx \#1
\{ \exp_not:n { \s__seq \__seq_item:n {\#2} } \exp_not:f { \exp_after:wN \__seq_put_left_aux:w \#1 } \}
\}
\cs_new:Npn \__seq_put_left_aux:w \s__seq { \exp_stop_f: }
\cs_generate_variant:Nn \seq_put_left:Nn { NV , Nv , No , Nx }
\cs_generate_variant:Nn \seq_gput_left:Nn { c , cV , cv , co , cx }
\cs_generate_variant:Nn \seq_put_left:Nn { c , cV , cv , co , cx }
```

(End definition for $\seq_put_left:Nn$, $\seq_gput_left:Nn$, and $\__seq_put_left_aux:w$. These functions are documented on page 143.)
Since there is no trailing marker, adding an item to the right of a sequence simply means wrapping it in `\__seq_item:n`.

\cs_new_protected:Npn \seq_put_right:Nn #1#2 { \tl_put_right:Nn #1 { \__seq_item:n {#2} } }
\cs_new_protected:Npn \seq_gput_right:Nn #1#2 { \tl_gput_right:Nn #1 { \__seq_item:n {#2} } }
\cs_generate_variant:Nn \seq_gput_right:Nn { NV , Nv , No , Nx }
\cs_generate_variant:Nn \seq_gput_right:Nn { c , cV , cv , co , cx }
\cs_generate_variant:Nn \seq_put_right:Nn { NV , Nv , No , Nx }
\cs_generate_variant:Nn \seq_put_right:Nn { c , cV , cv , co , cx }

(End definition for `\seq_put_right:Nn` and `\seq_gput_right:Nn`. These functions are documented on page 143.)

56.3 Modifying sequences

This function converts its argument to a proper sequence item in an x-expansion context.

\cs_new:Npn \__seq_wrap_item:n #1 { \exp_not:n { \__seq_item:n {#1} } }

(End definition for `\__seq_wrap_item:n`.)

An internal sequence for the removal routines.

\seq_new:N \l__seq_remove_seq

(End definition for `\l__seq_remove_seq`.)

Removing duplicates means making a new list then copying it.

\cs_new_protected:Npn \seq_remove_duplicates:N  { \__seq_remove_duplicates:NN \seq_set_eq:NN }
\cs_new_protected:Npn \seq_gremove_duplicates:N  { \__seq_remove_duplicates:NN \seq_gset_eq:NN }
\cs_new_protected:Npn \__seq_remove_duplicates:NN #1#2 { \seq_clear:N \l__seq_remove_seq \seq_map_inline:Nn #2 { \seq_if_in:NnF \l__seq_remove_seq {##1} { \seq_put_right:Nn \l__seq_remove_seq {##1} } } #1 #2 \l__seq_remove_seq }
\cs_generate_variant:Nn \seq_remove_duplicates:N { c }
\cs_generate_variant:Nn \seq_gremove_duplicates:N { c }

(End definition for `\seq_remove_duplicates:N`, `\seq_gremove_duplicates:N`, and `\__seq_remove_duplicates:NN`. These functions are documented on page 146.)

Removing all means making a new list then copying it.

\cs_new_protected:Npn \seq_remove_all:Nn \seq_remove_all:cn \seq_gremove_all:Nn \seq_gremove_all:cn \__seq_remove_all_aux:NNn

The idea of the code here is to avoid a relatively expensive addition of items one at a time to an intermediate sequence. The approach taken is therefore similar to that in `\__seq_pop_right:NNN`, using a “flexible” x-type expansion to do most of the work. As `\tl_if_eq:nnT` is not expandable, a two-part strategy is needed. First, the x-type expansion uses `\str_if_eq:nnT` to find potential matches. If one is found, the expansion is halted and the necessary set up takes place to use the `\tl_if_eq:nnT` test. The x-type is started
again, including all of the items copied already. This happens repeatedly until the entire sequence has been scanned. The code is set up to avoid needing and intermediate scratch list: the lead-off x-type expansion (#1 #2 {#2}) ensures that nothing is lost.

\begin{verbatim}
\cs_new_protected:Npn \seq_remove_all:Nn \__seq_remove_all_aux:Nn \__kernel_tl_set:Nx \cs_new_protected:Npn \seq_gremove_all:Nn \__seq_remove_all_aux:Nn \__kernel_tl_gset:Nx \cs_new_protected:Npn \__seq_remove_all_aux:NNn #1#2#3
\__seq_push_item_def:n
{ \str_if_eq:nnT {##1} {#3}
  { \if_false: { \fi: }
    \tl_set:Nn \l__seq_internal_b_tl {##1}
    #1 #2
    { \if_false: } \fi:
    \exp_not:o {#2}
    \tl_if_eq:NNT \l__seq_internal_a_tl \l__seq_internal_b_tl
    { \use_none:nn }
  }
  \__seq_wrap_item:n {##1}
}
\tl_set:Nn \l__seq_internal_a_tl {#3}
\__seq_pop_item_def:
\cs_generate_variant:Nn \seq_remove_all:Nn { c }
\cs_generate_variant:Nn \seq_gremove_all:Nn { c }
\end{verbatim}

(End definition for \seq_remove_all:Nn, \seq_gremove_all:Nn, and \__seq_remove_all_aux:NNn. These functions are documented on page 146.)

Previously, \seq_reverse:N was coded by collecting the items in reverse order after an \exp_stop_f: marker.

\begin{verbatim}
\cs_new_protected:Npn \seq_reverse:N \__seq_reverse_item:nwn
\cs_set_eq:NN \@@_item:n \@@_reverse_item:nw
\tl_set:Nf #2 { #2 \exp_stop_f: }
\tl_set:Nf \@@_item:n \l__seq_reverse_item:nw \exp_stop_f: \end{verbatim}

At first, this seems optimal, since we can forget about each item as soon as it is placed after \exp_stop_f:. Unfortunately, \TeX's usual tail recursion does not take place in this case: since the following \__seq_reverse_item:nw only reads tokens until \exp_stop_f:, and never reads the \@@_item:n {#1} left by the previous call, \TeX cannot remove that previous call from the stack, and in particular must retain the various macro parameters in memory, until the end of the replacement text is reached. The stack is thus
only flushed after all the \_seq_reverse_item\_NW are expanded. Keeping track of the arguments of all those calls uses up a memory quadratic in the length of the sequence. \TeX{} can then not cope with more than a few thousand items.

Instead, we collect the items in the argument of \exp_not\_N. The previous calls are cleanly removed from the stack, and the memory consumption becomes linear.

\begin{verbatim}
\cs_new_protected:Npn \seq_reverse:N { \__seq_reverse:NN \__kernel_tl_set:Nx }
\cs_new_protected:Npn \seq_greverse:N { \__seq_reverse:NN \__kernel_tl_gset:Nx }
\cs_new_protected:Npn \__seq_reverse:NN #1 #2 { \cs_set_eq:NN \__seq_tmp:w \__seq_item:n \cs_set_eq:NN \__seq_item:n \__seq_reverse_item:nwn #1 #2 { \exp_not:n { } } \cs_set_eq:NN \__seq_item:n \__seq_tmp:w }
\cs_generate_variant:Nn \seq_reverse:N { c }
\cs_generate_variant:Nn \seq_greverse:N { c }
\end{verbatim}

(End definition for \seq_reverse:N and others. These functions are documented on page 146.)

\begin{verbatim}
\seq_sort:Nn \seq_sort:cn \seq_gsort:Nn \seq_gsort:cn
\cs_new:Npn \__seq_shuffle:NN #1 #2 \exp_not:n \__seq_shuffle_item:n
\g__seq_internal_seq
\end{verbatim}

We apply the Fisher–Yates shuffle, storing items in \toks registers. We use the primitive \tex_uniformdeviate:D for speed reasons. Its non-uniformity is of order its argument divided by $2^{28}$, not too bad for small lists. For sequences with more than 13 elements there are more possible permutations than possible seeds ($13! > 2^{28}$) so the question of uniformity is somewhat moot. The integer variables are declared in \l3int: load-order issues.

\begin{verbatim}
\cs_if_exist:NTF \tex_uniformdeviate:D
\end{verbatim}

56.4 Sequence conditionals

\seq_if_empty_p:N \seq_if_empty_p:c \seq_if_empty:NTF \seq_if_empty:cTF

Similar to token lists, we compare with the empty sequence.

\begin{verbatim}
\prg_new_conditional:Npnn \seq_if_empty:N #1 { p , T , F , TF }
\seq_if_empty:NTF { \if_meaning:w #1 \c_empty_seq \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End definition for \seq_if_empty:NTF. This function is documented on page 146.)

\begin{verbatim}
\seq_shuffle:N \seq_shuffle:c \seq_gshuffle:N \seq_gshuffle:c \__seq_shuffle:NN \_seq_shuffle_item:n \g__seq_internal_seq
\end{verbatim}
The approach here is to define \_\_seq_item:n to compare its argument with the test sequence. If the two items are equal, the mapping is terminated and \group_end: \prg_return_true: is inserted after skipping over the rest of the recursion. On the other hand, if there is no match then the loop breaks, returning \prg_return_false:. Everything is inside a group so that \_\_seq_item:n is preserved in nested situations.
56.5 Recovering data from sequences

The two pop functions share their emptiness tests. We also use a common emptiness test for all branching get and pop functions.

```latex
\texttt{\textcolor{red}{\textbf{\_seq_pop:NNNN}} \texttt{\_seq_pop_TF:NNNN}}
```

Getting an item from the left of a sequence is pretty easy: just trim off the first item after \_\texttt{\_seq_item:n} at the start. We append a \texttt{\_no_value} item to cover the case of an empty sequence.

```latex
\texttt{\textcolor{red}{\textbf{\_seq_get_left:NN}} \texttt{\_seq_get_left:cN}}
```

\texttt{\_seq_get_left:wnw}

(End definition for \texttt{\_seq_pop:NNNN} and \texttt{\_seq_pop_TF:NNNN}.)

(End definition for \texttt{\_seq_pop:NNNN} and \texttt{\_seq_pop_TF:NNNN}.)
\__kernel_tl_set:Nx #2
{ \exp_after:wN \__seq_get_left:wnw
  \__seq_item:n { \q_no_value } \s__seq_stop
}
\cs_new:Npn \__seq_get_left:wnw #1 \__seq_item:n #2#3 \s__seq_stop
{ \exp_not:n {#2} }
\cs_generate_variant:Nn \seq_get_left:NN { c }

\__seq_pop_left:NNN #1#2#3
{ \exp_after:wN \__seq_pop_left:wnwNNN #2 \s__seq_stop #1#2#3 }
\cs_generate_variant:Nn \seq_pop_left:NN { c }
\cs_generate_variant:Nn \seq_gpop_left:NN { c }

\\__seq_get_left:wnw
\end definition for \seq_get_left:NN and \__seq_get_left:wnw. This function is documented on page 143.

\\seq_pop_left:NN \seq_gpop_left:NN
\seq_get_right:NN \seq_get_right:cN
\__seq_get_right_loop:nw
\__seq_get_right_end:NnN
\end definition for \seq_get_left:NN and others. These functions are documented on page 143.

The approach to popping an item is pretty similar to that to get an item, with the only difference being that the sequence itself has to be redefined. This makes it more sensible to use an auxiliary function for the local and global cases.

\cs_new_protected:Npn \seq_pop_left:NN
{ \__seq_pop:NNNN \__seq_pop_left:NNN \tl_set:Nn }
\cs_new_protected:Npn \seq_gpop_left:NN
{ \__seq_pop:NNNN \__seq_pop_left:NNN \tl_gset:Nn }
\cs_new_protected:Npn \__seq_pop_left:NNN #1#2#3
{ \exp_after:wN \__seq_pop_left:wnwNNN #2 \s__seq_stop #1#2#3 }
\cs_new_protected:Npn \__seq_pop_left:wnwNNN
#1 \__seq_item:n #2#3 \s__seq_stop #4#5#6
{ #4 #5 { #1 #3 }
  \tl_set:Nn #6 {#2} }
\cs_generate_variant:Nn \seq_pop_left:NN { c }
\cs_generate_variant:Nn \seq_gpop_left:NN { c }

\\seq_get_right:NN \seq_get_right:cN
\__seq_get_right_loop:nw
\__seq_get_right_end:NnN
First remove \s__seq and prepend \q_no_value. The first argument of \__seq_get_right_loop:nw is the last item found, and the second argument is empty until the end of the loop, where it is code that applies \exp_not:n to the last item and ends the loop.

\cs_new_protected:Npn \seq_get_right:NN \_#2
{ \\kernel_tl_set:Nx \_2
  \exp_after:wN \use_i_ii:nnn
  \exp_after:wN \__seq_get_right_loop:nw
  \exp_after:wN \q_no_value
  \_1
  \\__seq_get_right_end:NnN \\__seq_item:n
}
\cs_new:Npn \__seq_get_right_end:NnN #1#2#3
\exp_not:n {#2}
\cs_generate_variant:Nn \seq_get_right:NN { c }
\cs_generate_variant:Nn \seq_gpop_left:NN { c }

789
The approach to popping from the right is a bit more involved, but does use some of the same ideas as getting from the right. What is needed is a “flexible length” way to set a token list variable. This is supplied by the { \if_false: } \fi: \...\if_false: { \fi: } construct. Using an x-type expansion and a “non-expanding” definition for \__seq_item:n, the left-most \( n - 1 \) entries in a sequence of \( n \) items are stored back in the sequence. That needs a loop of unknown length, hence using the strange \if_false: way of including braces. When the last item of the sequence is reached, the closing brace for the assignment is inserted, and \tl_set:Nn \#3 is inserted in front of the final entry. This therefore does the pop assignment. One more iteration is performed, with an empty argument and \use_none:nn, which finally stops the loop.

Getting from the left or right with a check on the results. The first argument to \__seq_-pop_TF:NNNN is left unused.

(End definition for \seq_get_right:NNN, \__seq_get_right_loop:nn, and \__seq_get_right_end:NnN. This function is documented on page 143.)

(End definition for \seq_pop_right:NN and others. These functions are documented on page 144.)
More or less the same for popping.

\seq_pop_left:NNTF
\seq_pop_left:NN
\seq_gpop_left:NNTF
\seq_gpop_left:NN
\seq_pop_right:NNTF
\seq_pop_right:NN
\seq_gpop_right:NNTF
\seq_gpop_right:NN

The idea here is to find the offset of the item from the left, then use a loop to grab the correct item. If the resulting offset is too large, then the argument delimited by \_\_seq_item:nwN is \prg_break: instead of being empty, terminating the loop and returning nothing at all.

\seq_item:Nn
\seq_item:cn
\_\_seq_item:wNn
\_\_seq_item:nN
\_\_seq_item:nwn
\_\_seq_item:nwn

(End definition for \seq_get_left:NNTF and \seq_get_right:NNTF. These functions are documented on page 145.)
Importantly, \texttt{\seq_item:Nn} only evaluates its argument once.

\begin{verbatim}
\cs_new:Npn \seq_rand_item:N #1
{ \seq_if_empty:NF #1
{ \seq_item:Nn #1 { \int_rand:nn { 1 } \seq_count:N #1 } } 
}
\cs_generate_variant:Nn \seq_rand_item:N { c }
\end{verbatim}

\subsection*{56.6 Mapping over sequences}

To break a function, the special token \texttt{\prg_break_point:Nn} is used to find the end of the code. Any ending code is then inserted before the return value of \texttt{\seq_map_break:n} is inserted.

\begin{verbatim}
\cs_new:Npn \seq_map_break:
{ \prg_map_break:Nn \seq_map_break: { } }
\cs_new:Npn \seq_map_break:n
{ \prg_map_break:Nn \seq_map_break: }
\end{verbatim}

The idea here is to apply the code of \texttt{#2} to each item in the sequence without altering the definition of \texttt{\_\_seq_item:n}. The even-numbered arguments of \texttt{\_\_seq_map_function:Nw} delimited by \texttt{\_\_seq_item:n} are almost always empty, except at the end of the loop where it is \texttt{\prg_break:}. This allows to break the loop without needing to do a (relatively-expensive) quark test.

\begin{verbatim}
\cs_new:Npn \_\_seq_map_function:Nw #1#2 #3 #4 #5 #6 #7 #8 #9
{ #2 #1 {#3} #4 #1 {#5} #6 #1 {#7} #8 #1 {#9} \_\_seq_map_function:Nw #1 }
\cs_generate_variant:Nn \_\_seq_map_function:Nw { c }
\end{verbatim}
The definition of \_\_seq_item:n needs to be saved and restored at various points within the mapping and manipulation code. That is handled here: as always, this approach uses global assignments.

\begin{verbatim}
\cs_new_protected:Npn \_\_seq_push_item_def:n
\{ \_\_seq_push_item_def:
\cs_gset:Npn \_\_seq_item:n ##1
\}
\cs_new_protected:Npn \_\_seq_push_item_def:x
\{ \_\_seq_push_item_def:
\cs_gset:Npx \_\_seq_item:n ##1
\}
\cs_new_protected:Npn \_\_seq_push_item_def:
\{ \int_gincr:N \g__kernel_prg_map_int
\cs_gset_eq:cN { __seq_map_ \int_use:N \g__kernel_prg_map_int :w }
\_\_seq_item:n
\}
\cs_new_protected:Npn \_\_seq_pop_item_def:
\{ \cs_gset_eq:Nc \_\_seq_item:n
\{ __seq_map_ \int_use:N \g__kernel_prg_map_int :w \}
\int_gdecr:N \g__kernel_prg_map_int
\}
\end{verbatim}

(End definition for \_\_seq_push_item_def:n, \_\_seq_push_item_def:, and \_\_seq_pop_item_def:.)

\texttt{\textbackslash seq\textunderscore map\textunderscore inline:Nn}

The idea here is that \_\_seq_item:n is already “applied” to each item in a sequence, and so an in-line mapping is just a case of redefining \_\_seq_item:n.

\begin{verbatim}
\cs_new_protected:Npn \seq_map_inline:Nn #1#2
\{ \_\_seq_push_item_def:n {#2}
\prg_break_point:Nn \seq_map_break: { \_\_seq_pop_item_def: }
\}
\cs_generate_variant:Nn \seq_map_inline:Nn { c }
\end{verbatim}

(End definition for \texttt{\textbackslash seq\textunderscore map\textunderscore inline:Nn}. This function is documented on page 147.)

\texttt{\textbackslash seq\textunderscore map\textunderscore tokens:Nn}

This is based on the function mapping but using the same tricks as described for \texttt{\textbackslash prop\textunderscore map\textunderscore tokens:Nn}. The idea is to remove the leading \s__seq and apply the tokens such that they are safe with the break points, hence the \texttt{\textbackslash use:n}.

\begin{verbatim}
\cs_new:Npn \seq_map_tokens:Nn #1#2
\exp_last_unbraced:Nno \use_i:nn { __seq_map_tokens:nw {#2} } #1
\prg_break:
\_\_seq_item:n \{ \_\_seq_item:n \{ \_\_seq_item:n \{ \_\_seq_item:n \{ \_\_seq_item:n \{ \_\_seq_item:n \{ \prg_break_point:
\prg_break_point:Nn \seq_map_break: { }
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}
\}

793
\seq_map_variable:NNn
\seq_map_variable:Ncn
\seq_map_variable:cNn
\seq_map_variable:ccn

This is just a specialised version of the in-line mapping function, using an x-type expansion for the code set up so that the number of \# tokens required is as expected.

\cs_new_protected:Npn \seq_map_variable:NNn #1#2#3
{\__seq_push_item_def:x
{\tl_set:Nn \exp_not:N #2 {##1}\exp_not:n {#3}}
#1\prg_break_point:Nn \seq_map_break: { \__seq_pop_item_def: }
}
\cs_generate_variant:Nn \seq_map_variable:NNn { Nc }
\cs_generate_variant:Nn \seq_map_variable:NNn { c , cc }

(End definition for \seq_map_variable:NNn. This function is documented on page 147.)

\seq_map_indexed_function:NN
\seq_map_indexed_inline:Nn
\__seq_map_indexed:nNN
\__seq_map_indexed:Nw

Similar to \seq_map_function:NN but we keep track of the item index as a ; delimited argument of \__seq_map_indexed:Nw.

\cs_new:Npn \seq_map_indexed_function:NN #1#2
{\__seq_map_indexed:NN #1#2\prg_break_point:Nn \seq_map_break: { }}
\cs_new_protected:Npn \seq_map_indexed_inline:Nn #1#2
{\int_gincr:N \g__kernel_prg_map_int \cs_gset_protected:cpn
{ \__seq_map_ \int_use:N \g__kernel_prg_map_int :w } #1\#2 \#2\exp_args:NNc \__seq_map_indexed:NN #1
\prg_break_point:Nn \seq_map_break:
{ \int_gdecr:N \g__kernel_prg_map_int }
}
\cs_new:Npn \__seq_map_indexed:Nw #1
{\int_gincr:N \g__kernel_prg_map_int \cs_gset_protected:cpn
{ \__seq_map_ \int_use:N \g__kernel_prg_map_int :w } \#1\#2 \#2
\exp_args:NNc \__seq_map_indexed:NW #1
\prg_break_point:Nn \seq_map_break:
{ \int_gdecr:N \g__kernel_prg_map_int }\}
\cs_new:Npn \__seq_map_tokens:Nn \seq_map_tokens:Nn { c }
\cs_new:Npn \__seq_map_tokens:nw #1
#2 \__seq_item:n #3
#4 \__seq_item:n #5
#6 \__seq_item:n #7
#8 \__seq_item:n #9
}{
#2 \use:n {#1} {#3}
#4 \use:n {#1} {#5}
#6 \use:n {#1} {#7}
#8 \use:n {#1} {#9}
\__seq_map_tokens:nw {#1}
}

(End definition for \seq_map_tokens:Nn and \__seq_map_tokens:nw. This function is documented on page 147.)

794
Very similar to \seq_set_filter:Nnn. We could actually merge the two within a single function, but it would have weird semantics.

\seq_set_map_x:NNn
\seq_gset_map_x:NNn
\__seq_set_map_x:NNNn

Similar to \seq_set_map_x:NNn, but prevents expansion of the <inline function>.

\seq_count:N
\seq_count:c
\__seq_count:w
\__seq_count_end:w

Since counting the items in a sequence is quite common, we optimize it by grabbing 8 items at a time and correspondingly adding 8 to an integer expression. At the end of the loop, #9 is \__seq_count_end:w instead of being empty. It removes 8+ and instead
places the number of \_\_seq_item:n that \_\_seq_count:w grabbed before reaching the end of the sequence.

\cs_new:Npn \seq_count:N \#1
\int_eval:n
\exp_after:wN \use_i:nn
\exp_after:wN \_\_seq_count:w
#1
\_\_seq_count:end:w \_\_seq_item:n 7
\_\_seq_count:end:w \_\_seq_item:n 6
\_\_seq_count:end:w \_\_seq_item:n 5
\_\_seq_count:end:w \_\_seq_item:n 4
\_\_seq_count:end:w \_\_seq_item:n 3
\_\_seq_count:end:w \_\_seq_item:n 2
\_\_seq_count:end:w \_\_seq_item:n 1
\_\_seq_count:end:w \_\_seq_item:n 0
\prg_break_point:
\end{definition}

56.7 Using sequences

\seq_use:Nnnn
\seq_use:cn
\seq_use:Nnn
\seq_use_setup:w
\seq_use:nwwn
\seq_use:Nn
\seq_use:cn

See \clist_use:Nnnn for a general explanation. The main difference is that we use \_\_seq_item:n as a delimiter rather than commas. We also need to add \_\_seq_item:n at various places, and \s__seq.

\cs_new:Npn \seq_use:Nnnn \#1\#2\#3\#4
\seq_if_exist:NTF \#1
\int_case:nnF { \seq_count:N \#1 }
\prg_break_point: \{#1\}
\cs_generate_variant:Nn \seq_count:N { c }

(End definition for \seq_count:N, \_\_seq_count:w, and \_\_seq_count_end:w. This function is documented on page 149.)
56.8 Sequence stacks

The same functions as for sequences, but with the correct naming.

\seq_push:Nn
\seq_push:NV
\seq_push:Nv
\seq_push:No
\seq_push:Nx
\seq_push:cn
\seq_push:cV
\seq_push:co
\seq_push:cx
\seq_gpush:Nn
\seq_gpush:NV
\seq_gpush:Nv
\seq_gpush:No
\seq_gpush:Nx
\seq_gpush:cn
\seq_gpush:cV
\seq_gpush:co
\seq_gpush:cx

(End definition for \seq_push:Nn and \seq_gpush:Nn. These functions are documented on page 151.)

\seq_get:NN
\seq_get:cN
\seq_pop:NN
\seq_pop:cN
\seq_gpop:NN
\seq_gpop:cN

In most cases, getting items from the stack does not need to specify that this is from the left. So alias are provided.
56.9 Viewing sequences

Apply the general \_kernel_chk_tl_type:NnnT.

(End definition for \seq_get:N, \seq_pop:N, and \seq_gpop:N. These functions are documented on page 150.)

(End definition for \seq_get:NNTF, \seq_pop:NNTF, and \seq_gpop:NNTF. These functions are documented on page 150.)

(End definition for \seq_get:NN, \seq_pop:NN, and \seq_gpop:NN. These functions are documented on page 150.)

(End definition for \seq_get:NNTF, \seq_pop:NNTF, and \seq_gpop:NNTF. These functions are documented on page 153.)

(End definition for \seq_show:N and others. These functions are documented on page 155.)
56.10 Scratch sequences

\l_tmpa_seq  Temporary comma list variables.
\l_tmpb_seq
\g_tmpa_seq
\g_tmpb_seq

(End definition for \l_tmpa_seq and others. These variables are documented on page 155.)

⟨/package⟩
Chapter 57

\l3int implementation

The following test files are used for this code: m3int001, m3int002, m3int03.

\c_max_register_int

Done in \l3basics.

(End definition for \c_max_register_int. This variable is documented on page 165.)

\__int_to_roman:w
\if_int_compare:w

Done in \l3basics.

(End definition for \__int_to_roman:w and \if_int_compare:w. This function is documented on page 166.)

\or:

Done in \l3basics.

(End definition for \or:. This function is documented on page 166.)

\int_value:w
\__int_eval:w
\__int_eval_end:
\if_int_odd:w
\if_case:w

Here are the remaining primitives for number comparisons and expressions.

(End definition for \int_value:w and others. These functions are documented on page 166.)

\s__int_mark
\s__int_stop

Scan marks used throughout the module.

(End definition for \s__int_mark and \s__int_stop.)

\__int_use_none_delimit_by_s_stop:w

Function to gobble until a scan mark.

(End definition for \__int_use_none_delimit_by_s_stop:w.)

\q__int_recursion_tail
\q__int_recursion_stop

Quarks for recursion.

(End definition for \q__int_recursion_tail and \q__int_recursion_stop.)
Functions to query quarks.

\__int_recursion_tail_stop_do:Nn \__int_recursion_tail_stop:N

(End definition for \__int_recursion_tail and \__int_recursion_stop.)

57.1 Integer expressions

\int_eval:n \int_eval:w

Wrapper for \__int_eval:w: can be used in an integer expression or directly in the input stream. It is very slightly faster to use \the rather than \number to turn the expression to a number. When debugging, we introduce parentheses to catch early termination (see \textdebug).

\cs_new:Npn \int_eval:n #1
\cs_new:Npn \int_eval:w { \tex_the:D \__int_eval:w #1 \__int_eval_end: }

(End definition for \int_eval:n and \int_eval:w. These functions are documented on page 155.)

\int_sign:n \__int_sign:Nw

See \int_abs:n. Evaluate the expression once (and when debugging is enabled, check that the expression is well-formed), then test the first character to determine the sign. This is wrapped in \int_value:w... \exp_stop_f: to ensure a fixed number of expansions and to avoid dealing with closing the conditionals.

\cs_new:Npn \int_abs:n #1
\cs_new:Npn \__int_abs:N #1 { \if_meaning:w - #1 \else: \exp_after:wN #1 \fi: }

(End definition for \int_sign:n and \__int_sign:Nw. This function is documented on page 155.)

\int_abs:n \__int_abs:N \int_max:nn \int_min:nn \__int_maxmin:wwN

Functions for min, max, and absolute value with only one evaluation. The absolute value is obtained by removing a leading sign if any. All three functions expand in two steps.

\cs_new:Npn \int_abs:n #1
\cs_new:Npn \__int_abs:N #1 { \if_meaning:w - #1 \else: \exp_after:wN #1 \fi: }

(End definition for \int_abs:n and \__int_abs:N. This function is documented on page 155.)
As \texttt{\_\_int_eval:w} rounds the result of a division we also provide a version that truncates the result. We use an auxiliary to make sure numerator and denominator are only evaluated once: this comes in handy when those are more expressions are expensive to evaluate (e.g., \texttt{\tl_count:n}). If the numerator \#1\#2 is 0, then we divide 0 by the denominator (this ensures that 0/0 is correctly reported as an error). Otherwise, shift the numerator \#1\#2 towards 0 by (|\#3\#4|−1)/2, which we round away from zero. It turns out that this quantity exactly compensates the difference between \LaTeX's rounding and the truncating behaviour that we want. The details are thanks to Heiko Oberdiek: getting things right in all cases is not so easy.

\section{\texttt{\_\_int_div_truncate:nn}}
\section{\texttt{\_\_int_mod:nn}}
\section{\texttt{\_\_int_div_truncate:NwNw}}
\section{\texttt{\_\_int_mod:ww}}

\end{definition}

\endinput
For the sake of completeness:

\cs_new:Npn \int_div_round:nn #1#2
\{ \int_value:w \__int_eval:w ( #1 ) / ( #2 ) \__int_eval_end: \}

Finally there's the modulus operation.

\cs_new:Npn \int_mod:nn #1#2
\{ \int_value:w \__int_eval:w \exp_after:wN \__int_mod:ww
\int_value:w \__int_eval:w #1 \exp_after:wN ;
\int_value:w \__int_eval:w #2 ;
\__int_eval_end: \}
\cs_new:Npn \__int_mod:ww #1; #2;
{ #1 - ( \__int_div_truncate:NwNw #1 ; #2 ; ) * #2 }

(End definition for \int_div_truncate:nn and others. These functions are documented on page 155.)

\kernel_int_add:nnnn

Equivalent to \int_eval:n {#1+#2+#3} except that overflow only occurs if the final result overflows \([-2^{31} + 1, 2^{31} - 1\). The idea is to choose the order in which the three numbers are added together. If \#1 and \#2 have opposite signs (one is in \([-2^{31} + 1, -1\] and the other in \([0, 2^{31} - 1\)) then \#1+\#2 cannot overflow so we compute the result as \#1+\#2+\#3. If they have the same sign, then either \#3 has the same sign and the order does not matter, or \#3 has the opposite sign and any order in which \#3 is not last will work. We use \#1+\#3+\#2.

\cs_new:Npn \int_new:N #1
\{ \__kernel_chk_if_free_cs:N #1
\cs:w newcount \cs_end: #1 \}
\cs_generate_variant:Nn \int_new:N { c }

(End definition for \__kernel_int_add:nnnn.)

57.2 Creating and initialising integers

\int_new:N
\int_new:c

Two ways to do this: one for the format and one for the \LaTeX{} package. In plain \TeX, \newcount (and other allocators) are \texttt{outer}: to allow the code here to work in "generic" mode this is therefore accessed by name. (The same applies to \newbox, \newdimen and so on.)

\cs_new_protected:Npn \int_new:N #1
\{ \__kernel_chk_if_free_cs:N #1
\cs:w newcount \cs_end: \#1 \}
\cs_generate_variant:Nn \int_new:N { c }

(End definition for \int_new:N. This function is documented on page 156.)
As stated, most constants can be defined as \texttt{chardef} or \texttt{mathchardef} but that’s engine dependent. As a result, there is some set up code to determine what can be done. No full engine testing just yet so everything is a little awkward. We cannot use \texttt{int_gset:Nn} because (when \texttt{check-declarations} is enabled) this runs some checks that constants would fail.

\begin{verbatim}
\cs_new_protected:Npn \int_const:Nn #1 #2 { \exp_args:Nx \__int_const:nN { \int_eval:n {#2} } #1 }
\cs_new_protected:Npn \__int_const:nN #1 #2 {
    \int_compare:nNnTF {#1} < \c_zero_int {
        \int_new:N #2 \tex_global:D
    } {
        \int_compare:nNnTF {#1} > \c__int_max_constdef_int {
            \int_new:N #2 \tex_global:D
        } {
            \__kernel_chk_if_free_cs:N #2 \tex_global:D \__int_constdef:Nw
        }
    }
    \__int_eval:w #1 \__int_eval_end:
}
\cs_generate_variant:Nn \int_const:Nn { c }
\if_int_odd:w 0 \cs_if_exist:NT \tex_luatexversion:D { 1 } \cs_if_exist:NT \tex_omathchardef:D { 1 } \cs_if_exist:NTF \tex_omathchardef:D { \cs_new_eq:NN \__int_constdef:Nw \tex_omathchardef:D } \else: \cs_new_eq:NN \__int_constdef:Nw \tex_mathchardef:D \tex_mathchardef:D \c__int_max_constdef_int 32767 - \fi:
\end{verbatim}

Functions that reset an \texttt{integer} register to zero.

\begin{verbatim}
\cs_new_protected:Npn \int_zero:N #1 { #1 = \c_zero_int }
\cs_new_protected:Npn \int_gzero:N #1 { \tex_global:D #1 = \c_zero_int }
\cs_generate_variant:Nn \int_zero:N { c }
\cs_generate_variant:Nn \int_gzero:N { c }
\end{verbatim}

Create a register if needed, otherwise clear it.
\cs_new_protected:Npn \int_gzero_new:N #1{\int_if_exist:NTF #1{\int_gzero:N #1}{\int_new:N #1}}
\cs_generate_variant:Nn \int_zero_new:N { c }
\cs_generate_variant:Nn \int_gzero_new:N { c }

(End definition for \int_zero_new:N and \int_gzero_new:N. These functions are documented on page 156.)

\int_set_eq:NN \int_set_eq:cN \int_set_eq:Nc \int_set_eq:cc \int_gset_eq:NN \int_gset_eq:cN \int_gset_eq:Nc \int_gset_eq:cc

Setting equal means using one integer inside the set function of another. Check that assigned integer is local/global. No need to check that the other one is defined as TeX does it for us.

\cs_new_protected:Npn \int_set_eq:NN #1#2{#1 = #2}
\cs_generate_variant:Nn \int_set_eq:NN { c , Nc , cc }
\cs_new_protected:Npn \int_gset_eq:NN #1#2{\tex_global:D #1 = #2}
\cs_generate_variant:Nn \int_gset_eq:NN { c , Nc , cc }

(End definition for \int_set_eq:NN and \int_gset_eq:NN. These functions are documented on page 156.)

\int_if_exist_p:N \int_if_exist_p:c \int_if_exist:N \int_if_exist:c \int_if_exist:TF \int_if_exist:TF

Copies of the \cs functions defined in l3basics.

\prg_new_eq_conditional:NNn \int_if_exist:N \cs_if_exist:N { TF , T , F , p }
\prg_new_eq_conditional:NNn \int_if_exist:c \cs_if_exist:c { TF , T , F , p }

(End definition for \int_if_exist:NTF. This function is documented on page 156.)

57.3 Setting and incrementing integers

\int_add:Nn \int_add:cn \int_gadd:Nn \int_gadd:cn \int_sub:Nn \int_sub:cn \int_gsub:Nn \int_gsub:cn

Adding and subtracting to and from a counter. Including here the optional \by would slow down these operations by a few percent.

\cs_new_protected:Npn \int_add:N #1{\tex_advance:D #1 \c_one_int}
\cs_new_protected:Npn \int_decr:N #1{\tex_advance:D #1 - \c_one_int}
\cs_new_protected:Npn \int_gincr:N #1{\tex_global:D \tex_advance:D #1 \c_one_int}
\cs_new_protected:Npn \int_gdecr:N #1{\tex_global:D \tex_advance:D #1 - \c_one_int}
\cs_generate_variant:Nn \int_add:N { c }
\cs_generate_variant:Nn \int_decr:N { c }
\cs_generate_variant:Nn \int_gincr:N { c }
\cs_generate_variant:Nn \int_gdecr:N { c }

(End definition for \int_add:N and others. These functions are documented on page 157.)

\int_incr:N \int_incr:c \int_gincr:N \int_gincr:c \int_decr:N \int_decr:c \int_gdecr:N \int_gdecr:c

Incrementing and decrementing of integer registers is done with the following functions.

\cs_new_protected:Npn \int_incr:N #1{\tex_advance:D #1 \c_one_int}
\cs_new_protected:Npn \int_decr:N #1{\tex_advance:D #1 - \c_one_int}
\cs_new_protected:Npn \int_gincr:N #1{\tex_global:D \tex_advance:D #1 \c_one_int}
\cs_new_protected:Npn \int_gdecr:N #1{\tex_global:D \tex_advance:D #1 - \c_one_int}
\cs_generate_variant:Nn \int_incr:N { #1 }
\cs_generate_variant:Nn \int_gincr:N { #1 }
\cs_generate_variant:Nn \int_decr:N { #1 }
\cs_generate_variant:Nn \int_gdecr:N { #1 }
\cs_generate_variant:Nn \int_incr:c { #1 }
\cs_generate_variant:Nn \int_gincr:c { #1 }
\cs_generate_variant:Nn \int_decr:c { #1 }
\cs_generate_variant:Nn \int_gdecr:c { #1 }
57.4 Using integers

As integers are register-based \TeX{} issues an error if they are not defined. While the =

sign is optional, this version with = is slightly quicker than without, while adding the

optional space after = slows things down minutely.

57.5 Integer expression conditionals

Those functions are used for comparison tests which use a simple syntax where only

one set of braces is required and additional operators such as != and >= are supported.

The tests first evaluate their left-hand side, with a trailing \_\_int_compare_error:. This marker is normally not expanded, but if the relation symbol is missing from the
test’s argument, then the marker inserts = (and itself) after triggering the relevant \TeX{}
error. If the first token which appears after evaluating and removing the left-hand side is

not a known relation symbol, then a judiciously placed \_\_int_compare_error:Nw gets expanded, cleaning up the end of the test and telling the user what the problem was.
Comparisons using a simple syntax where only one set of braces is required, additional operators such as != and >= are supported, and multiple comparisons can be performed at once, for instance 0 < 5 <= 1. The idea is to loop through the argument, finding one operand at a time, and comparing it to the previous one. The looping auxiliary \__int_compare:Nw reads one ⟨operand⟩ and one ⟨comparison⟩ symbol, and leaves roughly

\begin{verbatim}
\prg_return_false: \fi:
\reverse_if:N \if_int_compare:w ⟨operand⟩⟨comparison⟩ \__int_compare:Nw
\end{verbatim}

in the input stream. Each call to this auxiliary provides the second operand of the last call’s \if_int_compare:w. If one of the ⟨comparisons⟩ is false, the true branch of the \TeX{} conditional is taken (because of \reverse_if:N), immediately returning false as the result of the test. There is no \TeX{} conditional waiting the first operand, so we add an \if_false: and expand by hand with \int_value:w, thus skipping \prg_return_-false: on the first iteration.

Before starting the loop, the first step is to make sure that there is at least one relation symbol. We first let \TeX{} evaluate this left hand side of the (in)equality using \__int_eval:w. Since the relation symbols <, >, = and ! are not allowed in integer expressions, they would terminate the expression. If the argument contains no relation symbol, \__int_compare_error: is expanded, inserting = and itself after an error. In all cases, \__int_compare:w receives as its argument an integer, a relation symbol, and some more tokens. We then setup the loop, which is ended by the two odd-looking items e and {=nd_}, with a trailing \s__int_stop used to grab the entire argument when necessary.
When the last (operand) is seen, \_\_\_int_compare:NNw receives e and =nd as arguments, hence calling \_\_\_int_compare_end=:NNw to end the loop: return the result of the last comparison (involving the operand that we just found). When a normal relation is found, the appropriate auxiliary calls \_\_\_int_compare:nnN where #1 is \if_int_compare:w or \reverse_if:N \if_int_compare:w, #2 is the (operand), and #3 is one of <, =, or >. As announced earlier, we leave the (operand) for the previous conditional. If this conditional is true the result of the test is known, so we remove all tokens and return false. Otherwise, we apply the conditional #1 to the (operand) #2 and the comparison #3, and call \_\_\_int_compare:NNw to look for additional operands, after evaluating the following expression.

The actual comparisons are then simple function calls, using the relation as delimiter for a delimited argument and discarding \_\_\_int_compare_error:Nw (token) responsible for error detection.
More efficient but less natural in typing.

\begin{verbatim}
\prg_new_conditional:Nn \int_compare:nNn #1#2#3 { p , T , F , TF }
\{
  \if_int_compare:w \__int_eval:w #1 \__int_eval:w #2 \__int_eval:w #3 \__int_eval_end:
  \prg_return_true:
  \else:
  \prg_return_false:
  \fi:
\}
\end{verbatim}

For integer cases, the first task to fully expand the check condition. The overall idea is then much the same as for \texttt{\tl_case:nnTF} as described in \texttt{l3tl}.

\begin{verbatim}
\cs_new:Npn \int_case:nnTF #1 #2 #3 #4
\{
  \__int_case:nw {#1} #2 {#1} { } \s__int_mark {#3} \s__int_mark {#4} \s__int_stop
\}
\end{verbatim}

A predicate function.

\begin{verbatim}
\prg_new_conditional:Nn \int_if_odd:n #1 { p , T , F , TF }
\{
  \if_int_odd:w \__int_eval:w #1 \__int_eval_end:
  \prg_return_true:
  \else:
  \prg_return_false:
  \fi:
\}
\end{verbatim}

\begin{verbatim}
\prg_new_conditional:Nn \int_if_even:n #1 { p , T , F , TF }
\{
  \if_int_even:w \__int_eval:w #1 \__int_eval_end:
  \prg_return_true:
  \else:
  \prg_return_false:
  \fi:
\}
\end{verbatim}
\else:
  \prg_return_false:
\fi:
\prg_new_conditional:Npnn \int_if_even:n #1 { p , T , F , TF} 
{ 
  \reverse_if:N \if_int_odd:w \__int_eval:w #1 \__int_eval_end:
  \prg_return_true:
  \else:
  \prg_return_false:
  \fi:
}
\reverse_if:N \if_int_odd:w \__int_eval:w #1 \__int_eval_end:
\prg_return_true:
\else:
\prg_return_false:
\fi:

(End definition for \int_if_odd:nTF and \int_if_even:nTF. These functions are documented on page 159.)

57.6 Integer expression loops

These are quite easy given the above functions. The while versions test first and then execute the body. The do_while does it the other way round.

\cs_new:Npn \int_while_do:nn #1#2 
{ 
\int_compare:nT {#1} 
{ 
  \int_compare:nT {#1} {#2} 
}
\cs_new:Npn \int_until_do:nn #1#2 
{ 
\int_compare:nF {#1} 
{ 
  \int_compare:nF {#1} {#2} 
}
\cs_new:Npn \int_do_while:nn #1#2 
{ #2 
  \int_compare:nT {#1} 
} 
\cs_new:Npn \int_do_until:nn #1#2 
{ 
#2 
  \int_compare:nF {#1} 
}

(End definition for \int_while_do:nn and others. These functions are documented on page 160.)

\cs_new:Npn \int_while_do:nNnn #1#2#3#4 
{ #2 
  \int_compare:nT {#1} 
  \int_compare:nT {#1} {#2} 
} 
\cs_new:Npn \int_until_do:nNnn #1#2#3#4 
{ #2 
  \int_compare:nF {#1} 
  \int_compare:nF {#1} {#2} 
} 
\cs_new:Npn \int_do_while:nNnn #1#2#3#4 
{ 
#2 
  \int_compare:nT {#1} 
} 
\cs_new:Npn \int_do_until:nNnn #1#2#3#4 
{ 
#2 
  \int_compare:nF {#1} 
} 

(End definition for \int_while_do:nNnn and others. These functions are documented on page 160.)

\int_while_do:nNn 
\int_until_do:nNn 
\int_do_while:nNn 
\int_do_until:nNn 

As above but not using the more natural syntax.

\cs_new:Npn \int_while_do:nNnn #1#2#3#4 
{ #2 
  \int_compare:nT {#1} 
  \int_compare:nT {#1} {#2} 
} 
\cs_new:Npn \int_until_do:nNnn #1#2#3#4 
{ #2 
  \int_compare:nF {#1} 
  \int_compare:nF {#1} {#2} 
} 
\cs_new:Npn \int_do_while:nNnn #1#2#3#4 
{ 
#2 
  \int_compare:nT {#1} 
} 
\cs_new:Npn \int_do_until:nNnn #1#2#3#4 
{ 
#2 
  \int_compare:nF {#1} 
} 

810
Before all else, evaluate the initial value, step, and final value. Repeating a function by steps first needs a check on the direction of the steps. After that, do the function for the start value then step and loop around. It would be more symmetrical to test for a step size of zero before checking the sign, but we optimize for the most frequent case (positive step).

\int_step_function:nnnN
\__int_step:wwwN
\__int_step:NwwwN
\int_step_function:nN
\int_step_function:nnN

\textbf{57.7 Integer step functions}

(End definition for \texttt{\string\int\_while\_do:nnNN} and others. These functions are documented on page 160.)
The approach here is to build a function, with a global integer required to make the nesting safe (as seen in other in line functions), and map that function using \int_step_function:nnnN. We put a \prg_break_point:Nn so that map_break functions from other modules correctly decrement \g__kernel_prg_map_int before looking for their own break point. The first argument is \scan_stop:, so that no breaking function recognizes this break point as its own.
57.8 Formatting integers

\int_to_arabic:n  Nothing exciting here.

\int_to_symbols:nnn  \__int_to_symbols:nnnn  For conversion of integers to arbitrary symbols the method is in general as follows. The input number (#1) is compared to the total number of symbols available at each place (#2). If the input is larger than the total number of symbols available then the modulus is needed, with one added so that the positions don’t have to number from zero. Using an f-type expansion, this is done so that the system is recursive. The actual conversion function therefore gets a ‘nice’ number at each stage. Of course, if the initial input was small enough then there is no problem and everything is easy.

\int_to_alph:n  \int_to_Alph:n  These both use the above function with input functions that make sense for the alphabet in English.
\cs_new:Npn \int_to_alph:n #1 
{
\int_to_symbols:nnn {#1} { 26 }

\{
  { 1 } { a }
  { 2 } { b }
  { 3 } { c }
  { 4 } { d }
  { 5 } { e }
  { 6 } { f }
  { 7 } { g }
  { 8 } { h }
  { 9 } { i }
  {10} { j }
  {11} { k }
  {12} { l }
  {13} { m }
  {14} { n }
  {15} { o }
  {16} { p }
  {17} { q }
  {18} { r }
  {19} { s }
  {20} { t }
  {21} { u }
  {22} { v }
  {23} { w }
  {24} { x }
  {25} { y }
  {26} { z }
\}
}
\cs_new:Npn \int_to_Alph:n #1 
{
\int_to_symbols:nnn {#1} { 26 }

\{
  { 1 } { A }
  { 2 } { B }
  { 3 } { C }
  { 4 } { D }
  { 5 } { E }
  { 6 } { F }
  { 7 } { G }
  { 8 } { H }
  { 9 } { I }
  {10} { J }
  {11} { K }
  {12} { L }
  {13} { M }
  {14} { N }
  {15} { O }
  {16} { P }
  {17} { Q }
  {18} { R }
\}
}
Converting from base ten (#1) to a second base (#2) starts with computing #1: if it is a complicated calculation, we shouldn’t perform it twice. Then check the sign, store it, either - or \texttt{\textbackslash c_empty_tl}, and feed the absolute value to the next auxiliary function.

Here, the idea is to provide a recursive system to deal with the input. The output is built up after the end of the function. At each pass, the value in #1 is checked to see if it is less than the new base (#2). If it is, then it is converted directly, putting the sign back in front. On the other hand, if the value to convert is greater than or equal to the new base then the modulus and remainder values are found. The modulus is converted to a symbol and put on the right, and the remainder is carried forward to the next round.
Convert to a letter only if necessary, otherwise simply return the value unchanged. It would be cleaner to use \int_case:nn, but in our case, the cases are contiguous, so it is forty times faster to use the \if_case:w primitive. The first \exp_after:wN expands the conditional, jumping to the correct case, the second one expands after the resulting character to close the conditional. Since #1 might be an expression, and not directly a single digit, we need to evaluate it properly, and expand the trailing \fi:.

\cs_new:Npn \__int_to_letter:n #1
  \exp_after:wN \exp_after:wN
  \if_case:w \__int_eval:w #1 - 10 \__int_eval_end:
  \or: a
  \or: b
  \or: c
  \or: d
  \or: e
  \or: f
  \or: g
  \or: h
  \or: i
  \or: j
  \or: k
  \or: l
  \or: m
  \or: n
  \or: o
  \or: p
  \or: q
  \or: r

816
\int_to_bin:n \int_to_hex:n \int_to_Hex:n \int_to_oct:n

Wrappers around the generic function.

(End definition for \int_to_base:nn and others. These functions are documented on page 163.)
The \texttt{\_int_to_roman:w} primitive creates tokens of category code 12 (other). Usually, what is actually wanted is letters. The approach here is to convert the output of the primitive into letters using appropriate control sequence names. That keeps everything expandable. The loop is terminated by the conversion of the Q. 

\begin{verbatim}
\cs_new:Npn \int_to_roman:n #1
{ \exp_after:wN \__int_to_roman:N \__int_to_roman:w \int_eval:n {#1} Q }
\cs_new:Npn \__int_to_roman:N #1
{ \use:c { \__int_to_roman_#1:w } \__int_to_roman:w }
\cs_new:Npn \int_to_Roman:n #1
{ \exp_after:wN \__int_to_Roman_aux:N \__int_to_roman:w \int_eval:n {#1} Q }
\cs_new:Npn \__int_to_Roman_aux:N #1
{ \use:c { \__int_to_Roman_#1:w } \__int_to_Roman_aux:N }
\cs_new:Npn \__int_to_roman_i:w { i }
\cs_new:Npn \__int_to_roman_v:w { v }
\cs_new:Npn \__int_to_roman_x:w { x }
\cs_new:Npn \__int_to_roman_l:w { l }
\cs_new:Npn \__int_to_roman_c:w { c }
\cs_new:Npn \__int_to_roman_d:w { d }
\cs_new:Npn \__int_to_roman_m:w { m }
\cs_new:Npn \__int_to_roman_Q:w #1 { }
\cs_new:Npn \__int_to_Roman_i:w { I }
\cs_new:Npn \__int_to_Roman_v:w { V }
\cs_new:Npn \__int_to_Roman_x:w { X }
\cs_new:Npn \__int_to_Roman_l:w { L }
\cs_new:Npn \__int_to_Roman_c:w { C }
\cs_new:Npn \__int_to_Roman_d:w { D }
\cs_new:Npn \__int_to_Roman_m:w { M }
\cs_new:Npn \__int_to_Roman_Q:w #1 { }
\end{verbatim}

(End definition for \texttt{\int_to_roman:n} and others. These functions are documented on page 163.)

57.9 Converting from other formats to integers

Called as \texttt{\_int_pass_signs:wn \langle signs and digits \rangle \s__int_stop \{\langle code\rangle\}}, this function leaves in the input stream any sign it finds, then inserts the \langle code\rangle before the first
non-sign token (and removes \texttt{s\_int\_stop}). More precisely, it deletes any + and passes any – to the input stream, hence should be called in an integer expression.

\begin{verbatim}
\cs_new:Npn \_int_pass_signs:wn #1
{\if:w + \if:w - \exp_not:N #1 + \fi: \exp_not:N #1 \exp_after:wN \_int_pass_signs:wn \else: \exp_after:wN \_int_pass_signs_end:wn \exp_after:wN #1 \fi:}
\cs_new:Npn \_int_pass_signs_end:wn #1 \s__int_stop #2 { #2 #1 }
\end{verbatim}

(End definition for \texttt{\_int_pass_signs:wn} and \texttt{\_int_pass_signs_end:wn}.)

\begin{verbatim}
\int_from_alph:n \_int_from_alph:nN \_int_from_alph:N
\end{verbatim}

First take care of signs then loop through the input using the \texttt{recursion} quarks. The \texttt{\_int_from_alph:nN} auxiliary collects in its first argument the value obtained so far, and the auxiliary \texttt{\_int_from_alph:N} converts one letter to an expression which evaluates to the correct number.

\begin{verbatim}
\cs_new:Npn \int_from_alph:n #1
{\int_eval:n
{\exp_after:wN \_int_pass_signs:wn \tl_to_str:n {#1} \s__int_stop { \_int_from_alph:nN { 0 } } \q__int_recursion_tail \q__int_recursion_stop}
}
\cs_new:Npn \_int_from_alph:nN #1#2
{\__int_if_recursion_tail_stop_do:Nn #2 {#1} \exp_args:Nf \_int_from_alph:nN { \int_eval:n { #1 * 26 + \_int_from_alph:N #2 } }
}
\cs_new:Npn \_int_from_alph:N #1
{ '#1 - \int_compare:nNnTF { '#1 } < { 91 } { 64 } { 96 } }
\end{verbatim}

(End definition for \texttt{\int_from_alph:n}, \texttt{\_int_from_alph:nN}, and \texttt{\_int_from_alph:N}. This function is documented on page 163.)

\begin{verbatim}
\int_from_base:nn \_int_from_base:nnN \_int_from_base:N
\end{verbatim}

Leave the signs into the integer expression, then loop through characters, collecting the value found so far in the first argument of \texttt{\_int_from_base:nnN}. To convert a single character, \texttt{\_int_from_base:N} checks first for digits, then distinguishes lower from upper case letters, turning them into the appropriate number. Note that this auxiliary does not use \texttt{\int_eval:n}, hence is not safe for general use.

\begin{verbatim}
\cs_new:Npn \int_from_base:nn #1#2
{\int_eval:n
{\exp_after:wN \_int_pass_signs:wn \tl_to_str:n {#1} \s__int_stop { \_int_from_base:nnN { 0 } } \q__int_recursion_tail \q__int_recursion_stop}
}
\cs_new:Npn \_int_from_base:nnN #1#2
{\__int_if_recursion_tail_stop_do:Nn #2 {#1} \exp_args:Nf \_int_from_base:nnN { \int_eval:n { #1 * 26 + \_int_from_base:N #2 } }
}
\cs_new:Npn \_int_from_base:N #1
{ '#1 - \int_compare:nNnTF { '#1 } < { 91 } { 64 } { 96 } }
\end{verbatim}

819
\cs_new:Npn \__int_from_base:nnN #1#2#3
\{
\__int_if_recursion_tail_stop_do:Nn #3 {#1}
\exp_args:Nf \__int_from_base:nnN
\{ \int_eval:n { #1 * #2 + \__int_from_base:N #3 } \}
\}
\cs_new:Npn \__int_from_base:N #1
\{
\int_compare:nNnTF { '#1 } < { 58 }
{#1}
{ '#1 - \int_compare:nNnTF { '#1 } < { 91 } { 55 } { 87 } }
\}
\int_from_bin:n
\int_from_hex:n
\int_from_oct:n
Wrappers around the generic function.

\cs_new:Npn \int_from_bin:n #1
\{ \int_from_base:nn {#1} { 2 } \}
\cs_new:Npn \int_from_hex:n #1
\{ \int_from_base:nn {#1} { 16 } \}
\cs_new:Npn \int_from_oct:n #1
\{ \int_from_base:nn {#1} { 8 } \}

Constants used to convert from Roman numerals to integers.

\int_const:cn { c__int_from_roman_i_int } { 1 }
\int_const:cn { c__int_from_roman_v_int } { 5 }
\int_const:cn { c__int_from_roman_x_int } { 10 }
\int_const:cn { c__int_from_roman_l_int } { 50 }
\int_const:cn { c__int_from_roman_c_int } { 100 }
\int_const:cn { c__int_from_roman_d_int } { 500 }
\int_const:cn { c__int_from_roman_m_int } { 1000 }
\int_const:cn { c__int_from_roman_I_int } { 1 }
\int_const:cn { c__int_from_roman_V_int } { 5 }
\int_const:cn { c__int_from_roman_X_int } { 10 }
\int_const:cn { c__int_from_roman_L_int } { 50 }
\int_const:cn { c__int_from_roman_C_int } { 100 }
\int_const:cn { c__int_from_roman_D_int } { 500 }
\int_const:cn { c__int_from_roman_M_int } { 1000 }

The method here is to iterate through the input, finding the appropriate value for each letter and building up a sum. This is then evaluated by \TeX. If any unknown letter is found, skip to the closing parenthesis and insert \texttt{*0-1} afterwards, to replace the value by \texttt{-1}.  

820
57.10 Viewing integer

\int_show:N  Diagnostics.
\int_show:c
\__int_show:nN

(End definition for \int_show:N and \__int_show:nN. This function is documented on page 165.)

\int_show:n  We don’t use the \TeX{} primitive \texttt{\texttt{\texttt{\texttt{showthe}}} to show integer expressions: this gives a more unified output.
\__int_show:nN  \msg_show_eval:Nn \int_eval:n }

(End definition for \int_show:n. This function is documented on page 165.)

\int_log:N  Diagnostics.
\int_log:c

\__int_log:nN  \msg_show_eval:Nn \int_eval:n }

(End definition for \int_log:n. This function is documented on page 165.)
\int_log:n

Similar to \int_show:n.

\int_rand:nn

Defined in \texttt{l3fp-random}.

\c_zero_int \c_one_int

The zero is defined in \texttt{l3basics}.

\int_const:Nn \c_one_int { 1 }

(End definition for \c_zero_int and \c_one_int. These variables are documented on page 165.)

\c_max_int

The largest number allowed is $2^{31} - 1$

\int_const:Nn \c_max_int { 2 147 483 647 }

(End definition for \c_max_int. This variable is documented on page 165.)

\c_max_char_int

The largest character code is 1114111 (hexadecimal \texttt{10FFFF}) in Xe\TeX{} and Lua\TeX{} and 255 in other engines. In many places \texttt{pdflatex} and \texttt{upTeX} support larger character codes but for instance the values of \texttt{	extbackslash lccode} are restricted to $[0, 255]$.

\int_const:Nn \c_max_char_int
{ 
  \if_int_odd:w 0  
  \cs_if_exist:NT \tex_luatexversion:D { 1 }  
  \cs_if_exist:NT \tex_xelatexversion:D { 1 } -  
  \tex_if_exist:NT \tex_tikzversion:D \texttt{10FFFF}  
  \else:  
    \texttt{FF}  
  \fi:  
}

(End definition for \c_max_char_int. This variable is documented on page 165.)

\l_tmpa_int \l_tmpb_int \g_tmpa_int \g_tmpb_int

We provide two local and two global scratch counters, maybe we need more or less.

\int_new:N \l_tmpa_int  
\int_new:N \l_tmpb_int  
\int_new:N \g_tmpa_int  
\int_new:N \g_tmpb_int

(End definition for \l_tmpa_int and others. These variables are documented on page 165.)
57.14 Integers for earlier modules

\l_int_internal_a_int
\l_int_internal_b_int

\int_new:N \l_int_internal_a_int
\int_new:N \l_int_internal_b_int

(End definition for \l_int_internal_a_int and \l_int_internal_b_int.)

\endocument
Chapter 58

l3flag implementation

The following test files are used for this code: m3flag001.

58.1 Non-expandable flag commands

The height $h$ of a flag (initially zero) is stored by setting control sequences of the form \flag\langle name\rangle \langle integer\rangle to \relax for $0 \leq \langle integer\rangle < h$. When a flag is raised, a “trap” function \flag\langle name\rangle is called. The existence of this function is also used to test for the existence of a flag.

\flag_new:n

For each flag, we define a “trap” function, which by default simply increases the flag by 1 by letting the appropriate control sequence to \relax. This can be done expandably!

\flag_clear:n

Undefine control sequences, starting from the 0 flag, upwards, until reaching an undefined control sequence. We don’t use \cs_undefine:c because that would act globally. When the option check-declarations is used, check for the function defined by \flag_new:n.

(End definition for \flag_new:n. This function is documented on page 168.)
As for other datatypes, clear the \(\text{⟨flag⟩}\) or create a new one, as appropriate.

\begin{verbatim}
\texttt{\textbackslash cs\_new\_protected:Npn \textbackslash flag\_clear\_new:n \#1}
\end{verbatim}

\texttt{(End definition for \textbackslash flag\_clear\_new:n. This function is documented on page 169.)}

Show the height (terminal or log file) using appropriate \l3msg\ auxiliaries.

\begin{verbatim}
\texttt{\textbackslash cs\_new\_protected:Npn \textbackslash flag\_show:n \__flag\_show:Nn}
\end{verbatim}

\begin{verbatim}
\texttt{\textbackslash cs\_new\_protected:Npn \textbackslash flag\_log:n \__flag\_show:Nn \textbackslash tl\_log:n}
\end{verbatim}

\begin{verbatim}
\texttt{\textbackslash cs\_new\_protected:Npn \__flag\_show:Nn \#1}
\end{verbatim}

\texttt{(End definition for \textbackslash flag\_show:n, \textbackslash flag\_log:n, and \__flag\_show:Nn. These functions are documented on page 169.)}

### 58.2 Expandable flag commands

A flag exist if the corresponding trap \(\text{⟨flag name⟩}:n\) is defined.

\begin{verbatim}
\texttt{\textbackslash prg\_new\_conditional:Npnn \textbackslash flag\_if\_exist:n \#1 \ p, T, F, TF}
\end{verbatim}

\begin{verbatim}
\texttt{\cs\_if\_exist:cTF \{ \text{flag-#1} \}
\texttt{\{ \prg\_return\_true: \} \{ \prg\_return\_false: \}}
\end{verbatim}

\texttt{(End definition for \textbackslash flag\_if\_exist:nTF. This function is documented on page 169.)}

Test if the flag has a non-zero height, by checking the 0 control sequence.

\begin{verbatim}
\texttt{\textbackslash prg\_new\_conditional:Npnn \textbackslash flag\_if\_raised:n \#1 \ p, T, F, TF}
\end{verbatim}

\begin{verbatim}
\texttt{\if\_cs\_exist:w flag\-#1\-0 \cs\_end:
\prg\_return\_true:
\else:
\prg\_return\_false:
\fi:
\end{verbatim}

\texttt{(End definition for \textbackslash flag\_if\_raised:nTF. This function is documented on page 169.)}

Extract the value of the flag by going through all of the control sequences starting from 0.

\begin{verbatim}
\texttt{\textbackslash cs\_new:Npn \textbackslash flag\_height:n \#1 \__flag\_height\_loop:wn \textbackslash int\_value:w \textbackslash int\_eval:w 1 +}
\end{verbatim}

\begin{verbatim}
\texttt{\textbackslash \_flag\_height\_loop:wn \#1}
\end{verbatim}

\texttt{(End definition for \textbackslash flag\_height:n and \__flag\_height:wn. This function is documented on page 168.)}
\begin{verbatim}
\exp_after:wN \__flag_height_end:wn
\fi:
\__flag_height_loop:wn #1 ; #2
\__flag_height_end:wn #1 ; #2 {#1}
\cs_new:Npn \__flag_height_end:wn #1 ; #2 {#1}

(End definition for \flag_height:n, \__flag_height_loop:wn, and \__flag_height_end:wn. This function is documented on page 169.)

\flag_raise:n Simply apply the trap to the height, after expanding the latter.
\begin{verbatim}
\cs_new:Npn \flag_raise:n #1
\cs:w flag~#1 \exp_after:wN \cs_end:
\int_value:w \flag_height:n {#1} ;
\end{verbatim}

(End definition for \flag_raise:n. This function is documented on page 169.)
\end{verbatim}
Chapter 59

\texttt{l3clist} implementation

The following test files are used for this code: \texttt{m3clist002}.

\texttt{\langle package \rangle}
\texttt{\langle @@=clist \rangle}

\texttt{\c_empty_clist}  An empty comma list is simply an empty token list.
\texttt{\langle cs_new_eq:NN \c_empty_clist \c_empty_tl \rangle}

(End definition for \texttt{\c_empty_clist}. This variable is documented on page 179.)

\texttt{\l__clist_internal_clist}  Scratch space for various internal uses. This comma list variable cannot be declared as such because it comes before \texttt{clist_new:N}
\texttt{\langle tl_new:N \l__clist_internal_clist \rangle}

(End definition for \texttt{\l__clist_internal_clist}.)

\texttt{\s__clist_mark}  \texttt{\s__clist_stop}  Internal scan marks.
\texttt{\langle \scan_new:N \s__clist_mark \rangle}
\texttt{\langle \scan_new:N \s__clist_stop \rangle}

(End definition for \texttt{\s__clist_mark} and \texttt{\s__clist_stop}.)

\texttt{\__clist_use_none_delimit_by_s_mark:w}  \texttt{\__clist_use_none_delimit_by_s_stop:w}  \texttt{\__clist_use_i_delimit_by_s_stop:nw}  Functions to gobble up to a scan mark.
\texttt{\langle cs_new:Np \__clist_use_none_delimit_by_s_mark:w \#1 \s__clist_mark \{ \} \rangle}
\texttt{\langle cs_new:Np \__clist_use_none_delimit_by_s_stop:w \#1 \s__clist_stop \{ \} \rangle}
\texttt{\langle cs_new:Np \__clist_use_i_delimit_by_s_stop:nw \#1 \#2 \s__clist_stop \{#1\} \rangle}

(End definition for \texttt{\__clist_use_none_delimit_by_s_mark:w}, \texttt{\__clist_use_none_delimit_by_s_stop:w}, and \texttt{\__clist_use_i_delimit_by_s_stop:nw}.)

\texttt{\__clist_tmp:w}  A temporary function for various purposes.
\texttt{\langle cs_new_protected:Np \__clist_tmp:w \{ \} \rangle}

(End definition for \texttt{\__clist_tmp:w}.)
59.1 Removing spaces around items

\clist_trim_next:w  Called as \exp:w \clist_trim_next:w \prg_do_nothing: ⟨comma list⟩ ... it expands to {⟨trimmed item⟩} where the ⟨trimmed item⟩ is the first non-empty result from removing spaces from both ends of comma-delimited items in the ⟨comma list⟩. The \prg_do_nothing: marker avoids losing braces. The test for blank items is a somewhat optimized \tl_if_empty:oTF construction; if blank, another item is sought, otherwise trim spaces.

\clist_sanitize:n  \clist_sanitize:Nn

The auxiliary \clist_sanitize:Nn receives a delimiter (\c_empty_tl the first time, afterwards a comma) and that item as arguments. Unless we are done with the loop it calls \clist_wrap_item:w to unbrace the item (using a comma delimiter is safe since \#2 came from removing spaces from an argument delimited by a comma) and possibly re-brace it if needed.

\clist_if_wrap:nTF  \clist_if_wrap:w

True if the argument must be wrapped to avoid getting altered by some clist operations. That is the case whenever the argument

- starts or end with a space or contains a comma,
- is empty, or
- consists of a single braced group.

If the argument starts or ends with a space or contains a comma then one of the three arguments of \clist_if_wrap:w will have its end delimiter (partly) in one of the three copies of \#1 in \clist_if_wrap:nTF; this has a knock-on effect meaning that the result of the expansion is not empty; in that case, wrap. Otherwise, the argument is safe unless it starts with a brace group (or is empty) and it is empty or consists of a single n-type argument.
\texttt{\_clist_if_wrap:w} \\_clist_mark \#1 ~ \_clist_mark \#2 ~ \_clist_mark \#3 , \\ \texttt{\_clist_wrap_item:w} Safe items are put in \texttt{\exp_not:n}, otherwise we put an extra set of braces.

59.2 Allocation and initialisation

\texttt{\clist_new:N} \texttt{\clist_new:c} Internally, comma lists are just token lists.

\texttt{\clist_new:NN} \texttt{\clist_new:cc} Creating and initializing a constant comma list is done by sanitizing all items (stripping spaces and braces).

\texttt{\clist_clear:N} \texttt{\clist_clear:c} Clearing comma lists is just the same as clearing token lists.
Once again a copy from the token list functions.

\clist_set_eq:NN \clist_set_eq:cN \clist_set_eq:Nc \clist_set_eq:cc \clist_gset_eq:NN \clist_gset_eq:cN \clist_gset_eq:Nc \clist_gset_eq:cc

Setting a comma list from a comma-separated list is done using a simple mapping. Safe items are put in \exp_not:n, otherwise we put an extra set of braces. The first comma must be removed, except in the case of an empty comma-list.
Concatenating comma lists is not quite as easy as it seems, as there needs to be the correct addition of a comma to the output. So a little work to do.

\begin{verbatim}
\cs_new_protected:Npn \clist_concat:NNN { \__clist_concat:NNNN \__kernel_tl_set:Nx }
\cs_new_protected:Npn \clist_gconcat:NNN { \__clist_concat:NNNN \__kernel_tl_gset:Nx }
\cs_new_protected:Npn \__clist_concat:NNNN #1#2#3#4
{ #1 #2
{ \exp_not:o #3
\clist_if_empty:NF #3 { \clist_if_empty:NF #4 { , } }
\exp_not:o #4
}
}
\cs_generate_variant:Nn \clist_concat:NNN { ccc }
\cs_generate_variant:Nn \clist_gconcat:NNN { ccc }
\end{verbatim}

(End definition for \clist_concat:NNN, \clist_gconcat:NNN, and \__clist_concat:NNNN. These functions are documented on page 172.)

\begin{verbatim}
\prg_new_eq_conditional:NNn \clist_if_exist:N \cs_if_exist:N { TF , T , F , p }
\prg_new_eq_conditional:NNn \clist_if_exist:c \cs_if_exist:c { TF , T , F , p }
\end{verbatim}

(End definition for \clist_if_exist:NTF. This function is documented on page 172.)

59.3 Adding data to comma lists

\begin{verbatim}
\cs_new_protected:Npm \clist_concat:NNN \clist_set:Nn #1#2
{ \__clist_concat:NNNN \__kernel_tl_set:Nx { \__clist_sanitize:n {#2} } }
\cs_new_protected:Npm \clist_gconcat:NNN \clist_set:Nn #1#2
{ \__clist_concat:NNNN \__kernel_tl_gset:Nx { \__clist_sanitize:n {#2} } }
\end{verbatim}

(End definition for \clist_set:Nn and \clist_gset:Nn. These functions are documented on page 172.)

Everything is based on concatenation after storing in \l__clist_internal_clist. This avoids having to worry here about space-trimming and so on.

\begin{verbatim}
\cs_new_protected:Npm \clist_put_left:Nn \clist_set:Nn #1#2
{ \__clist_put_left:NNNn \__clist_concat:NNN \__kernel_tl_set:Nx { \__clist_sanitize:n {#2} } }
\cs_new_protected:Npm \clist_gput_left:Nn \clist_set:Nn #1#2
{ \__clist_put_left:NNNn \__clist_gconcat:NNN \__kernel_tl_set:Nx { \__clist_sanitize:n {#2} } }
\end{verbatim}

(End definition for \clist_set:Nn and \clist_gset:Nn. These functions are documented on page 172.)
Comma lists as stacks

Getting an item from the left of a comma list is pretty easy: just trim off the first item using the comma. No need to trim spaces as comma-list variables are assumed to have "cleaned-up" items. (Note that grabbing a comma-delimited item removes an outer pair of braces if present, exactly as needed to uncover the underlying item.)

An empty clist leads to \texttt{\q_no_value} otherwise grab until the first comma and assign to the variable. The second argument of \texttt{\_clist_pop:wwNN} is a comma list ending in a comma and \texttt{\s__clist_mark}, unless the original clist contained exactly one item: then the argument is just \texttt{\s__clist_mark}. The next auxiliary picks either \texttt{\exp_not:n} or \texttt{\use_none:n} as #2, ensuring that the result can safely be an empty comma list.
\clist_get:NN { T , F , TF } \clist_get:cN { T , F , TF } \clist_pop:NN { T , F , TF } \clist_pop:cN { T , F , TF } \clist_gpop:NN { T , F , TF } \clist_gpop:cN { T , F , TF }

The same, as branching code: very similar to the above.

\prg_new_protected_conditional:Npnn \clist_get:NN #1 #2 { T , F , TF } \prg_new_protected_conditional:Npnn \clist_get:cN #1 #2 { T , F , TF } \prg_new_protected_conditional:Npnn \clist_pop:NN #1 #2 { T , F , TF } \prg_new_protected_conditional:Npnn \clist_pop:cN #1 #2 { T , F , TF } \prg_new_protected_conditional:Npnn \clist_gpop:NN #1 #2 { T , F , TF } \prg_new_protected_conditional:Npnn \clist_gpop:cN #1 #2 { T , F , TF }

(End definition for \clist_pop:NN and others. These functions are documented on page 177.)
Modifying comma lists

Removing duplicates means making a new list then copying it.

```latex
\clist_remove_duplicates:N
\clist_gremove_duplicates:N
```

(End definition for \clist_remove_duplicates:N and \clist_gremove_duplicates:N.)
The method used here for safe items is very similar to \texttt{\textbackslash tl\_replace\_all:Nnn}. However, if the item contains commas or leading/trailing spaces, or is empty, or consists of a single brace group, we know that it can only appear within braces so the code would fail; instead just convert to a sequence and do the removal with \texttt{l3seq} code (it involves somewhat elaborate code to do most of the work expandably but the final token list comparisons non-expandably).

For “safe” items, build a function delimited by the \emph{⟨item⟩} that should be removed, surrounded with commas, and call that function followed by the expanded comma list, and another copy of the \emph{⟨item⟩}. The loop is controlled by the argument grabbed by \texttt{\_\_clist\_remove\_all:w}: when the item was found, the \texttt{\_\_clist\_mark} delimiter used is the one inserted by \texttt{\_\_clist\_tmp:w}, and \texttt{\_\_clist\_use\_none\_delimit\_by\_s\_stop:w} is deleted. At the end, the final \emph{⟨item⟩} is grabbed, and the argument of \texttt{\_\_clist\_tmp:w} contains \texttt{\_\_clist\_mark}: in that case, \texttt{\_\_clist\_remove\_all:w} removes the second \texttt{\_\_clist\_mark} (inserted by \texttt{\_\_clist\_tmp:w}), and lets \texttt{\_\_clist\_use\_none\_delimit\_by\_s\_stop:w} act.

No brace is lost because items are always grabbed with a leading comma. The result of the first assignment has an extra leading comma, which we remove in a second assignment. Two exceptions: if the clist lost all of its elements, the result is empty, and we shouldn’t remove anything; if the clist started up empty, the first step happens to turn it into a single comma, and the second step removes it.

\begin{verbatim}
\cs_new_protected:Npn \clist_remove_all:Nn       \list_remove_all:cn       \clist_gremove_all:Nn       \_\_clist_remove_all:NNNn \_\_clist_remove_all:w       \_\_clist_remove_all:
{ \__clist_remove_all:NNNn \clist_set_from_seq:NN \_\_kernel_tl_set:Nx }{ \__clist_remove_all:NNNn \clist_gset_from_seq:NN \_\_kernel_tl_gset:Nx }\cs_new_protected:Npn \__clist_remove_all:NNNn #1#2#3#4
\__clist_if_wrap:nTF {#4}
{ \seq_set_from_clist:NN \_\_clist_internal_remove_seq #3 \seq_remove_all:Nn \_\_clist_internal_remove_seq [#4] \_\_clist_internal_remove_seq }
\cs_set:Npn \_\_clist_tmp:w #1 , #4 , \__clist_remove_all:
\exp_after:wN \_\_clist_remove_all: #3 , \_\_clist_stop
\clist_if_empty:NF \_\_clist_remove_all:
\exp_args:No \exp_not:o
\end{verbatim}
Use \clist_reverse:n in an x-expanding assignment. The extra work that \clist_reverse:n does to preserve braces and spaces would not be needed for the well-controlled case of N-type comma lists, but the slow-down is not too bad.

The reversed token list is built one item at a time, and stored between \s__clist_stop and \s__clist_mark, in the form of “? followed by zero or more instances of “(item)i”. We start from a comma list “(item)1,...,(item)n”. During the loop, the auxiliary \__clist_reverse:wwNww receives “?(item)i” as #1, “(itemi+1),...,(item)n” as #2, \__clist_reverse:wwNww as #3, what remains until \s__clist_stop as #4, and “(itemi−1),...,(item)1”, as #5. The auxiliary moves #1 just before #5, with a comma, and calls itself (#3). After the last item is moved, \__clist_reverse:wwNww receives “\s__clist_mark \__clist_reverse:wwNww !” as its argument #1, thus \__clist_reverse_end:ww as its argument #3. This second auxiliary cleans up until the marker !, removes the trailing comma (introduced when the first item was moved after \s__clist_stop), and leaves its argument #1 within \exp_not:n. There is also a need to remove a leading comma, hence \exp_not:o and \use_none:n. There is also a need to remove a leading comma, hence \exp_not:o and \use_none:n.

(End definition for \clist_reverse:N and \__clist_reverse:wwNww. These functions are documented on page 173.)
59.6 Comma list conditionals

Simple copies from the token list variable material.

\clist_if_empty_p:N
\clist_if_empty_p:c
\clist_if_empty:N { T , F , TF }
\clist_if_empty:c { T , F , TF }
(End definition for \clist_if_empty:N. This function is documented on page 174.)

\clist_if_in:Nn
\clist_if_in:NV
\clist_if_in:No
\clist_if_in:cn
\clist_if_in:cV
\clist_if_in:co
\clist_if_in:nn
\clist_if_in:nV
\clist_if_in:no
\__clist_if_in_return:nnN
For “safe” items, we simply surround the comma list, and the item, with commas, then use the same code as for \tl_if_in:N. For “unsafe” items we follow the same route as \seq_if_in:Nn, mapping through the list a comparison function. If found, return true and remove \prg_return_false:

\prg_new_protected_conditional:Nnn \clist_if_in:Nn #1 \#2 { T , F , TF }
\exp_args:No \__clist_if_in_return:nnN #1 \#2
\prg_new_protected_conditional:Nppn \l__clist_internal_clist {#1}
\exp_args:No \__clist_if_in_return:nnN \l__clist_internal_clist {#2}
\clist_set:Nn \l__clist_internal_clist {#1}
59.7 Mapping over comma lists

If the variable is empty, the mapping is skipped (otherwise, that comma-list would be seen as consisting of one empty item). Then loop over the comma-list, grabbing eight comma-delimited items at a time. The end is marked by \s__clist_stop, which may not appear in any of the items. Once the last group of eight items has been reached, we go through them more slowly using \_cllist_map_function_end:w. The auxiliary function \_cllist_map_function:Nw is also used in some other clist mappings.
The \texttt{\_clist_map_function:nN} function is a bit more awkward, since spaces must be trimmed from each item. Space trimming is again based on \texttt{\_clist_trim_next:w}. The auxiliary \texttt{\_clist_map_function:nN} receives as arguments the function, and the next non-empty item (after space trimming but before brace removal). One level of braces is removed by \texttt{\_clist_map_unbrace:wn}.

Inline mapping is done by creating a suitable function “on the fly”: this is done globally to avoid any issues with \TeX’s groups. We use a different function for each level of nesting. Since the mapping is non-expandable, we can perform the space-trimming needed by the \texttt{n} version simply by storing the comma-list in a variable. We don’t need a different comma-list for each nesting level: the comma-list is expanded before the mapping starts.
\clist_map_variable:NNn
\clist_map_variable:cn
\clist_map_variable:nNn
\__clist_map_variable:Nnn

The N-type version is a straightforward application of \clist_map_tokens:Nn, calling \__clist_map_variable:Nnn for each item to assign the variable and run the user's code. The n-type version is not implemented in terms of the n-type function \clist_map_tokens:Nn, because here we are allowed to clean up the n-type comma list non-expandably.

\cs_new_protected:Npn \clist_map_variable:NNn #1#2#3
{ \clist_map_tokens:Nn #1 { \__clist_map_variable:Nnn #2 {#3} } }
\cs_generate_variant:Nn \clist_map_variable:NNn { c }
\cs_new_protected:Npn \__clist_map_variable:Nnn #1#2#3
{ \tl_set:Nn #1 {#3} #2 }
\cs_new_protected:Npn \clist_map_variable:nNn #1
{ \clist_set:Nn \l__clist_internal_clist {#1} \clist_map_variable:NNn \l__clist_internal_clist }
(End definition for \clist_map_variable:NNn, \clist_map_variable:nNn, and \__clist_map_variable:Nnn. These functions are documented on page 175.)

\clist_map_tokens:Nn
\clist_map_tokens:cn
\__clist_map_tokens:nw
\__clist_map_tokens_end:w

Essentially a copy of \clist_map_function:NN with braces added.
\cs_new:Npn \clist_map_tokens:Nn #1#2
{ \clist_if_empty:NF #1
\exp_last_unbraced:Nno \__clist_map_tokens:nw {#2} #1 ,
\s__clist_stop , \s__clist_stop , \s__clist_stop , \s__clist_stop ,
\prg_break_point:Nn \clist_map_break: { \int_gdec:N \g__kernel_prg_map_int }
}
\cs_new_protected:Npm \clist_map_tokens:nw #1 #2, #3, #4, #5, #6, #7, #8, #9,
{ \__clist_use_none_delimit_by_s_stop:w #9 \__clist_map_tokens_end:w \s__clist_stop
 \use:n {#1} {#2} \use:n {#1} {#3} \use:n {#1} {#4} \use:n {#1} {#5}
 \use:n {#1} {#6} \use:n {#1} {#7} \use:n {#1} {#8} \use:n {#1} {#9}
 \__clist_map_tokens:nw {#1} }
\clist_map_tokens:nn
Similar to \clist_map_function:nN but with a different way of grabbing items because we cannot use \exp_after:wN to pass the \textit{code}.

\clist_map_break:
The break statements use the general \prg_map_break:Nn mechanism.

\clist_count:N
Counting the items in a comma list is done using the same approach as for other token count functions: turn each entry into a +1 then use integer evaluation to actually do the mathematics. In the case of an \textit{n}-type comma-list, we could of course use \clist_map-_function:nN, but that is very slow, because it carefully removes spaces. Instead, we loop manually, and skip blank items (but not \texttt{}, hence the extra spaces).

(End definition for \clist_map_tokens:nn and \_\clist_map_tokens_n:nw. This function is documented on page 175.)
\cs_new:Npn \__clist_count:w #1 , \s__clist_stop \prg_break: , \prg_break_point: 
\cs_new:Npn \clist_count:n ##1 
\int_eval:n \{ \__clist_count:w #1 , ##1 , \s__clist_stop \prg_break: , \prg_break_point: \} 
\exp_args:No \__clist_tmp:w \c_space_tl 
\clist_use:Nnnn #1#2#3#4 
\clist_if_exist:NTF #1 
\int_case:nnF { \clist_count:N #1 } 
\exp_after:wN \__clist_use:wwn #1 , \s__clist_stop \tl_if_blank:nF {##1} { + 1 } 
\__clist_use:wwn #1 
\exp_args:No \__clist_tmp:w \c_space_tl 

(End definition for \clist_count:N and others. These functions are documented on page 176.)

### 59.8 Using comma lists

First check that the variable exists. Then count the items in the comma list. If it has none, output nothing. If it has one item, output that item, brace stripped (note that space-trimming has already been done when the comma list was assigned). If it has two, place the \langle separator between two \rangle in the middle.

Otherwise, \__clist_use:wwwwnn takes the following arguments: 1: a \langle separator \rangle, 2, 3, 4: three items from the comma list (or quarks), 5: the rest of the comma list, 6: a \langle continuation \rangle function (use_ii or use_iii with its \langle separator \rangle argument), 7: junk, and 8: the temporary result, which is built in a brace group following \s__clist_stop. The \langle separator \rangle and the first of the three items are placed in the result, then we use the \langle continuation \rangle, placing the remaining two items after it. When we begin this loop, the three items really belong to the comma list, the first \s__clist_mark is taken as a delimiter to the use_ii function, and the continuation is use_ii itself. When we reach the last two items of the original token list, \s__clist_mark is taken as a third item, and now the second \s__clist_mark serves as a delimiter to use_ii, switching to the other \langle continuation \rangle, use_iii, which uses the \langle separator between final two \rangle.
Items are grabbed by \clist_use:Nw, which detects blank items with a \tl_if_empty:oTF test (in which case it recurses). Non-blank items are either the end of the list, in which case the argument #1 of \clist_use:Nw is used to properly end the list, or are normal items, which must be trimmed and properly unbraced. As we find successive items, the long list of \clist_use:Nw calls gets shortened and we end up calling \clist_use_more:w once we have found 3 items. This auxiliary leaves the first-found item and the general separator, and calls \clist_use:Nw to find more items. A subtlety is that we use \clist_use_end:w both in the case of a two-item list and for the last two items of a general list: to get the correct separator, \clist_use_more:w replaces the separator-of-two by the last-separator when called, namely as soon as we have found three items.
59.9 Using a single item

To avoid needing to test the end of the list at each step, we first compute the \( \langle \text{length} \rangle \) of the list. If the item number is 0, less than \(-\langle \text{length} \rangle\), or more than \(\langle \text{length} \rangle\), the result is empty. If it is negative, but not less than \(-\langle \text{length} \rangle\), add \(\langle \text{length} \rangle + 1\) to the item number before performing the loop. The loop itself is very simple, return the item if the counter reached 1, otherwise, decrease the counter and repeat.

\[
\begin{align*}
\textbf{clist_item:NN} & \ \\
\textbf{clist_item:cn} & \\
\textbf{clist_item:nnN} & \\
\textbf{clist_item:ffoN} & \\
\textbf{clist_item:ffnN} & \\
\end{align*}
\]

\(\text{clist_item:NN}\) loop:nnw

\[
\begin{align*}
\text{cs_new:Npn} \ \text{clist_item:NN} \ #1#2 & \\
\{ & \text{clist_count:N} #1 \\
\{ & \text{int_eval:n} \ #2 \} \\
\#1 & \\
\_\text{clist_item:NN_loop:nnw} & \\
\} & \\
\text{cs_new:Npn} \ \_\text{clist_item:nnN} \ #1\#2#3#4 & \\
\{ & \text{int_compare:nNnTF} \ #2 < 0 \\
\{ & \text{int_compare:nNnTF} \ #2 < \ - \#1 \\
\{ & \_\text{clist_use:non_delimit_by_s:stop} \\
\{ & \text{exp_args:Nf} \ #4 \ \text{int_eval:n} \ \#2 + 1 + \#1 \} \\
\} & \\
\} & \\
\} & , \#3, \_\text{s_clist_stop} \\
\text{cs_generate_variant:Nn} \ \_\text{clist_item:nnN} \ \{ \text{ffo, ff} \} & \\
\text{cs_new:Npn} \ \_\text{clist_item:NN_loop:nnw} \ #1 \#2, & \\
\}
\end{align*}
\]
\int_compare:nNnTF {#1} = 0
\{ \__clist_use_i_delimit_by_s_stop:nw \{ \exp_not:n \{#2\} \}
\{ \exp_args:Nf \__clist_item_N_loop:nw \{ \int_eval:n \{ #1 - 1 \} \} \}
\}
\cs_generate_variant:Nn \clist_item:Nn \{ c \}

(End definition for \clist_item:Nn, \__clist_item:nwN, and \__clist_item_N_loop:nw. This function is documented on page 178.)

This starts in the same way as \clist_item:Nn by counting the items of the comma list. The final item should be space-trimmed before being brace-stripped, hence we insert a couple of odd-looking \prg_do_nothing: to avoid losing braces. Blank items are ignored.

\cs_new:Npn \clist_item:nn \__clist_item_n:nw \__clist_item_n_loop:nw \__clist_item_n_end:n \__clist_item_n_strip:n \__clist_item_n_strip:w

The N-type function is not implemented through the n-type function for efficiency: for instance comma-list variables do not require space-trimming of their items. Even testing for emptyness of an n-type comma-list is slow, so we count items first and use that both for the emptiness test and the pseudo-random integer. Importantly, \clist_item:Nn and \clist_item:mm only evaluate their argument once.

\cs_new:Npn \clist_rand_item:n \__clist_rand_item_n:nw \__clist_rand_item_n_end:n \__clist_rand_item_n_strip:n \__clist_rand_item_n_strip:w

(End definition for \clist_item:mm and others. This function is documented on page 178.)
\end{verbatim}

59.10 Viewing comma lists

Apply the general \_\_kernel\_chk\_tl\_type:NnnT with \exp\_not:o #2 serving as a dummy code to prevent a check performed by this auxiliary.

\begin{verbatim}
\cs_new:Npn \clist\_show:N { \__clist\_show:NN \msg\_show:nnxxxx }
\cs_new_protected:Npn \clist\_log:N { \__clist\_show:NN \msg\_log:nnxxxx }
\cs_new_protected:Npn \__clist\_show:NN #1#2
{ \__kernel\_chk\_tl\_type:NnnT #2 { clist } { \exp\_not:o #2 }
  { \int\_compare:nNnTF { \clist\_count:N #2 }
    = { \exp\_args:No \clist\_count:n #2 }
    { #1 { clist } { show }
      { \token\_to\_str:N #2 }
      { \clist\_map\_function:nN {#2} \msg\_show\_item:n }
      { } { } }
    } { }
  } { \msg\_error:nxxx { clist } { non\_clist }
    { \token\_to\_str:N #2 } { \tl\_to\_str:N #2 } }
}
\end{verbatim}

A variant of the above: no existence check, empty first argument for the message.

\begin{verbatim}
\cs_new_protected:Npn \clist\_show:n { \__clist\_show:Nn \msg\_show:nxxxx }
\cs_new_protected:Npn \clist\_log:n { \__clist\_show:Nn \msg\_log:nxxxx }
\cs_new_protected:Npn \__clist\_show:Nn \clist\_log:n { \__clist\_show:Nn \msg\_log:nxxxx }
\__clist\_show:Nn #1#2
{ #1 { clist } { show }
  { } { \clist\_map\_function:nN {#2} \msg\_show\_item:n } { } { }
}
\end{verbatim}

(End definition for \clist\_show:N, \clist\_log:N, and \__clist\_show:NN. These functions are documented on page 179.)
59.11 Scratch comma lists

\l_tmpa_clist  Temporary comma list variables.
\l_tmpb_clist  \clist_new:N \l_tmpa_clist
\g_tmpa_clist  \clist_new:N \l_tmpb_clist
\g_tmpb_clist  \clist_new:N \g_tmpa_clist
\clist_new:N \g_tmpb_clist

(End definition for \l_tmpa_clist and others. These variables are documented on page 179.)

⟨/package⟩
Chapter 60

l3token implementation

60.1 Internal auxiliaries

\s__char_stop Internal scan mark.
\scan_new:N \s__char_stop
(End definition for \s__char_stop.)

\q__char_no_value Internal recursion quarks.
\quark_new:N \q__char_no_value
(End definition for \q__char_no_value.)

\__char_quark_if_no_value:p:N \__char_quark_if_no_value:NTF Functions to query recursion quarks.
\__kernel_quark_new_conditional:Nn \__char_quark_if_no_value:N { TF }
(End definition for \__char_quark_if_no_value:NTF.)

60.2 Manipulating and interrogating character tokens

Simple wrappers around the primitives.
\cs_new_protected:Npn \char_set_catcode:nn \char_value_catcode:n \char_show_value_catcode:n
{ \tex_catcode:D \int_eval:n {#1} = \int_eval:n {#2} \exp_stop_f: }
\cs_new:Npn \char_value_catcode:n #1
{ \tex_the:D \tex_catcode:D \int_eval:n {#1} \exp_stop_f: }
\cs_new_protected:Npn \char_show_value_catcode:n #1
{ \exp_args:Nf \tl_show:n { \char_value_catcode:n {#1} } }
(End definition for \char_set_catcode:nn, \char_value_catcode:n, and \char_show_value_catcode:n. These functions are documented on page 183.)
End definition for \char_set_catcode_escape:N and others. These functions are documented on page 182.
\cs_new_protected:Npn \char_set_catcode_math_subscript:n #1 \{ \char_set_catcode:nn {#1} { 8 } \}
\cs_new_protected:Npn \char_set_catcode_ignore:n #1 \{ \char_set_catcode:nn {#1} { 9 } \}
\cs_new_protected:Npn \char_set_catcode_space:n #1 \{ \char_set_catcode:nn {#1} { 10 } \}
\cs_new_protected:Npn \char_set_catcode_letter:n #1 \{ \char_set_catcode:nn {#1} { 11 } \}
\cs_new_protected:Npn \char_set_catcode_other:n #1 \{ \char_set_catcode:nn {#1} { 12 } \}
\cs_new_protected:Npn \char_set_catcode_active:n #1 \{ \char_set_catcode:nn {#1} { 13 } \}
\cs_new_protected:Npn \char_set_catcode_comment:n #1 \{ \char_set_catcode:nn {#1} { 14 } \}
\cs_new_protected:Npn \char_set_catcode_invalid:n #1 \{ \char_set_catcode:nn {#1} { 15 } \}

(End definition for \char_set_catcode_escape:n and others. These functions are documented on page 183.)

\char_set_mathcode:nn \char_value_mathcode:n \char_show_value_mathcode:n
\char_set_lccode:nn \char_value_lccode:n \char_show_value_lccode:n
\char_set_uccode:nn \char_value_uccode:n \char_show_value_uccode:n
\char_set_sfcode:nn \char_value_sfcode:n \char_show_value_sfcode:n

Pretty repetitive, but necessary!

\seq_new:N \l_char_special_seq
\seq_set_split:Nnn \l_char_special_seq { } { 850 }

\l_char_active_seq \l_char_special_seq

Two sequences for dealing with special characters. The first is characters which may be active, the second longer list is for “special” characters more generally. Both lists are escaped so that for example bulk code assignments can be carried out. In both cases, the order is by ASCII character code (as is done in for example \ExplSyntaxOn).

\seq_new:N \l_char_special_seq
\seq_set_split:Nnn \l_char_special_seq { } { 850 }

(End definition for \char_set_mathcode:nn and others. These functions are documented on page 184.)
\char_set_active_eq:NN
\char_set_active_eq:Nc
\char_gset_active_eq:NN
\char_gset_active_eq:Nc
\char_set_active_eq:nc
\char_gset_active_eq:nc
\__char_tmp:nN { char_set_active_eq } \cs_set_eq:NN
\__char_tmp:nN { char_gset_active_eq } \cs_gset_eq:NN
\group_end:
\cs_generate_variant:Nn \char_set_active_eq:NN { Nc }
\cs_generate_variant:Nn \char_gset_active_eq:NN { Nc }
\cs_generate_variant:Nn \char_set_active_eq:nc { nc }
\cs_generate_variant:Nn \char_gset_active_eq:nc { nc }
\group_end:
\cs_new_eq:NN \__char_int_to_roman:w \tex_romannumeral:D
\__char_int_to_roman:w
\__char_generate:nn
\__char_generate_aux:nn
\__char_generate_auxii:nn
\l__char_tmp_tl
\__char_generate_invalid_catcode:

\char_generate:nn
\__char_generate_aux:nn
\__char_generate_auxii:nn
\l__char_tmp_tl
\__char_generate_invalid_catcode:

60.3 Creating character tokens

Four simple functions with very similar definitions, so set up using an auxiliary. These are similar to \LaTeX's \letcharcode primitive.

\group_begin:
\char_set_catcode_active:N \^^@
\cs_set_protected:Npn \__char_tmp:nN #1#2
{ \cs_new_protected:cpn { #1 :nN } ##1
{ \group_begin:
\char_set_lccode:nn { '\^^@ } { ##1 }
\tex_lowercase:D { \group_end: #2 ^^@ } }
\cs_new_protected:cpx { #1 :NN } ##1
{ \exp_not:c { #1 : nN } { '##1 } }
\__char_tmp:nN { char_set_active_eq } \cs_set_eq:NN
\__char_tmp:nN { char_gset_active_eq } \cs_gset_eq:NN
\group_end:
\cs_generate_variant:Nn \char_set_active_eq:NN { Nc }
\cs_generate_variant:Nn \char_gset_active_eq:NN { Nc }
\cs_generate_variant:Nn \char_set_active_eq:nc { nc }
\cs_generate_variant:Nn \char_gset_active_eq:nc { nc }
\group_end:
\cs_new_eq:NN \__char_int_to_roman:w \tex_romannumeral:D
\__char_int_to_roman:w
\__char_generate:nn
\__char_generate_aux:nn
\__char_generate_auxii:nn
\l__char_tmp_tl
\__char_generate_invalid_catcode:

\char_generate:nn
\__char_generate_aux:nn
\__char_generate_auxii:nn
\l__char_tmp_tl
\__char_generate_invalid_catcode:

For efficiency in 8-bit engines, we use the faster primitive approach to making roman numerals.

\cs_new_eq:NN \__char_int_to_roman:w \tex_romannumeral:D
\__char_int_to_roman:w

The aim here is to generate characters of (broadly) arbitrary category code. Where possible, that is done using engine support (Xe\TeX, \LaTeX). There are though various issues which are covered below. At the interface layer, turn the two arguments into integers up-front so this is only done once.

\cs_new:Npn \char_generate:nn #1#2
{ \exp:w \exp_after:wN \__char_generate_aux:w
\int_value:w \int_eval:n {#1} \exp_after:wN ;
\int_value:w \int_eval:n {#2} ;
}
Before doing any actual conversion, first some special case filtering. Spaces are out here as LuaTeX emulation only makes normal (charcode 32 spaces). However, \^^@ is filtered out separately as that can’t be done with macro emulation either, so is flagged up separately. That done, hand off to the engine-dependent part.

\begin{verbatim}
\cs_new:Npn \__char_generate_aux:w #1 ; #2 ;
{ \if_int_compare:w #2 = 10 \exp_stop_f:
  \if_int_compare:w #1 = \c_zero_int
  \msg_expandable_error:nn { char } { null-space }
  \else:
  \msg_expandable_error:nn { char } { space }
  \fi:
  \else:
  \if_int_odd:w 0
  \if_int_compare:w #2 < 1 \exp_stop_f: 1 \fi:
  \if_int_compare:w #2 = 5 \exp_stop_f: 1 \fi:
  \if_int_compare:w #2 = 9 \exp_stop_f: 1 \fi:
  \if_int_compare:w #2 > 13 \exp_stop_f: 1 \fi:
  \msg_expandable_error:nn { char }
  { invalid-catcode }
  \else:
  \if_int_odd:w 0
  \if_int_compare:w #1 < \c_zero_int 1 \fi:
  \if_int_compare:w #1 > \c_max_char_int 1 \fi:
  \exp_end:
  \fi:
  \else:
  \__char_generate_aux:nnw {#1} {#2}
  \fi:
  \exp_end:
\tl_new:N \l__char_tmp_tl
\end{verbatim}

Engine-dependent definitions are now needed for the implementation. For LuaTeX and XeLaTeX there is engine-level support. They can do cases that macro emulation can’t. All of those are filtered out here using a primitive-based boolean expression to avoid fixing the category code of the null character used in the false branch (for 8-bit engines). The final level is the basic definition at the engine level: the arguments here are integers so there is no need to worry about them too much. Older versions of XeLaTeX cannot generate active characters so we filter that: at some future stage that may change: the slightly odd ordering of auxiliaries reflects that.

\begin{verbatim}
\group_begin:
\char_set_catcode_active:N \^^L
\cs_set:Npn \^^L { }
\char_set_catcode_other:n { 0 }
\if_int_odd:w 0
\sys_if_engine_luatex:T { 1 }
\sys_if_engine_xetex:T { 1 } \exp_stop_f:
\cs_new:Npn \__char_generate_aux:nnw #1#2\#3 \exp_end:
{ \#3
\tl_new:N \l__char_tmp_tl
\end{verbatim}

852
For engines where \Ucharcat isn’t available or emulated, we have to work in macros, and cover only the 8-bit range. The first stage is to build up a \tl containing \^\^@ with each category code that can be accessed in this way, with an error set up for the other cases. This is all done such that it can be quickly accessed using a \if_case:w low-level conditional. There are a few things to notice here. As \^\^L is \outer we need to locally set it to avoid a problem. To get open/close braces into the list, they are set up using \if_false: pairing and are then x-type expanded together into the desired form.

\begin{verbatim}
\tl_set:Nn \l__char_tmp_tl { \exp_not:N \or: }
\char_set_catcode_group_begin:n { 0 } % {
\tl_put_right:Nn \l__char_tmp_tl { \^\^@ \if_false: } }
\char_set_catcode_group_end:n { 0 }
\tl_put_right:Nn \l__char_tmp_tl { \^\^\^\^@ \\
\tl_set:Nn \l__char_tmp_tl { \l__char_tmp_tl } }
\char_set_catcode_math_toggle:n { 0 }
\tl_put_right:No \l__char_tmp_tl { \use:n { \or: } \^\^@ }
\char_set_catcode_math_superscript:n { 0 }
\tl_put_right:No \l__char_tmp_tl { \or: \^\^\^\^@ }
\char_set_catcode_math_subscript:n { 0 }
\tl_put_right:No \l__char_tmp_tl { \or: \^\^\^\^@ }
\char_set_catcode_space:n { 0 }
\tl_put_right:No \l__char_tmp_tl { \or: \^\^\^\^\^\^@ }
\char_set_catcode_letter:n { 0 }
\tl_put_right:No \l__char_tmp_tl { \or: \^\^\^\^\^\^\^\^\^\^@ }
\char_set_catcode_other:n { 0 }
\tl_put_right:No \l__char_tmp_tl { \or: \^\^\^\^\^\^\^\^\^\^\^\^\^@ }
\char_set_catcode_active:n { 0 }
\end{verbatim}

For making spaces, there needs to be an o-type expansion of a \use:n (or some other tokenization) to avoid dropping the space.
Convert the above temporary list into a series of constant token lists, one for each character code, using \text{\texttt{tex_lowercase:D}} to convert `^^@` in each case. The x-type expansion ensures that \text{\texttt{tex_lowercase:D}} receives the contents of the token list. `^^@` is awkward hence this is done in three parts: up to `^^L`, `^^L` itself and above `^^L`. Notice that at this stage `^^@` is active.

\cs_set_protected:Npn \__char_tmp:n #1
\begin{verbatim}
\char_set_lccode:nn { 0 } {#1}
\char_set_lccode:nn { 32 } {#1}
\exp_args:Nx \tex_lowercase:D
\begin{verbatim}
\tl_const:Nn \exp_not:c { c__char_ \__char_int_to_roman:w #1 _tl }
\exp_not:o \l__char_tmp_tl
\end{verbatim}
\end{verbatim}
\int_step_function:nnN { 0 } { 11 } \__char_tmp:n
\group_begin:
\tl_replace_once:Nnn \l__char_tmp_tl { ^^@ } { \ERROR }
\__char_tmp:n { 12 }
\group_end:
\int_step_function:nnN { 13 } { 255 } \__char_tmp:n
\end{verbatim}

As \TeX{} is very unhappy if it finds an alignment character inside a primitive \texttt{\halign} even when skipping false branches, some precautions are required. \TeX{} is happy if the token is hidden between braces within \texttt{\if_false: ... \fi:].

\cs_new:Npn \__char_generate_aux:nnw #1#2#3 \exp_end:
\begin{verbatim}
\if_false: { \fi:
\exp_after:wN \exp_after:wN
\exp_after:wN \exp_end:
\exp_after:wN \exp_after:wN
\if_case:w #2
\exp_last_unbraced:Nv \exp_stop_f:
\{ c__char_ \__char_int_to_roman:w #1 _tl }
\or: }
\fi:
\fi:
\group_end:
\end{verbatim}

(End definition for \texttt{\char_generate:nn} and others. This function is documented on page \pageref{char_generate:nn}.)

\char_to_utfviii_bytes:n
\begin{verbatim}
\char_to_utfviii_bytes_auxi:n
\char_to_utfviii_bytes_auxii:n
\char_to_utfviii_bytes_outputi:n
\char_to_utfviii_bytes_outputii:n
\char_to_utfviii_bytes_outputiii:n
\char_to_utfviii_bytes_outputiv:n
\char_to_utfviii_bytes_outputv:n
\char_to_utfviii_bytes_outputvi:n
\char_to_utfviii_bytes_outputvn
\char_to_utfviii_bytes_outputvm
\char_to_utfviii_bytes_end:
\end{verbatim}

This code converts a codepoint into the correct UTF-8 representation. In terms of the algorithm itself, see \url{https://en.wikipedia.org/wiki/UTF-8} for the octet pattern.

\cs_new:Npn \char_to_utfviii_bytes:n #1
\begin{verbatim}
\exp_args:Nf \__char_to_utfviii_bytes_auxi:n
\{ \int_eval:n {#1} \}
\end{verbatim}

\cs_new:Npn \__char_to_utfviii_bytes_auxi:n
\begin{verbatim}
\{ \int_eval:n {#1} \}
\end{verbatim}

854
\[ \text{\texttt{\_char_to_utfviii_bytes_outputi:nw}} \]

\[ \text{\texttt{\_char_to_utfviii_bytes_outputii:nw}} \]

\[ \text{\texttt{\_char_to_utfviii_bytes_outputiii:nw}} \]

\[ \text{\texttt{\_char_to_utfviii_bytes_outputiv:nw}} \]

\[ \text{\texttt{\_char_to_utfviii_bytes_end:}} \]

\[ \text{\texttt{\_char_to_utfviii_bytes_output:}} \]

\[ \text{\texttt{\_char_to_utfviii_bytes:}} \]
\char_to_nfd:N
\__char_to_nfd:n
\__char_to_nfd:Nw

Look up any NFD and recursively produce the result.

\cs_new:Npn \char_to_nfd:N #1
\cs_set_eq:NN \__char_to_nfd:n \char_to_nfd:N
\cs_new:Npn \__char_to_nfd:Nw #1#2#3 \s__char_stop

End definition for \char_to_nfd:N, \__char_to_nfd:n, and \__char_to_nfd:Nw. This function is documented on page 307.

\char_lowercase:N
\char_uppercase:N
\char_titlecase:N
\char_foldcase:N
\__char_change_case:nNN
\__char_change_case:nN
\__char_change_case_multi:nN
\__char_change_case_multi:vN
\__char_change_case:NNNNw
\__char_change_case:NNN
\__char_change_case:NNNN
\__char_change_case:NN
\__char_change_case_catcode:N
\char_str_lowercase:N
\char_str_uppercase:N
\char_str_titlecase:N
\char_str_foldcase:N
\__char_str_change_case:nNN
\__char_str_change_case:nN

To ensure that the category codes produced are predictable, every character is regenerated even if it is otherwise unchanged. This makes life a little interesting when we might have multiple output characters: we have to grab each of them and case change them in reverse order to maintain f-type expandability.

\cs_new:Npn \char_lowercase:N #1
\{ \__char_change_case:NN { lower } \char_value_lccode:n #1 \}

856
\cs_new:Npn \char_uppercase:N #1
\cs_new:Npn \char_titlecase:N #1
\cs_new:Npn \char_foldcase:N #1
\cs_new:Npn \__char_change_case:nNN { upper \token_to_str:N #1 _t1 \exp_args:Nf \__char_change_case:nN { #2 { \token_to_str:N #3 } } #3 }
\cs_new:Npn \__char_change_case:NNN #1#2#3
\cs_new:Npn \__char_change_case:NNNN #1#2#3#4
\cs_new:Npn \__char_quark_if_no_value:NNTF #1#2#3
Same story for the string version, except category code is easier to follow. This of course makes this version significantly faster.
\cs_new:Npn \__char_str_change_case:nN #1#2
{\int_compare:nNnTF {#1} = 0
{\tl_to_str:n {#2}}
{\char_generate:nn {#1} {12}}}
\bool_lazy_or:nnF
{\cs_if_exist_p:N \tex_luatexversion:D}
{\cs_if_exist_p:N \tex_XeTeXversion:D}
{\cs_set:Npn \__char_str_change_case:nN #1#2
{\tl_to_str:n {#2}}}

(End definition for \char_lowercase:N and others. These functions are documented on page 182.)

\c_catcode_other_space_tl

Create a space with category code 12: an “other” space.

\tl_const:Nx \c_catcode_other_space_tl \char_generate:nn {‘} {12}

(End definition for \c_catcode_other_space_tl. This function is documented on page 182.)

60.4 Generic tokens

\s__token_mark
\s__token_stop

Internal scan marks.

\scan_new:N \s__token_mark
\scan_new:N \s__token_stop

(End definition for \s__token_mark and \s__token_stop.)

\token_to_meaning:N
\token_to_meaning:c
\token_to_str:N
\token_to_str:c
\c_group_begin_token
\c_group_end_token
\c_math_toggle_token
\c_alignment_token
\c_parameter_token
\c_math_superscript_token
\c_math_subscript_token
\c_space_token
\c_catcode_letter_token
\c_catcode_other_token

These are all defined in l3basics, as they are needed “early”. This is just a reminder!

(End definition for \token_to_meaning:N and \token_to_str:N. These functions are documented on page 186.)

We define these useful tokens. For the brace and space tokens things have to be done by hand: the formal argument spec. for \cs_new_eq:NN does not cover them so we do things by hand. (As currently coded it would work with \cs_new_eq:NN but that’s not really a great idea to show off: we want people to stick to the defined interfaces and that includes us.) So that these few odd names go into the log when appropriate there is a need to hand-apply the \__kernel_chk_if_free_cs:N check.

\group_begin:
\__kernel_chk_if_free_cs:N \c_group_begin_token
\tex_global:D \tex_let:D \c_group_begin_token \}
\__kernel_chk_if_free_cs:N \c_group_end_token
\tex_global:D \tex_let:D \c_group_end_token \}
\char_set_catcode_math_toggle:N \*
\cs_new_eq:NN \c_math_toggle_token *
\char_set_catcode_alignment:N \*
\cs_new_eq:NN \c_alignment_token *
\cs_new_eq:NN \c_parameter_token #
\cs_new_eq:NN \c_math_superscript_token ^
\char_set_catcode_math_subscript:N \*
\cs_new_eq:NN \c_math_subscript_token *
\_kernel_chk_if_free_cs:N \c_space_token
\use:n \{ \tex_global:D \tex_lig:D \c_space_token = \~ \} \~
\cs_new_eq:NN \c_catcode_letter_token a
\cs_new_eq:NN \c_catcode_other_token 1
\group_end:

(End definition for \c_group_begin_token and others. These functions are documented on page 185.)

\c_catcode_active_tl
Not an implicit token!
\group_begin:
\char_set_catcode_active:N \*
\tl_const:Nn \c_catcode_active_tl { \exp_not:N * }
\group_end:

(End definition for \c_catcode_active_tl. This variable is documented on page 185.)

60.5 Token conditionals

\token_if_group_begin_p:N
\token_if_group_begin:N\TF
Check if token is a begin group token. We use the constant \c_group_begin_token for this.
\prg_new_conditional:Nnn \token_if_group_begin:N \#1 { p , T , F , \TF }
\prg_new_conditional:Nnn \token_if_group_begin:N \#1 { p , T , F , \TF }
{ \if_catcode:w \exp_not:N \#1 \c_group_begin_token
 \prg_return_true: \else: \prg_return_false: \fi: }

(End definition for \token_if_group_begin:NTF. This function is documented on page 186.)

\token_if_group_end_p:N
\token_if_group_end:N\TF
Check if token is an end group token. We use the constant \c_group_end_token for this.
\prg_new_conditional:Nnn \token_if_group_end:N \#1 { p , T , F , \TF }
\prg_new_conditional:Nnn \token_if_group_end:N \#1 { p , T , F , \TF }
{ \if_catcode:w \exp_not:N \#1 \c_group_end_token
 \prg_return_true: \else: \prg_return_false: \fi: }

(End definition for \token_if_group_end:NTF. This function is documented on page 186.)

\token_if_math_toggle_p:N
\token_if_math_toggle:N\TF
Check if token is a math shift token. We use the constant \c_math_toggle_token for this.
\prg_new_conditional:Nnn \token_if_math_toggle:N \#1 { p , T , F , \TF }
\prg_new_conditional:Nnn \token_if_math_toggle:N \#1 { p , T , F , \TF }
{ \if_catcode:w \exp_not:N \#1 \c_math_toggle_token
 \prg_return_true: \else: \prg_return_false: \fi: }

(End definition for \token_if_math_toggle:NTF. This function is documented on page 186.)

\token_if_alignment_p:N
\token_if_alignment:N\TF
Check if token is an alignment tab token. We use the constant \c_alignment_token for this.
\prg_new_conditional:Nnn \token_if_alignment:N \#1 { p , T , F , \TF }
\prg_new_conditional:Nnn \token_if_alignment:N \#1 { p , T , F , \TF }
{ \if_catcode:w \exp_not:N \#1 \c_alignment_token
 \prg_return_true: \else: \prg_return_false: \fi: }

(End definition for \token_if_alignment:NTF. This function is documented on page 186.)
Check if token is a parameter token. We use the constant \texttt{c\_parameter\_token} for this. We have to trick \LaTeX{} a bit to avoid an error message: within a group we prevent \texttt{c\_parameter\_token} from behaving like a macro parameter character. The definitions of \prg\_new\_conditional\_Npnn are global, so they remain after the group.

\begin{verbatim}
\groupBegin
\cs_set_eq:NN \c_parameter_token \scan_stop:
\prg_new_conditional:Npnn \token_if_parameter:N #1 { p , T , F , TF }
\begin{Verbatim}
\{ \if_catcode:w \exp_not:N #1 \c_parameter_token \prg_return_true: \else: \prg_return_false: \fi:
\}
\groupEnd
\end{verbatim}

Check if token is a math superscript token. We use the constant \texttt{c\_math\_superscript\_token} for this.

\begin{verbatim}
\prg_new_conditional:Npnn \token_if_math_superscript:N #1 { p , T , F , TF }
\begin{Verbatim}
\{ \if_catcode:w \exp_not:N #1 \c_math_superscript_token \prg_return_true: \else: \prg_return_false: \fi:
\}
\end{verbatim}

Check if token is a math subscript token. We use the constant \texttt{c\_math\_subscript\_token} for this.

\begin{verbatim}
\prg_new_conditional:Npnn \token_if_math_subscript:N #1 { p , T , F , TF }
\begin{Verbatim}
\{ \if_catcode:w \exp_not:N #1 \c_math_subscript_token \prg_return_true: \else: \prg_return_false: \fi:
\}
\end{verbatim}

Check if token is a space token. We use the constant \texttt{c\_space\_token} for this.

\begin{verbatim}
\prg_new_conditional:Npnn \token_if_space:N #1 { p , T , F , TF }
\begin{Verbatim}
\{ \if_catcode:w \exp_not:N #1 \c_space_token \prg_return_true: \else: \prg_return_false: \fi:
\}
\end{verbatim}

Check if token is a letter token. We use the constant \texttt{c\_catcode\_letter\_token} for this.

\begin{verbatim}
\prg_new_conditional:Npnn \token_if_letter:N #1 { p , T , F , TF }
\begin{Verbatim}
\{ \if_catcode:w \exp_not:N #1 \c_catcode_letter_token \prg_return_true: \else: \prg_return_false: \fi:
\}
\end{verbatim}

(End definition for \texttt{\token\_if\_alignment\_NTF}. This function is documented on page 187.)

(End definition for \texttt{\token\_if\_parameter\_NTF}. This function is documented on page 187.)

(End definition for \texttt{\token\_if\_math\_superscript\_NTF}. This function is documented on page 187.)

(End definition for \texttt{\token\_if\_math\_subscript\_NTF}. This function is documented on page 187.)

(End definition for \texttt{\token\_if\_space\_NTF}. This function is documented on page 187.)

(End definition for \texttt{\token\_if\_letter\_NTF}. This function is documented on page 187.)
\token_if_other_p:N  \token_if_other:NTF Check if token is an other char token. We use the constant \c_catcode_other_token for this.
\begin{verbatim}
18912 \prg_new_conditional:Nnnn \token_if_other:N #1 { p , T , F , TF }
18913 { 18914 \if_catcode:w \exp_not:N #1 \c_catcode_other_token
18915 \prg_return_true: \else: \prg_return_false: \fi: 18916 }
\end{verbatim}
(End definition for \token_if_other:NTF. This function is documented on page 187.)

\token_if_active_p:N  \token_if_active:NTF Check if token is an active char token. We use the constant \c_catcode_active_tl for this. A technical point is that \c_catcode_active_tl is in fact a macro expanding to \exp_not:N *, where * is active.
\begin{verbatim}
18917 \prg_new_conditional:Nnnn \token_if_active:N #1 { p , T , F , TF }
18918 { 18919 \if_catcode:w \exp_not:N #1 \c_catcode_active_tl
18920 \prg_return_true: \else: \prg_return_false: \fi: 18921 }
\end{verbatim}
(End definition for \token_if_active:NTF. This function is documented on page 187.)

\token_if_eq_meaning_p:NN  \token_if_eq_meaning:NNTF Check if the tokens #1 and #2 have same meaning.
\begin{verbatim}
18922 \prg_new_conditional:Nnnn \token_if_eq_meaning:NN #1#2 { p , T , F , TF }
18923 { 18924 \if_meaning:w #1 #2
18925 \prg_return_true: \else: \prg_return_false: \fi: 18926 }
\end{verbatim}
(End definition for \token_if_eq_meaning:NNTF. This function is documented on page 188.)

\token_if_eq_catcode_p:NN  \token_if_eq_catcode:NNTF Check if the tokens #1 and #2 have same category code.
\begin{verbatim}
18927 \prg_new_conditional:Nnnn \token_if_eq_catcode:NN #1#2 { p , T , F , TF }
18928 { 18929 \if_catcode:w \exp_not:N #1 \exp_not:N #2
18930 \prg_return_true: \else: \prg_return_false: \fi: 18931 }
\end{verbatim}
(End definition for \token_if_eq_catcode:NNTF. This function is documented on page 187.)

\token_if_eq_charcode_p:NN  \token_if_eq_charcode:NNTF Check if the tokens #1 and #2 have same character code.
\begin{verbatim}
18932 \prg_new_conditional:Nnnn \token_if_eq_charcode:NN #1#2 { p , T , F , TF }
18933 { 18934 \if_charcode:w \exp_not:N #1 \exp_not:N #2
18935 \prg_return_true: \else: \prg_return_false: \fi: 18936 }
\end{verbatim}
(End definition for \token_if_eq_charcode:NNTF. This function is documented on page 187.)

\token_if_macro_p:N  \token_if_macro:NTF When a token is a macro, \token_to_meaning:N always outputs something like \long macro:#1->#1 so we could naively check to see if the meaning contains \_. However, this can fail the five \...mark primitives, whose meaning has the form \...mark:\langle user material \rangle. The problem is that the \langle user material \rangle can contain \_. However, only characters, macros, and marks can contain the colon character. The idea is thus to grab until the first :, and analyse what is left. However, macros can have
any combination of \long, \protected or \outer (not used in \LaTeX3) before the string \texttt{macro}:. We thus only select the part of the meaning between the first \texttt{ma} and the first following \\texttt{.} If this string is \texttt{cro}, then we have a macro. If the string is \texttt{rk}, then we have a mark. The string can also be \texttt{cro parameter character} for a colon with a weird category code (namely the usual category code of \texttt{#}). Otherwise, it is empty.

This relies on the fact that \texttt{long}, \texttt{protected}, \texttt{outer} cannot contain \texttt{ma}, regardless of the escape character, even if the escape character is \texttt{m}... Both \texttt{ma} and : must be of category code 12 (other), so are detokenized.

\begin{verbatim}
\use:x
\do\prg_new_conditional:Npnn \exp_not:N \token_if_macro:N ##1 { p, T, F, TF }
\do\do\{ p, T, F, TF \}
\do\{ \exp_not:N \exp_after:wN \exp_not:N \__token_if_macro_p:w
\do\exp_not:N \token_to_meaning:N ##1 \tl_to_str:n { ma : }
\do\s__token_stop
\do\cs_new:Npn \exp_not:N \__token_if_macro_p:w
\do##1 \tl_to_str:n { ma } ##2 \c_colon_str ##3 \s__token_stop
\do\{ \str_if_eq:nnTF { #2 } { cro } { \prg_return_true: } { \prg_return_false: } \fi:
\do\end{verbatim}

(End definition for \texttt{token_if_macro:NTF and \_\_token_if_macro_p:w}. This function is documented on page 188.)

\texttt{\token_if_cs_p:N, \token_if_cs:NTF} Check if token has same catcode as a control sequence. This follows the same pattern as for \texttt{token_if_letter:N} \etc. We use \texttt{\scan_stop:} for this.

\begin{verbatim}
\do\prg_new_conditional:Npnn \exp_not:N \token_if_cs:N #1 { p, T, F, TF }
\do\do\{ p, T, F, TF \}
\do\{ \if_catcode:w \exp_not:N #1 \scan_stop:
\do\exp_not:N \token_to_meaning:N \#1 \tl_to_str:n { #2 \c_colon_str #3 \s__token_stop
\do\{ \prg_return_true: \}
\do\{ \prg_return_false: \}
\do\end{verbatim}

(End definition for \texttt{token_if_cs:NTF}. This function is documented on page 188.)

\texttt{\token_if_expandable_p:N, \token_if_expandable:NTF} Check if token is expandable. We use the fact that \TeX temporarily converts \texttt{\exp_not:N \{token\}} into \texttt{\scan_stop:} if \texttt{\{token\}} is expandable. An \texttt{undefined} token is not considered as expandable. No problem nesting the conditionals, since the third \texttt{\#1} is only skipped if it is non-expandable (hence not part of \TeX’s conditional apparatus).

\begin{verbatim}
\do\prg_new_conditional:Npnn \exp_not:N \token_if_expandable:N #1 { p, T, F, TF }
\do\do\{ p, T, F, TF \}
\do\{ \exp_after:wN \if_meaning:w \exp_not:N \#1 \prg_return_false:
\do\else:
\do\{ \if_cs_exist:N #1 \prg_return_true: \}
\do\else:
\do\{ \prg_return_false: \}
\do\fi:
\end{verbatim}

863
These auxiliary functions are used below to define some conditionals which detect whether the \meaning of their argument begins with a particular string. Each auxiliary takes an argument delimited by a string, a second one delimited by \_\_token_stop, and returns the first one and its delimiter. This result is eventually compared to another string. Note that the “font” auxiliary is delimited by a space followed by “font”. This avoids an unnecessary check for the \font primitive below.

Each of these conditionals tests whether its argument’s \meaning starts with a given string. This is essentially done by having an auxiliary grab an argument delimited by the string and testing whether the argument was empty. Of course, a copy of this string must first be added to the end of the \meaning to avoid a runaway argument in case it does not contain the string. Two complications arise. First, the escape character is not fixed, and cannot be included in the delimiter of the auxiliary function (this function cannot be defined on the fly because tests must remain expandable): instead the first argument of the auxiliary (plus the delimiter to avoid complications with trailing spaces) is compared using \str_if_eq:eeTF to the result of applying \token_to_str:N to a control sequence. Second, the \meaning of primitives such as \dimen or \dimendef starts in the same way as registers such as \dimen123, so they must be tested for.

Characters used as delimiters must have catcode 12 and are obtained through \tl_to_str:n. This requires doing all definitions within x-expansion. The temporary function \_\_token_tmp:w used to define each conditional receives three arguments: the name of the conditional, the auxiliary’s delimiter (also used to name the auxiliary), and the string to which one compares the auxiliary’s result. Note that the \meaning of a protected long macro starts with \protected\long macro, with no space after \protected but a space after \long, hence the mixture of \token_to_str:N and \token_to_str:n.

For the first six conditionals, \cs_if_exist:cT turns out to be false (thanks to the leading space for \font), and the code boils down to a string comparison between
the result of the auxiliary on the \texttt{meaning} of the conditional’s argument \texttt{###1}, and \texttt{#3}. Both are evaluated at run-time, as this is important to get the correct escape character.

The other five conditionals have additional code that compares the argument \texttt{###1} to two \TeX{} primitives which would wrongly be recognized as registers otherwise. Despite using \TeX{}’s primitive conditional construction, this does not break when \texttt{###1} is itself a conditional, because branches of the conditionals are only skipped if \texttt{###1} is one of the two primitives that are tested for (which are not \TeX{} conditionals).

\begin{verbatim}
\group_begin:
\cs_set_protected:Npn \__token_tmp:w #1#2#3
  \use:x
  \prg_new_conditional:Npnn \exp_not:c { token_if_ #1 :N } ###1
    { p , T , F , TF }
    { \cs_if_exist:cT { tex_ #2 :D }
        { \exp_not:N \if_meaning:w ###1 \exp_not:c { tex_ #2 :D }
            \exp_not:N \prg_return_false:
            \exp_not:N \else:
            \exp_not:N \if_meaning:w ###1 \exp_not:c { tex_ #2 def:D }
            \exp_not:N \prg_return_false:
            \exp_not:N \else:
            \exp_not:N \str_if_eq:eeTF
            { \exp_not:N \exp_after:wN
              \exp_not:c { __token_delimit_by_ #2 :w }
              \exp_not:N \token_to_meaning:N ###1
            }{ \tl_to_str:n {#2} \s__token_stop }
        { \exp_not:n {#3} }
        { \exp_not:N \prg_return_false: }
        \cs_if_exist:cT { tex_ #2 :D }
        { \exp_not:N \fi:
            \exp_not:N \fi:
        }
    }
    \__token_tmp:w { chardef } { char" } { \token_to_str:N \char" }
  \__token_tmp:w { mathchardef } { char" } { \token_to_str:N \mathchar" }
  \__token_tmp:w { long_macro } { macro } { \tl_to_str:n { \long } macro }
  \__token_tmp:w { protected_macro } { macro }
  \__token_tmp:w { protected_long_macro } { macro }
  \__token_to_str:N \protected \tl_to_str:n { \long } macro
  \__token_tmp:w { dim_register } { dimen } { \token_to_str:N \dimen }
  \__token_tmp:w { int_register } { count } { \token_to_str:N \count }
  \__token_tmp:w { muskip_register } { muskip } { \token_to_str:N \muskip }
\end{verbatim}

865
We filter out macros first, because they cause endless trouble later otherwise.

Primitives are almost distinguished by the fact that the result of \token_to_-meaning:N is formed from letters only. Every other token has either a space (e.g., the letter A), a digit (e.g., \count{123}) or a double quote (e.g., \char{"A}).

Ten exceptions: on the one hand, \tex Undefined:D is not a primitive, but its meaning is undefined, only letters; on the other hand, \space, \italiccorr, \hyphen, \firstmark, \topmark, \botmark, \splitfirstmark, \splitbotmark, and \nullfont are primitives, but have non-letters in their meaning.

We start by removing the two first (non-space) characters from the meaning. This removes the escape character (which may be nonexistent depending on \endlinechar), and takes care of three of the exceptions: \space, \italiccorr and \hyphen, whose meaning is at most two characters. This leaves a string terminated by some \s__token_stop.

The meaning of each one of the five \...mark primitives has the form ⟨letters⟩:(user material). In other words, the first non-letter is a colon. We remove everything after the first colon.

We are now left with a string, which we must analyze. For primitives, it contains only letters. For non-primitives, it contains either ", or a space, or a digit. Two exceptions remain: \tex Undefined:D, which is not a primitive, and \nullfont, which is a primitive.

Spaces cannot be grabbed in an undelimited way, so we check them separately. If there is a space, we test for \nullfont. Otherwise, we go through characters one by one, and stop at the first character less than ‘A’ (this is not quite a test for “only letters”, but is close enough to work in this context). If this first character is : then we have a primitive, or \tex Undefined:D, and if it is ” or a digit, then the token is not a primitive.

For LuaTeX we use a different implementation which just looks at the command code for the token and compares it to a list of non-primitives. Again, \nullfont is a special case because it is the only primitive with the normally non-primitive \setfont command code.

In LuaMetaTeX some of the command names are different, so we check for both versions. The first one is always the LuaTeX version.
local mode_below_biggest_char = {}
local index_not_nil = {}
local mode_not_null = {}
local non_primitive = {
    [cmd'left_brace'] = true,
    [cmd'right_brace'] = true,
    [cmd'math_shift'] = true,
    [cmd'mac_param' or cmd'parameter'] = mode_below_biggest_char,
    [cmd'sup_mark' or cmd'superscript'] = true,
    [cmd'sub_mark' or cmd'subscript'] = true,
    [cmd'endv' or cmd-ignore'] = true,
    [cmd'spacer'] = true,
    [cmd'letter'] = true,
    [cmd'other_char'] = true,
    [cmd'tab_mark' or cmd'alignment_tab'] = mode_below_biggest_char,
    [cmd'char_given'] = true,
    [cmd'math_given' or 'math_char_given'] = true,
    [cmd'xmath_given' or 'math_char_xgiven'] = true,
    [cmd'set_font'] = mode_not_null,
    [cmd'undefined_cs'] = true,
    [cmd'call'] = true,
    [cmd'long_call' or cmd'protected_call'] = true,
    [cmd'outer_call' or cmd'tolerant_call'] = true,
    [cmd'long_outer_call' or cmd'tolerant_protected_call'] = true,
    [cmd'assign_glue' or cmd'register_glue'] = index_not_nil,
    [cmd'assign_mu_glue' or cmd'register_mu_glue'] = index_not_nil,
    [cmd'assign_toks' or cmd'register_toks'] = index_not_nil,
    [cmd'assign_int' or cmd'register_int'] = index_not_nil,
    [cmd'assign_attr' or cmd'register_attribute'] = true,
    [cmd'assign_dimen' or cmd'register_dimen'] = index_not_nil,
}

luacmd("__token_if_primitive_lua:N", function()
local tok = get_next()
local is_non_primitive = non_primitive[get_command(tok)]
return put_next(
    is_non_primitive == true
    and false_tok
    or is_non_primitive == nil
    and true_tok
    or is_non_primitive == mode_not_null
    and (get_mode(tok) == 0 and true_tok or false_tok)
    or is_non_primitive == index_not_nil
    and (get_index(tok) and false_tok or true_tok)
    or is_non_primitive == mode_below_biggest_char
    and (get_mode(tok) > biggest_char and true_tok or false_tok)
end, "global")
\use:x

{ p , T , F , TF } \exp_not:N \token_if_primitive:N ##1
\exp_not:N \token_if_macro:NTF ##1
\exp_not:N \prg_return_false:
\exp_not:N \exp_after:wN \exp_not:N \__token_if_primitive:NNw ##1
\tl_to_str:n { : : : } \s__token_stop
\cs_new:Npn \exp_not:N \__token_if_primitive_space:w #1 ~ { }
\cs_new:Npn \__token_if_primitive_nullfont:N #1
{ \if_meaning:w \tex_nullfont:D #1 \prg_return_true: \else: \prg_return_false: \fi: }
\cs_new:Npn \__token_if_primitive_loop:N #1
{ \if_int_compare:w #1 < \c__token_A_int % \exp_after:wN \__token_if_primitive:Nw \exp_after:wN \__token_if_primitive_loop:N \else: \exp_after:wN \__token_if_primitive_loop:N \fi: }
\cs_new:Npn \__token_if_primitive_nullfont:Nw #1 \a__token_stop
\cs_new:Npn \__token_if_primitive_space:w #1 \a__token_stop
\cs_new:Npn \__token_if_primitive_nullfont:Nw #1 \a__token_stop
\cs_new:Npn \__token_if_primitive_loop:Nw #1 \a__token_stop
\if:w : #1 \exp_after:wN \__token_if_primitive_undefined:N \else: \exp_after:wN \use_none:n \fi:
The aim here is to allow the case statement to be evaluated using a known number of expansion steps (two), and without needing to use an explicit “end of recursion” marker. That is achieved by using the test input as the final case, as this is always true. The trick is then to tidy up the output such that the appropriate case code plus either the true or false branch code is inserted.
To tidy up the recursion, there are two outcomes. If there was a hit to one of the cases searched for, then \#1 is the code to insert, \#2 is the next case to check on and \#3 is all of the rest of the cases code. That means that \#4 is the true branch code, and \#5 tidies up the spare \texttt{\_token\_mark} and the false branch. On the other hand, if none of the cases matched then we arrive here using the “termination” case of comparing the search with itself. That means that \#1 is empty, \#2 is the first \texttt{\_token\_mark} and so \#4 is the false code (the true code is mopped up by \#3).

\begin{verbatim}
\cs_new:Npn \_token_case_end:nw #1#2#3 \_token\_mark #4#5 \_token\_stop
{ \exp_end: #1 #4 }
\end{verbatim}

(End definition for \texttt{\_token\_case\_catcode:NnTF} and others. These functions are documented on page 190.)

### 60.6 Peeking ahead at the next token

Peeking ahead is implemented using a two part mechanism. The outer level provides a defined interface to the lower level material. This allows a large amount of code to be shared. There are four cases:

1. peek at the next token;
2. peek at the next non-space token;
3. peek at the next token and remove it;
4. peek at the next non-space token and remove it.

\texttt{\_l\_peek\_token} \quad Storage tokens which are publicly documented: the token peeked.

\texttt{\_g\_peek\_token} \quad (End definition for \texttt{\_l\_peek\_token} and \texttt{\_g\_peek\_token}. These variables are documented on page 190.)

\texttt{\_l\_peek\_search\_token} \quad The token to search for as an implicit token: cf. \texttt{\_l\_peek\_search\_tl}.

\texttt{\_l\_peek\_search\_tl} \quad (End definition for \texttt{\_l\_peek\_search\_token}.)

\texttt{\_l\_peek\_search\_tl} \quad The token to search for as an explicit token: cf. \texttt{\_l\_peek\_search\_token}.

\texttt{\_l\_peek\_search\_tl} \quad (End definition for \texttt{\_l\_peek\_search\_tl}.)

\texttt{\_peek\_true:w} \quad Functions used by the branching and space-stripping code.

\texttt{\_peek\_true\_aux:w} \quad (End definition for \texttt{\_peek\_true:w} and others.)
\__peek_mark
\__peek_stop

Internal scan marks.
\begin{verbatim}
\scan_new:N \s__peek_mark
\scan_new:N \s__peek_stop
\end{verbatim}

(End definition for \s__peek_mark and \s__peek_stop.)

\_peek_use_none_delimit_by_s_stop:w

Functions to gobble up to a scan mark.
\begin{verbatim}
\cs_new:Npn \_peek_use_none_delimit_by_s_stop:w #1 \s__peek_stop { }
\end{verbatim}

(End definition for \_peek_use_none_delimit_by_s_stop:w)

\peek_after:Nw \peek_gafter:Nw

Simple wrappers for \futurelet: no arguments absorbed here.
\begin{verbatim}
\cs_new_protected:Npn \peek_after:Nw { \tex_futurelet:D \l_peek_token }
\cs_new_protected:Npn \peek_gafter:Nw { \tex_global:D \tex_futurelet:D \g_peek_token }
\end{verbatim}

(End definition for \peek_after:Nw and \peek_gafter:Nw. These functions are documented on page 190.)

\__peek_true_remove:w

A function to remove the next token and then regain control.
\begin{verbatim}
\cs_new_protected:Npn \__peek_true_remove:w {
\tex_afterassignment:D \__peek_true_aux:w
\cs_set_eq:NN \__peek_tmp:w \exp_not:n {#1}
\group_align_safe_begin:
\cs_set:Npn \__peek_true_aux:w { \peek_after:Nw \__peek_remove_spaces: }
\__peek_true_aux:w
\group_align_safe_end:
\exp_after:wN \__peek_false:w
\fi:
}
\end{verbatim}

(End definition for \__peek_true_remove:w)

\peek_remove_spaces:n \__peek_remove_spaces:

Repeatedly use \__peek_true_remove:w to remove a space and call \__peek_true_remove:w.
\begin{verbatim}
\cs_new_protected:Npn \peek_remove_spaces:n #1 \__peek_remove_spaces:
\cs_set:Npx \__peek_false:w { \exp_not:n {#1} }
\group_align_safe_begin:
\cs_set:Npn \__peek_true_aux:w { \peek_after:Nw \__peek_remove_spaces: }
\__peek_true_aux:w
\group_align_safe_end:
\exp_after:wN \__peek_false:w
\fi:
}
\end{verbatim}

(End definition for \peek_remove_spaces:n and \__peek_remove_spaces:. This function is documented on page 308.)
The generic functions store the test token in both implicit and explicit modes, and the `true` and `false` code as token lists, more or less. The two branches have to be absorbed here as the input stream needs to be cleared for the peek function itself. Here, `#1` is \_\_\_peek\_true\_remove:w when removing the token and \_\_\_peek\_true\_aux:w otherwise.

\begin{verbatim}
\cs_new_protected:Npn \_\_\_peek\_token\_generic\_aux:NNNTF \#1\#2\#3\#4\#5
\group_align_safe_begin:
\cs_set_eq:NN \l__peek_search_token \#3
\tl_set:Nn \l__peek_search_tl {\#3}
\cs_set:Npx \_\_\_peek\_true\_aux:w
\exp_not:N \group_align_safe_end:
\exp_not:n {\#4}
\cs_set_eq:NN \_\_\_peek\_true:w \#1
\cs_set:Npx \_\_\_peek\_false:w
\exp_not:N \group_align_safe_end:
\exp_not:n \{\#5\}
\peek_after:Nw \#2
\}
\end{verbatim}

\begin{verbatim}
(End definition for \_\_\_peek\_token\_generic\_aux:NNNTF.)
\end{verbatim}

For token removal there needs to be a call to the auxiliary function which does the work.

\begin{verbatim}
\cs_new_protected:Npn \_\_\_peek\_token\_generic:NNTF \__\_peek\_true\_aux:w
\cs_new_protected:Npn \_\_\_peek\_token\_generic:NNT \#1\#2\#3
\{ \_\_\_peek\_token\_generic\_aux:NNNTF \__\_peek\_true\_aux:w \}
\cs_new_protected:Npn \_\_\_peek\_token\_remove\_generic:NNTF \__\_peek\_true\_remove:w
\cs_new_protected:Npn \_\_\_peek\_token\_remove\_generic:NNT \#1\#2\#3
\{ \_\_\_peek\_token\_remove\_generic\_aux:NNNTF \__\_peek\_true\_remove:w \}
\cs_new_protected:Npn \_\_\_peek\_execute\_branches\_meaning:
\{ \_\_\_peek\_execute\_branches\_meaning:
\}
\end{verbatim}

\begin{verbatim}
(End definition for \_\_\_peek\_token\_generic:NNTF and \_\_\_peek\_token\_remove\_generic:NNTF.)
\end{verbatim}

The meaning test is straight forward.

\begin{verbatim}
\cs_new:Npn \_\_\_peek\_execute\_branches\_meaning:
\{
\if_meaning:w \l_peek_token \l__peek\_search\_token
\exp_after:wN \__\_peek\_true:w
\else:
\exp_after:wN \__\_peek\_false:w
\fi:
\}
\end{verbatim}

(End definition for \_\_\_peek\_execute\_branches\_meaning:.)

872
The catcode and charcode tests are very similar, and in order to use the same auxiliaries we do something a little bit odd, firing \if_catcode:w and \if_charcode:w before finding the operands for those tests, which are only given in the \auxii:N and \auxiii: auxiliaries. For our purposes, three kinds of tokens may follow the peeking function:

- control sequences which are not equal to a non-active character token (e.g., macro, primitive);
- active characters which are not equal to a non-active character token (e.g., macro, primitive);
- explicit non-active character tokens, or control sequences or active characters set equal to a non-active character token.

The first two cases are not distinguishable simply using \TeX’s \futurelet, because we can only access the \meaning of tokens in that way. In those cases, detected thanks to a comparison with \scan_stop:, we grab the following token, and compare it explicitly with the explicit search token stored in \_\_peek_search_tl. The \exp_not:N prevents outer macros (coming from non-\LPX code) from blowing up. In the third case, \_\_peek_token is good enough for the test, and we compare it again with the explicit search token. Just like the peek token, the search token may be of any of the three types above, hence the need to use the explicit token that was given to the peek function.

\begin{verbatim}
\cs_new:Npn \_\_peek_execute_branches_catcode:
\{ \if_catcode:w \_\_peek_execute_branches_catcode_aux: \}
\cs_new:Npn \_\_peek_execute_branches_charcode:
\{ \if_charcode:w \_\_peek_execute_branches_catcode_aux: \}
\cs_new:Npn \_\_peek_execute_branches_catcode_aux:
\{ \if_catcode:w \exp_not:N \l_peek_token \scan_stop:
\exp_after:wN \exp_after:wN \exp_after:wN \_\_peek_execute_branches_catcode_auxii:N \exp_after:wN \exp_not:N \else: \exp_after:wN \exp_not:N \l_peek_search_tl \exp_after:wN \exp_not:N \__peek_true:w \else: \exp_after:wN \exp_not:N \__peek_false:w \fi: \}
\cs_new:Npn \_\_peek_execute_branches_catcode_auxii:N #1
\{ \exp_not:N #1 \exp_after:wN \exp_not:N \l_peek_search_tl \exp_after:wN \exp_not:N \__peek_true:w \else: \exp_after:wN \exp_not:N \__peek_false:w \fi: \}
\cs_new:Npn \_\_peek_execute_branches_catcode_auxiii:
\{ \exp_not:N \l_peek_token \exp_after:wN \exp_not:N \l_peek_search_tl \exp_after:wN \exp_not:N \__peek_true:w \else: \exp_after:wN \exp_not:N \__peek_false:w \fi: \}
\end{verbatim}

873
The public functions themselves cannot be defined using \texttt{prg\_new\_conditional:Npnn}. Instead, the TF, T, F variants are defined in terms of corresponding variants of \texttt{\_\_peek\_token\_generic:NNTF} or \texttt{\_\_peek\_token\_remove\_generic:NNTF}, with first argument one of \texttt{\_\_peek\_execute\_branches\_catcode:}, \texttt{\_\_peek\_execute\_branches\_charcode:}, or \texttt{\_\_peek\_execute\_branches\_meaning:}.

To ignore spaces, remove them using \texttt{\_\_peek\_remove\_spaces:n} before running the tests.
All tokens are N-type tokens, except in four cases: begin-group tokens, end-group tokens, space tokens with character code 32, and outer tokens. Since \l_peek_token might be outer, we cannot use the convenient bool_if:nTF function, and must resort to the old trick of using \ifodd to expand a set of tests. The false branch of this test is taken if the token is one of the first three kinds of non-N-type tokens (explicit or implicit), thus we call \__peek_false:w. In the true branch, we must detect outer tokens, without impacting performance too much for non-outer tokens. The first filter is to search for outer in the \meaning of \l_peek_token. If that is absent, \__peek_use_none_delimit_by_s_stop:w cleans up, and we call \__peek_true:w. Otherwise, the token can be a non-outer macro or a primitive mark whose parameter or replacement text contains outer, it can be the primitive \outer, or it can be an outer token. Macros and marks would have ma in the part before the first occurrence of outer; the meaning of \outer has nothing after outer, contrarily to outer macros; and that covers all cases, calling \__peek_true:w or \__peek_false:w as appropriate. Here, there is no ⟨search token⟩, so we feed a dummy \scan_stop: to the \__peek_token_generic:NTF function.
(End definition for \texttt{\_\_peek\_type:TF} and others. This function is documented on page 193.)

\textit{\texttt{\_\_peek\_type:TF}}
Chapter 61

**l3prop implementation**

The following test files are used for this code: m3prop001, m3prop002, m3prop003, m3prop004, m3show001.

A property list is a macro whose top-level expansion is of the form

\__prop \__prop_pair:wn (key) \__prop {value} 
...
\__prop_pair:wn (keyn) \__prop {valuen}

where \__prop is a scan mark (equal to \scan_stop:), and \__prop_pair:wn can be used to map through the property list.

\__prop The internal token used at the beginning of property lists. This is also used after each \langle key \rangle (see \__prop_pair:wn).

(End definition for \__prop.)

\__prop_pair:wn \__prop_pair:wn (key) \__prop {item}

The internal token used to begin each key–value pair in the property list. If expanded outside of a mapping or manipulation function, an error is raised. The definition should always be set globally.

(End definition for \__prop_pair:wn.)

\l__prop_internal_tl Token list used to store new key–value pairs to be inserted by functions of the \prop-put:Nnn family.

(End definition for \l__prop_internal_tl.)
\__prop_split:NnTF \__prop_split:NnTF \{property list\} \{\langle key\rangle\} \{\langle true code\rangle\} \{\langle false code\rangle\}

Splits the \{property list\} at the \{key\}, giving three token lists: the \{extract\} of \{property list\} before the \{key\}, the \{value\} associated with the \{key\} and the \{extract\} of the \{property list\} after the \{value\}. Both \{extracts\} retain the internal structure of a property list, and the concatenation of the two \{extracts\} is a property list. If the \{key\} is present in the \{property list\} then the \{true code\} is left in the input stream, with \#1, \#2, and \#3 replaced by the first \{extract\}, the \{value\}, and the second extract. If the \{key\} is not present in the \{property list\} then the \{false code\} is left in the input stream, with no trailing material. Both \{true code\} and \{false code\} are used in the replacement text of a macro defined internally, hence macro parameter characters should be doubled, except \#1, \#2, and \#3 which stand in the \{true code\} for the three extracts from the property list. The \{key\} comparison takes place as described for \str_if_eq:nn.

\s__prop A private scan mark is used as a marker after each key, and at the very beginning of the property list.
\s__prop_pair:wn The delimiter is always defined, but when misused simply triggers an error and removes its argument.
\l__prop_internal_tl Token list used to store the new key–value pair inserted by \prop_put:Nnn and friends.
\c_empty_prop An empty prop.
\s__prop_mark \s__prop_stop Internal scan marks.
\q__prop_recursion_tail \q__prop_recursion_stop Internal recursion quarks.
\__kernel_quark_new_test:N \__prop_if_recursion_tail_stop:n Functions to query recursion quarks.
\__prop_if_recursion_tail_stop:o

61.1 Internal auxiliaries

Updated: 2013-01-08
61.2 Allocation and initialisation

Property lists are initialized with the value `\c_empty_prop`.

The same idea for clearing.

Once again a simple variation of the token list functions.

These are simply copies from the token list functions.

We can now initialize the scratch variables.

Property list used by `\prop_concat:NNN`, `\prop_set_from_keyval:Nn` and others.
Combine two property lists. We cannot use a simple \texttt{tl\_concat:NNN} because there may be some duplicate keys between the two property lists.

\begin{verbatim}
\cs_new_protected:Npn \prop_concat:NNN { \__prop_concat:NNNN \prop_set_eq:NN }
\cs_generate_variant:Nn \prop_concat:NNN { ccc }
\cs_new_protected:Npn \prop_gconcat:NNN { \__prop_concat:NNNN \prop_gset_eq:NN }
\cs_generate_variant:Nn \prop_gconcat:NNN { ccc }
\cs_new_protected:Npn \__prop_concat:NNNN #1#2#3#4
\prop_set_eq:NN \l__prop_internal_prop #3
\prop_map_inline:Nn #4 { \prop_put:Nnn \l__prop_internal_prop {##1} {##2} }
#1 #2 \l__prop_internal_prop
\end{verbatim}

(End definition for \texttt{\_\_prop\_concat:NNN}, \texttt{\prop\_concat:NNN}, and \texttt{\prop\_gconcat:NNN}. These functions are documented on page 200.)

To avoid tracking throughout the loop the variable name and whether the assignment is local/global, do everything in a scratch variable and empty it afterwards to avoid wasting memory. Loop through items separated by commas, with \texttt{prg\_do\_nothing}: to avoid losing braces. After checking for termination, split the item at the first and then at the second \texttt{=} (which ought to be the first of the trailing \texttt{=} that we added). For both splits trim spaces and call a function (first \texttt{\_\_prop\_from\_keyval\_key:w} then \texttt{\_\_prop\_from\_keyval\_value:w}), followed by the trimmed material, \texttt{\_\_prop\_mark}, the subsequent part of the item, and the trailing \texttt{=}’s and \texttt{\_\_prop\_stop}. After finding the \texttt{⟨key⟩} just store it after \texttt{\_\_prop\_stop}. After finding the \texttt{⟨value⟩} ignore completely empty items (both trailing \texttt{=} were used as delimiters and all parts are empty); if the remaining part \texttt{#2} consists exactly of the second trailing \texttt{=} (namely there was exactly one \texttt{=} in the item) then output one key–value pair for the property list; otherwise complain about a missing or extra \texttt{=}.

\begin{verbatim}
\cs_new_protected:Npm \prop_set_from_keyval:Nn #1 \ propel #1
\cs_generate_variant:Nn \prop_set_from_keyval:Nn { c }
\cs_new_protected:Npn \prop_gset_from_keyval:Nn #1 \ propel #1
\cs_generate_variant:Nn \prop_gset_from_keyval:Nn { c }
\cs_new_protected:Npn \prop_const_from_keyval:Nn #1#2
\prop_set_from_keyval:Nn \l__prop_internal_prop {#2}
\tl_const:Nx #1 { \exp_not:o \l__prop_internal_prop }
\prop_clear:N \l__prop_internal_prop
\end{verbatim}

880
61.3 Accessing data in property lists

This function is used by most of the module, and hence must be fast. It receives a ⟨property list⟩, a ⟨key⟩, a ⟨true code⟩ and a ⟨false code⟩. The aim is to split the ⟨property list⟩ at the given ⟨key⟩ into the ⟨extract1⟩ before the key–value pair, the ⟨value⟩ associated with the ⟨key⟩ and the ⟨extract2⟩ after the key–value pair. This is done using a delimited function, whose definition is as follows, where the ⟨key⟩ is turned into a string.

\cs_set:Npn \__prop_split_aux:w #1 \__prop_pair:wn ⟨key⟩ \s__prop #2 \s__prop_mark #4 #5 \s__prop_stop { #4 {⟨true code⟩} {⟨false code⟩} }

If the ⟨key⟩ is present in the property list, \__prop_split_aux:w’s #1 is the part before the ⟨key⟩, #2 is the ⟨value⟩, #3 is the part after the ⟨key⟩, #4 is \use_i:nn, and #5 is additional tokens that we do not care about. The ⟨true code⟩ is left in the input stream, and can use the parameters #1, #2, #3 for the three parts of the property list as desired. Namely, the original property list is in this case #1 \__prop_pair:wn ⟨key⟩ \s__prop ⟨#2⟩ #3.

If the ⟨key⟩ is not there, then the ⟨function⟩ is \use_ii:nn, which keeps the ⟨false code⟩.

\cs_set:Npn \__prop_split_aux:w #1
Deleting from a property starts by splitting the list. If the key is present in the property list, the returned value is ignored. If the key is missing, nothing happens.

```
\cs_new_protected:Npn \prop_remove:Nn #1#2
{ \__prop_split:NnTF #1 {#2} { \tl_set:Nn #1 { ##1 ##3 } } \tl_set:Nn #3 { \q_no_value } }
```

```
\cs_generate_variant:Nn \prop_remove:Nn { NV , Nv , No } \cs_generate_variant:Nn \prop_remove:Nn { c , cV , cv , co }
```

(End definition for \prop_remove:Nn and \prop_gremove:Nn. These functions are documented on page 201.)

Getting an item from a list is very easy: after splitting, if the key is in the property list, just set the token list variable to the return value, otherwise to \q_no_value.

```
\cs_new_protected:Npn \prop_get:NnN #1#2#3
{ \__prop_split:NnTF #1 {#2} { \tl_set:Nn #3 { ##2 } } \tl_set:Nn #3 { \q_no_value } }
```

```
\cs_generate_variant:Nn \prop_get:NnN { NV , Nv , No } \cs_generate_variant:Nn \prop_get:NnN { c , cV , cv , co }
```

(End definition for \prop_get:NnN. This function is documented on page 200.)

Popping a value also starts by doing the split. If the key is present, save the value in the token list and update the property list as when deleting. If the key is missing, save \q_no_value in the token list.

```
\cs_new_protected:Npn \prop_pop:NnN #1#2#3
{ \__prop_split:NnTF #1 {#2} { \tl_set:Nn #3 { ##2 } } \tl_set:Nn #3 { \q_no_value } }
```

```
\cs_generate_variant:Nn \prop_pop:NnN { NV , Nv , No } \cs_generate_variant:Nn \prop_pop:NnN { c , cV , cv , co }
```

(End definition for \prop_pop:NnN. This function is documented on page 200.)
\prop_item:Nn Getting the value corresponding to a key in a property list in an expandable fashion simply uses \prop_map_tokens:Nn to go through the property list. The auxiliary \_prop_item:nnn receives the search string \#1, the key \#2 and the value \#3 and returns as appropriate.

\prop_map_tokens:Nn \prop_map_tokens:Nn \prop_map_tokens:Nn \prop_map_tokens:Nn

\__prop_item:nnn { \exp_after:wN \__prop_item:nnn { \tl_to_str:n {#2} } }

\prop_map_function:NN \_prop_item:nnn \prop_map_function:NN \prop_map_function:NNn

\prop_pop:NnN Popping an item from a property list, keeping track of whether the key was present or not, is implemented as a conditional. If the key was missing, neither the property list, nor the token list are altered. Otherwise, \prg_return_true: is used after the assignments.
\prg_new_protected_conditional:Npnn \prop_pop:NnN #1#2#3 { T , F , TF }
{
  \_\prop_split:NnTF #1 {#2}
  {
    \tl_set:Nn \l__prop_internal_tl {##2}
    \tl_set:Nn #1 { ##1 ##3 }
    \prg_return_true:
  }
  { \prg_return_false: }
}
\prg_new_protected_conditional:Npnn \prop_gpop:NnN #1#2#3 { T , F , TF }
{
  \_\prop_split:NnTF #1 {#2}
  {
    \tl_set:Nn \l__prop_internal_tl {##2}
    \tl_gset:Nn #1 { ##1 ##3 }
    \prg_return_true:
  }
  { \prg_return_false: }
}
\prg_generate_conditional_variant:Nnn \prop_pop:NnN { c } { T , F , TF }
\prg_generate_conditional_variant:Nnn \prop_gpop:NnN { c } { T , F , TF }

(End definition for \prop_pop:NnTF and \prop_gpop:NnTF. These functions are documented on page 202.)

\prop_put:Nnn \prop_put:NnV \prop_put:Nnx \prop_put:NVn \prop_put:NVV \prop_put:NVx \prop_put:Nvx \prop_put:Non \prop_put:Noo \prop_put:Nxx \prop_put:cnn \prop_put:cnV \prop_put:cno \prop_put:cnx \prop_put:cVn \prop_put:cVV \prop_put:cVx \prop_put:cvx \prop_put:con \prop_put:coo \prop_put:cxx \prop_gput:Nnn \prop_gput:NnV \prop_gput:Nnx \prop_gput:NVn \prop_gput:NVV \prop_gput:NVx \prop_gput:Nvx \prop_gput:Non \prop_gput:Noo \prop_gput:Nxx \prop_gput:cnn \prop_gput:cnV \prop_gput:cno \prop_gput:cnx \prop_gput:cVn \prop_gput:cVV \prop_gput:cVx \prop_gput:cvx \prop_gput:con \prop_gput:coo \prop_gput:cxx

\__prop_put:NNnn
\cs_new_protected:Npn \prop_put:Nnn { \__prop_put:NNnn \__kernel_tl_set:Nx }
\cs_new_protected:Npn \prop_gput:Nnn { \__prop_put:NNnn \__kernel_tl_gset:Nx }
\cs_new_protected:Npn \__prop_put:NNnn #1#2#3#4
{
  \tl_set:Nn \l__prop_internal_tl {
    \exp_not:N \__prop_pair:wn \tl_to_str:n {#3}
    \s__prop { \exp_not:n {#4} }
  }
  \__prop_split:NnTF #2 {#3}
  { #1 #2 { \exp_not:n {##1} \l__prop_internal_tl \exp_not:n {##3} } }
  { #1 #2 { \exp_not:o {#2} \l__prop_internal_tl } }
}
\cs_generate_variant:Nn \prop_put:Nnn { NnV , Nno , Nnx , NV , NVV , NVx , Nvx , No , Noo , Nxx }
\cs_generate_variant:Nn \prop_gput:Nnn { NnV , Nno , Nnx , NV , NVV , NVx , Nvx , No , Noo , Nxx }
\cs_generate_variant:Nn \cs_new_protected:Npn { c , cnV , cno , cnx , cV , cVV , cVx , cvx , co , coo , cxx }
\cs_generate_variant:Nn \cs_new_protected:Npn { c , cnV , cno , cnx , cV , cVV , cVx , cvx , co , coo , cxx }
\cs_generate_variant:Nn \cs_new_protected:Npn { c , cnV , cno , cnx , cV , cVV , cVx , cvx , co , coo , cxx }
\cs_generate_variant:Nn \cs_new_protected:Npn { c , cnV , cno , cnx , cV , cVV , cVx , cvx , co , coo , cxx }

Since the branches of \_\prop_split:NnTF are used as the replacement text of an internal macro, and since the ⟨key⟩ and new ⟨value⟩ may contain arbitrary tokens, it is not safe to include them in the argument of \_\prop_split:NnTF. We thus start by storing in \l__prop_internal_tl tokens which (after x-expansion) encode the key–value pair. This variable can safely be used in \_\prop_split:NnTF. If the ⟨key⟩ was absent, append the new key–value to the list. Otherwise concatenate the extracts ##1 and ##3 with the new key–value pair \l__prop_internal_tl. The updated entry is placed at the same spot as the original ⟨key⟩ in the property list, preserving the order of entries.
Adding conditionally also splits. If the key is already present, the three brace groups given by \_\_prop_split:NnTF are removed. If the key is new, then the value is added, being careful to convert the key to a string using \tl_to_str:n.

(End definition for \prop_put:Nnn and others. These functions are documented on page 199.)

61.4 Property list conditionals

Copies of the cs functions defined in l3basics.

(End definition for \prop_if_exist:NTF. This function is documented on page 201.)

Same test as for token lists.

(End definition for \prop_if_empty:NTF. This function is documented on page 201.)

Testing expandably if a key is in a property list requires to go through the key–value pairs one by one. This is rather slow, and a faster test would be
but \_prop_split:NnTF is non-expandable. Instead, we use \prop_map_tokens:Nn to compare the search key to each key in turn using \str_if_eq:ee, which is expandable.

Getting the value corresponding to a key, keeping track of whether the key was present or not, is implemented as a conditional (with side effects). If the key was absent, the token list is not altered.

\prop_get:NnNTF
\prop_get:NVNTF
\prop_get:NvNTF
\prop_get:NVNTF
\prop_get:NoNTF
\prop_get:cNNTF
\prop_get:cVNTF
\prop_get:coNTF

(End definition for \prop_if_in:NnTF and \_prop_if_in:nnn. This function is documented on page 201.)

61.5 Recovering values from property lists with branching

\prop_map_function:NN
\prop_map_function:Nc
\prop_map_function:cN
\prop_map_function:cc
\_prop_map_function:Nw

The even-numbered arguments of \_prop_map_function:Nw are keys, hence have string catcodes, except at the end where they are \fi: \prop_map_break: . The \fi: ends the

886
\iffalse: \(\text{even}\) \fi: construction and we jump out of the loop. No need for any quark test.

\begin{verbatim}
\cs_new:Npn \prop_map_function:NN #1#2
  \exp_after:wN \use_i_ii:nnn
  \exp_after:wN \__prop_map_function:Nw
  \exp_after:wN #2 #1
  \__prop_pair:wn \fi: \prop_map_break: \s__prop { }
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__prop_map_function:Nw #1
  \__prop_pair:wn #2 \s__prop #3
  \__prop_pair:wn #4 \s__prop #5
  \__prop_pair:wn #6 \s__prop #7
  \__prop_pair:wn #8 \s__prop #9
  \__prop_map_function:Nw #1
\end{verbatim}

(End definition for \prop_map_function:NN and \__prop_map_function:Nw. This function is documented on page 202.)

\begin{verbatim}
\prop_map_inline:Nn \prop_map_inline:cn
\end{verbatim}

Mapping in line requires a nesting level counter. Store the current definition of \__prop_pair:wn, and define it anew. At the end of the loop, revert to the earlier definition. Note that besides pairs of the form \__prop_pair:wn \(\langle\text{key}\rangle \\s__prop \{\langle\text{value}\rangle\}\), there are a leading and a trailing tokens, but both are equal to \scan_stop:, hence have no effect in such inline mapping. Such \scan_stop: could have affected ligatures if they appeared during the mapping.

\begin{verbatim}
\cs_new_protected:Npn \prop_map_inline:Nn #1#2
  \cs_gset_eq:cN { __prop_map_ \int_use:N \g__kernel_prg_map_int :wn } \__prop_pair:wn
  \int_g incr:N \g__kernel_prg_map_int
  \cs_gset_protected:Npn \__prop_pair:wn ##1 \s__prop ##2 {#2}
  \prg_break_point:Nn \prop_map_break: { }
  \int_gdecr:N \g__kernel_prg_map_int
  \cs_gset_eq:Nc \__prop_pair:wn { __prop_map_ \int_use:N \g__kernel_prg_map_int :wn }
\end{verbatim}

(End definition for \prop_map_inline:Nn. This function is documented on page 203.)
The mapping is very similar to \texttt{\prop_map_function:NN}. The \texttt{\use_i:nn} removes the leading \texttt{\s__prop}. The odd construction \texttt{\use:n {#1}} allows \texttt{#1} to contain any token without interfering with \texttt{\prop_map_break:}. The loop stops when the (key) between \texttt{\__prop_pair:wn} and \texttt{\s__prop} is \texttt{\fi: \prop_map_break:} instead of being a string.

\begin{verbatim}
\cs_new:Npn \prop_map_tokens:Nn #1 #2
{ \exp_last_unbraced:Nno \use_i:nn { \__prop_map_tokens:nw {#2} } #1 \__prop_pair:wn \fi: \prop_map_break: \s__prop { } \__prop_pair:wn \fi: \prop_map_break: \s__prop { } \__prop_pair:wn \fi: \prop_map_break: \s__prop { } \__prop_pair:wn \fi: \prop_map_break: \s__prop { } \prg_break_point:Nn \prop_map_break: { } }
\end{verbatim}

(\emph{End definition for \texttt{\prop_map_tokens:Nn} and \texttt{\__prop_map_tokens:nw}. This function is documented on page \pageref{prop_map_tokens}.})

\begin{verbatim}
\cs_new:Npn \prop_map_break: \prg_map_break:Nn \prop_map_break: { }
\cs_new:Npn \prop_map_break:n \prg_map_break:Nn \prop_map_break: 
\end{verbatim}

(\emph{End definition for \texttt{\prop_map_break:} and \texttt{\prop_map_break:n}. These functions are documented on page \pageref{prop_map_break}.})

\section{61.7 Viewing property lists}

Apply the general \texttt{\__kernel_chk_tl_type:NnnT}. Contrarily to sequences and comma lists, we use \texttt{msg_show_item:nn} to format both the key and the value for each pair.

\begin{verbatim}
\cs_new_protected:Npn \prop_show:N \prop_show:c \prop_log:N \prop_log:c \__prop_show:NN \__prop_show_validate:w
\cs_new_protected:Npn \prop_show:Nn \__prop_show:NN \msg_show:nnxxxx 
\cs_generate_variant:Nn \prop_show:N { c }
\cs_new_protected:Npn \prop_log:Nn \__prop_show:NN \msg_log:nnxxxx 
\cs_generate_variant:Nn \prop_log:N { c }
\cs_new_protected:Npn \prop_show:NN \__prop_show:NN \__prop_show:NN \prop_show:NN #1#2
{ \__kernel_chk_tl_type:NnnT \s__prop } \pskprop_last_ipair:wwww 
\end{verbatim}

888
\exp_after:wN \use_i:nn \exp_after:wN \__prop_show_validate:w #2
\__prop_pair:wn \q_recursion_tail \s__prop \{ \} \q_recursion_stop
\}
\{
#1 \{\ prop \} \{\ show \}
\{\ \token_to_str:N \#2 \}
\{\ \prop_map_function:NN \#2 \ms__prop \}
\{\ \{} \{\ \}
\}
\cs_new:Npn \__prop_show_validate:w \#1 \__prop_pair:wn \#2 \s__prop \#3
\{
\quark_if_recursion_tail_stop:n \#2
\exp_not:N \__prop_pair:wn \tl_to_str:n \#2 \s__prop \exp_not:n \{ \#3 \}
\}
\__prop_show_validate:w
\}
\}
\{
\}
\}
\cs_new:Npn \__prop_show_validate:w \#1 \__prop_pair:wn \#2 \s__prop \#3
\{
\quark_if_recursion_tail_stop:n \#2
\exp_not:N \__prop_pair:wn \tl_to_str:n \#2 \s__prop \exp_not:n \{ \#3 \}
\}
\__prop_show_validate:w
\}

(End definition for \prop_show:N and others. These functions are documented on page 204.)

\end{package}
Chapter 62

\texttt{l3skip implementation}

62.1 Length primitives renamed

Primitives renamed.
\begin{verbatim}
\if_dim:w \__dim_eval:w \__dim_eval_end: \cs_new_eq:NN \if_dim:w \tex_ifdim:D
\cs_new_eq:NN \__dim_eval:w \tex_dimexpr:D \cs_new_eq:NN \__dim_eval_end: \tex_relax:D
\end{verbatim}
(End definition for \texttt{\if_dim:w}, \texttt{\__dim_eval:w}, and \texttt{\__dim_eval_end:}. This function is documented on page 220.)

62.2 Internal auxiliaries

Internal scan marks.
\begin{verbatim}
\scan_new:N \s__dim_mark \scan_new:N \s__dim_stop
\end{verbatim}
(End definition for \texttt{\s__dim_mark} and \texttt{\s__dim_stop}.)

Functions to gobble up to a scan mark.
\begin{verbatim}
\cs_new:Npn \__dim_use_none_delimit_by_s_stop:w #1 \s__dim_stop { }
\end{verbatim}
(End definition for \texttt{\__dim_use_none_delimit_by_s_stop:w}.)

62.3 Creating and initialising \texttt{dim} variables

Allocating (\texttt{dim}) registers ...
\begin{verbatim}
\dim_new:N \dim_new:c
\end{verbatim}
Contrarily to integer constants, we cannot avoid using a register, even for constants. We cannot use \texttt{\textbackslash dim_gset:Nn} because debugging code would complain that the constant is not a global variable. Since \texttt{\textbackslash dim_const:Nn} does not need to be fast, use \texttt{\textbackslash dim_eval:n} to avoid needing a debugging patch that wraps the expression in checking code.

\begin{verbatim}
\cs_new_protected:Npn \dim_const:Nn #1#2
\{ \dim_new:N #1 \tex_global:D #1 = \dim_eval:n {#2} \scan_stop: \}
\end{verbatim}

Reset the register to zero. Using \texttt{\textbackslash c_zero_skip} deals with the case where the variable passed is incorrectly a skip (for example a \LaTeX{} length). Besides, these functions are then simply copied for \texttt{\textbackslash skip_zero:N} and related functions.

\begin{verbatim}
\cs_new_protected:Npn \dim_zero:N #1 { #1 = \c_zero_skip }\end{verbatim}

Create a register if needed, otherwise clear it.

\begin{verbatim}
\cs_new_protected:Npn \dim_zero_new:N #1
\{ \dim_if_exist:NTF #1 { \dim_zero:N #1 } { \dim_new:N #1 } \}
\end{verbatim}

62.4 Setting dim variables

Setting dimensions is easy enough but when debugging we want both to check that the variable is correctly local/global and to wrap the expression in some code. The \texttt{\textbackslash scan_stop:} deals with the case where the variable passed is a skip (for example a \LaTeX{} length).

\begin{verbatim}
\cs_new_protected:Npn \dim_set:Nn \dim_set:cn \dim_gset:Nn \dim_gset:cn
\{ \_\_\_\_\_dim_eval:w #1 \_\_\_\_\_dim_eval_end: \scan_stop: \}
\end{verbatim}
\dim_set_eq:NN \dim_set_eq:cN \dim_set_eq:cc
(End definition for \dim_set_eq:NN and \dim_set_eq:NN. These functions are documented on page 207.)
\dim_add:Nn \dim_add:cn \dim_gadd:Nn \dim_gadd:cn
\dim_sub:Nn \dim_sub:cn \dim_gsub:Nn \dim_gsub:cc
(End definition for \dim_add:Nn and others. These functions are documented on page 207.)
\dim_abs:n \_dim_abs:N \dim_max:nn \dim_min:nn \_dim_maxmin:wwN
Functions for min, max, and absolute value with only one evaluation. The absolute value is evaluated by removing a leading - if present.
\dim_abs:n
{ \exp_after:wN \_dim_abs:N \dim_use:N \__dim_eval:w \__dim_eval_end: \scan_stop: }
(End definition for \_dim_abs:N.)
62.5 Utilities for dimension calculations
With dimension expressions, something like \(10 \text{ pt} \times \left( \frac{5 \text{ pt}}{10 \text{ pt}} \right)\) does not work. Instead, the ratio part needs to be converted to an integer expression. Using \(\text{int\_value:w}\) forces everything into \(\text{sp}\), avoiding any decimal parts.

\(\text{End definition for } \dim\text{abs:n} \text{ and others. These functions are documented on page 207.}\)

62.6 Dimension expression conditionals

Simple comparison.

This code is adapted from the \(\text{int\_compare:nTF}\) function. First make sure that there is at least one relation operator, by evaluating a dimension expression with a trailing \(\_\dim\text{compare_error:}\). Just like for integers, the looping auxiliary \(\_\dim\text{compare:wNN}\) closes a primitive conditional and opens a new one. It is actually easier to
grab a dimension operand than an integer one, because once evaluated, dimensions all
end with pt (with category other). Thus we do not need specific auxiliaries for the three
“simple” relations <, =, and >.

\prg_new_conditional:Npnn \dim_compare:n \#1 { p , T , F , TF }
\{ \exp_after:wN \__dim_compare:w
\dim_use:N \__dim_eval:w \#1 \__dim_compare_error:
\}
\cs_new:Npn \__dim_compare:w #1 \__dim_compare_error:
\{ \exp_after:wN \if_false: \exp:w \exp_end_continue_f:w
\__dim_compare:wNN #1 ? { = \__dim_compare_end:w \else: } \s__dim_stop
\}
\exp_args:Nno \use:nn
{ \cs_new:Npn \__dim_compare:wNN #1 } { \tl_to_str:n { pt } \#2 \#3 }
\cs_new_protected:Npn \__dim_compare_error:
{ \if_int_compare:w \c_zero_int \c_zero_int \fi:
\__dim_compare_error:
}
\end {definition for \dim_compare:nTF and others. This function is documented on page 209.}

For dimension cases, the first task to fully expand the check condition. The over all idea
is then much the same as \str_case:nnTF as described in l3basics.

\cs_new:Npn \dim_case:nnTF \#1 \dim_case:nw \__dim_case:end:nw
\cs_new:Npn \dim_case:nnTF \#1 \dim_case:nw
\{ \exp:w \exp_args:Nf \__dim_case:nnTF { \dim_eval:n \#1 } \}
\cs_new:Npn \dim_case:nnT \#1 \#2 \#3
\end {definition for \dim_compare:nTF and others. This function is documented on page 209.}

894
62.7 Dimension expression loops

\dim_while_do:nn while_do and do_while functions for dimensions. Same as for the int type only the names have changed.

(End definition for \dim_case:nnTF and others. This function is documented on page 210.)
\begin{verbatim}
\cs_new:Npn \dim_do_until:nn #1#2
\{
  \dim_compare:nF {#1} {\dim_do_until:nn {#1} {#2} }
\}
\end{verbatim}

(End definition for \dim_while_do:nn and others. These functions are documented on page 211.)

\dim_while_do:nNnn \dim_until_do:nNnn \dim_do_while:nNnn \dim_do_until:nNnn
while_do and do_while functions for dimensions. Same as for the int type only the
names have changed.

\begin{verbatim}
\cs_new:Npn \dim_while_do:nNnn #1#2#3#4
\{
  \dim_compare:nNnT {#1} #2 {#3}
  \{
    #4
  \}
\}
\cs_new:Npn \dim_until_do:nNnn #1#2#3#4
\{
  \dim_compare:nNnF {#1} #2 {#3}
  \{
    #4
  \}
\}
\cs_new:Npn \dim_do_while:nNnn #1#2#3#4
\{
  #4
  \dim_compare:nNnT {#1} #2 {#3}
  \{
    \dim_do_while:nNnn {#1} #2 {#3} {#4} 
  \}
\}
\cs_new:Npn \dim_do_until:nNnn #1#2#3#4
\{
  #4
  \dim_compare:nNnF {#1} #2 {#3}
  \{
    \dim_do_until:nNnn {#1} #2 {#3} {#4} 
  \}
\}
\end{verbatim}

(End definition for \dim_while_do:nNnn and others. These functions are documented on page 211.)

62.8 Dimension step functions
Before all else, evaluate the initial value, step, and final value. Repeating a function by
steps first needs a check on the direction of the steps. After that, do the function for the
start value then step and loop around. It would be more symmetrical to test for a step
size of zero before checking the sign, but we optimize for the most frequent case (positive
step).

\begin{verbatim}
\cs_new:Npn \dim_step_function:nnnN \__dim_step:wwN \__dim_step:NnnN
\\dim_step_function:nnnN
\end{verbatim}
The approach here is to build a function, with a global integer required to make the nesting safe (as seen in other in line functions), and map that function using \texttt{\dim_step_function:nnnN}. We put a \texttt{\prg_break_point:Nn} so that \texttt{map_break} functions from other modules correctly decrement \texttt{\g__kernel_prg_map_int} before looking for their own break point. The first argument is \texttt{\scan_stop:}, so that no breaking function recognizes this break point as its own.
2004 \cs_new_protected:Npn \__dim_step:NNnnnn #1#2#3#4#5#6
2005 { #1 #2 ##1 {#6}
2006 \dim_step_function:nnn {#3} {#4} {#5} #2
2007 \prg_break_point:Nn \scan_stop: { \int_gdecr:N \g__kernel_prg_map_int }
2008 }

(End definition for \dim_step_inline:nnnn, \dim_step_variable:nnnNn, and \__dim_step:NNnnnn. These functions are documented on page 211.)

62.9 Using dim expressions and variables

\dim_eval:n Evaluating a dimension expression expandably.
2004 \cs_new:Npn \dim_eval:n #1
2005 { \dim_use:N \__dim_eval:w #1 \__dim_eval_end: }

(End definition for \dim_eval:n. This function is documented on page 212.)

\dim_sign:n See \dim_abs:n. Contrarily to \int_sign:n the case of a zero dimension cannot be distinguished from a positive dimension by looking only at the first character, since 0.2pt and 0pt start the same way. We need explicit comparisons. We start by distinguishing the most common case of a positive dimension.
2004 \cs_new:Npn \dim_sign:n #1
2005 { \int_value:w \exp_after:wN \__dim_sign:Nw \dim_use:N \__dim_eval:w #1 \__dim_eval_end: ; \exp_stop_f: }
2006 \cs_new:Npn \__dim_sign:Nw #1#2 ;
2007 { #1 > \c_zero_dim
2008 \else:
2009 \if_meaning:w - #1
2010 -1
2011 \else:
2012 0
2013 \fi:
2014 \fi: }

(End definition for \dim_sign:n and \__dim_sign:Nw. This function is documented on page 212.)

\dim_use:N Accessing a (dim). We hand-code the c variant for some speed gain.
2004 \cs_new_eq:NN \dim_use:N \tex_the:D
2005 \cs_new:Npn \dim_use:c #1 { \tex_the:D \cs:w #1 \cs_end: }

(End definition for \dim_use:N. This function is documented on page 212.)

\dim_to_decimal:n A function which comes up often enough to deserve a place in the kernel. Evaluate the dimension expression #1 then remove the trailing pt. When debugging is enabled, the argument is put in parentheses as this prevents the dimension expression from terminating early and leaving extra tokens lying around. This is used a lot by low-level manipulations.
\cs_new:Npn \dim_to_decimal:n #1 \\
{ \\
\exp_after:wN \__dim_to_decimal:w \dim_use:N \__dim_eval:w \dim_use:N \__dim_eval_end: \\
} \\
\use:x \\
{ \\
\cs_new:Npn \exp_not:N \__dim_to_decimal:w ##1 . ##2 \tl_to_str:n { pt } \\
##1 \int_compare:nNnTF {#2} > \c_zero_int \\
{ #1 . #2 } \\
{ #1 } \\
} \\
(End definition for \dim_to_decimal:n and \__dim_to_decimal:w. This function is documented on page 212.)

\dim_to_decimal_in_bp:n Conversion to big points is done using a scaling inside \__dim_eval:w as \epsilon-\TeX does that using 64-bit precision. Here, 800/803 is the integer fraction for 72/72.27. This is a common case so is hard-coded for accuracy (and speed).

\cs_new:Npn \dim_to_decimal_in_bp:n #1 \\
{ \dim_to_decimal:n { ( #1 ) * 800 / 803 } } \\
(End definition for \dim_to_decimal_in_bp:n. This function is documented on page 213.)

\dim_to_decimal_in_sp:n Another hard-coded conversion: this one is necessary to avoid things going off-scale.

\cs_new:Npn \dim_to_decimal_in_sp:n #1 \\
{ \int_value:w \__dim_eval:w \dim_use:N \__dim_eval:w \dim_use:N \__dim_eval_end: } \\
(End definition for \dim_to_decimal_in_sp:n. This function is documented on page 213.)

\dim_to_decimal_in_unit:nn An analogue of \dim_ratio:nn that produces a decimal number as its result, rather than a rational fraction for use within dimension expressions.

\cs_new:Npn \dim_to_decimal_in_unit:nn #1#2 \\
{ \dim_to_decimal:n \\
\lpt * \\
\dim_ratio:nn {#1} {#2} } \\
(End definition for \dim_to_decimal_in_unit:nn. This function is documented on page 213.)

\dim_to_fp:n Defined in l3fp-convert, documented here.

(End definition for \dim_to_fp:n. This function is documented on page 213.)
62.10 Viewing \texttt{dim} variables

\begin{verbatim}
\dim_show:N \dim_show:c
\end{verbatim}

Diagnostics.

\begin{verbatim}
\cs_new_eq:NN \dim_show:N \__kernel_register_show:N
\cs_generate_variant:Nn \dim_show:N { c }
\end{verbatim}

(End definition for \texttt{dim\_show:N}. This function is documented on page 213.)

\dim_show:n

Diagnostics. We don’t use the \TeX\ primitive \texttt{\textbackslash showthe} to show dimension expressions: this gives a more unified output.

\begin{verbatim}
\cs_new_protected:Npn \dim_show:n
\end{verbatim}

(End definition for \texttt{dim\_show:n}. This function is documented on page 214.)

\dim_log:N \dim_log:c \dim_log:n

Diagnostics. Redirect output of \texttt{dim\_show:n} to the log.

\begin{verbatim}
\cs_new_eq:NN \dim_log:N \__kernel_register_log:N
\cs_new_eq:NN \dim_log:c \__kernel_register_log:c
\cs_new_protected:Npn \dim_log:n
\end{verbatim}

(End definition for \texttt{dim\_log:N} and \texttt{dim\_log:n}. These functions are documented on page 214.)

62.11 Constant dimensions

\begin{verbatim}
\c_zero_dim \c_max_dim
\end{verbatim}

Constant dimensions.

\begin{verbatim}
\dim_const:Nn \c_zero_dim { 0 pt }
\dim_const:Nn \c_max_dim { 16383.99999 pt }
\end{verbatim}

(End definition for \texttt{c\_zero\_dim} and \texttt{c\_max\_dim}. These variables are documented on page 214.)

62.12 Scratch dimensions

\begin{verbatim}
\l_tmpa_dim \l_tmpb_dim \g_tmpa_dim \g_tmbp_dim
\end{verbatim}

We provide two local and two global scratch registers, maybe we need more or less.

\begin{verbatim}
\dim_new:N \l_tmpa_dim \dim_new:N \l_tmpb_dim \dim_new:N \g_tmpa_dim \dim_new:N \g_tmbp_dim
\end{verbatim}

(End definition for \texttt{l\_tmpa\_dim} and others. These variables are documented on page 214.)

62.13 Creating and initialising skip variables

\begin{verbatim}
\s__skip_stop
\end{verbatim}

Internal scan marks.

\begin{verbatim}
\scan_new:N \s__skip_stop
\end{verbatim}

(End definition for \texttt{s\_skip\_stop}.)
\textbf{\texttt{\textbackslash skip\_new:N}} Allocation of a new internal registers.
\begin{verbatim}
\cs_new_protected:Npn \skip_new:N #1 20114
{ \__kernel_chk_if_free_cs:N #1 20117
  \cs:w newskip \cs_end: #1 20118 }
\cs_generate_variant:Nn \skip_new:N { c }
\end{verbatim}
\textit{(End definition for \texttt{\textbackslash skip\_new:N}. This function is documented on page 214.)}

\textbf{\texttt{\textbackslash skip\_const:Nn}} Contrarily to integer constants, we cannot avoid using a register, even for constants. See \texttt{\textbackslash dim\_const:Nn} for why we cannot use \texttt{\textbackslash skip\_get:Nn}.
\begin{verbatim}
\cs_new_protected:Npn \skip_const:Nn #1#2 20120
{ \skip_new:N #1 20135
  \tex_global:D #1 = \skip_eval:n {#2} \scan_stop: 20124 }
\cs_generate_variant:Nn \skip_const:Nn { c }
\end{verbatim}
\textit{(End definition for \texttt{\textbackslash skip\_const:Nn}. This function is documented on page 215.)}

\textbf{\texttt{\textbackslash skip\_zero:N}} Reset the register to zero.
\begin{verbatim}
\cs_new_eq:NN \skip_zero:N \dim_zero:N
\cs_new_eq:NN \skip_gzero:N \dim_gzero:N
\cs_generate_variant:Nn \skip_zero:N { c }
\cs_generate_variant:Nn \skip_gzero:N { c }
\end{verbatim}
\textit{(End definition for \texttt{\textbackslash skip\_zero:N} and \texttt{\textbackslash skip\_gzero:N}. These functions are documented on page 215.)}

\textbf{\texttt{\textbackslash skip\_zero\_new:N}} Create a register if needed, otherwise clear it.
\begin{verbatim}
\cs_new_protected:Npn \skip_zero_new:N #1 20130
{ \skip_if_exist:NTF #1 { \skip_zero:N #1 } { \skip_new:N #1 } }
\cs_new_protected:Npn \skip_gzero_new:N #1 20131
{ \skip_if_exist:NTF #1 { \skip_gzero:N #1 } { \skip_new:N #1 } }
\cs_generate_variant:Nn \skip_zero_new:N { c }
\cs_generate_variant:Nn \skip_gzero_new:N { c }
\end{verbatim}
\textit{(End definition for \texttt{\textbackslash skip\_zero\_new:N} and \texttt{\textbackslash skip\_gzero\_new:N}. These functions are documented on page 215.)}

\textbf{\texttt{\textbackslash skip\_if\_exist:p:N}} Copies of the \texttt{cs} functions defined in \texttt{l3basics}.
\begin{verbatim}
\prg_new_eq_conditional:NNn \skip_if_exist:N \cs_if_exist:N
{ TF , T , F , p }
\prg_new_eq_conditional:NNn \skip_if_exist:c \cs_if_exist:c
{ TF , T , F , p }
\end{verbatim}
\textit{(End definition for \texttt{\textbackslash skip\_if\_exist:p:N}. This function is documented on page 215.)}
62.14 Setting skip variables

\skip_set:Nn
\skip_set:cn
\skip_gset:Nn
\skip_gset:cn

Much the same as for dimensions.

\cs_new_protected:Npn \skip_set:Nn #1#2
\cs_new_protected:Npn \skip_gset:Nn #1#2
\cs_generate_variant:Nn \skip_set:Nn { c }
\cs_generate_variant:Nn \skip_gset:Nn { c }

(End definition for \skip_set:Nn and \skip_gset:Nn. These functions are documented on page 215.)

\skip_set_eq:NN
\skip_set_eq:cN
\skip_set_eq:Nc
\skip_set_eq:cc
\skip_gset_eq:NN
\skip_gset_eq:cN
\skip_gset_eq:Nc
\skip_gset_eq:cc

All straightforward.

\cs_new_protected:Npn \skip_add:Nn #1#2
\cs_new_protected:Npn \skip_gadd:Nn #1#2
\cs_generate_variant:Nn \skip_add:Nn { c }
\cs_generate_variant:Nn \skip_gadd:Nn { c }
\cs_new_protected:Npn \skip_sub:Nn #1#2
\cs_new_protected:Npn \skip_gsub:Nn #1#2
\cs_generate_variant:Nn \skip_sub:Nn { c }
\cs_generate_variant:Nn \skip_gsub:Nn { c }

(End definition for \skip_add:Nn and others. These functions are documented on page 215.)

\skip_if_eq_p:nn
\skip_if_eq:nn

Comparing skips means doing two expansions to make strings, and then testing them.

As a result, only equality is tested.

\prg_new_conditional:Npnn \skip_if_eq:nn #1#2 { p , T , F , TF }
\str_if_eq:eeTF { \skip_eval:n {#1} } { \skip_eval:n {#2} }
\prg_return_true: 
\prg_return_false: 

(End definition for \skip_if_eq:nnTF. This function is documented on page 216.)

62.15 Skip expression conditionals

\skip_if_eq:nn
\skip_if_eq:nnTF

Comparing skips means doing two expansions to make strings, and then testing them.

As a result, only equality is tested.

\prg_new_conditional:Nn \skip_if_eq:nn #1#2 { p , T , F , TF }
\str_if_eq:eeTF { \skip_eval:n {#1} } { \skip_eval:n {#2} }
\prg_return_true: 
\prg_return_false: 

(End definition for \skip_if_eq:nnTF. This function is documented on page 216.)
With \textsf{\LaTeX}, we have an easy access to the order of infinities of the stretch and shrink components of a skip. However, to access both, we either need to evaluate the expression twice, or evaluate it, then call an auxiliary to extract both pieces of information from the result. Since we are going to need an auxiliary anyways, it is quicker to make it search for the string \textsf{fil} which characterizes infinite glue.

\begin{verbatim}
\cs_set_protected:Npn \__skip_tmp:w #1
{
\prg_new_conditional:Npnn \skip_if_finite:n {p,T,F,TF}
{
\exp_after:wN \__skip_if_finite:wwNw
\skip_use:N \tex_glueexpr:D ##1 ; \prg_return_false:
#1 ; \prg_return_true: \s__skip_stop
}
\cs_new:Npn \__skip_if_finite:wwNw ##1 #1 ##2 ; ##3 ##4 \s__skip_stop {##3}
}
\exp_args:No \__skip_tmp:w { \tl_to_str:n { fil } }
\end{verbatim}

(End definition for \texttt{\skip_if_finite:nTF} and \texttt{\__skip_if_finite:wwNw}. This function is documented on page 216.)

\section{Using skip expressions and variables}

\subsection{Evaluating a skip expression expandably}

\begin{verbatim}
\cs_new:Npn \skip_eval:n #1
{ \skip_use:N \tex_glueexpr:D #1 \scan_stop: }
\end{verbatim}

(End definition for \texttt{\skip_eval:n}. This function is documented on page 216.)

\subsection{Accessing a \langle skip\rangle}

\begin{verbatim}
\cs_new_eq:NN \skip_use:N \dim_use:N
\cs_new_eq:NN \skip_use:c \dim_use:c
\end{verbatim}

(End definition for \texttt{\skip_use:N}. This function is documented on page 216.)

\section{Inserting skips into the output}

\subsection{Inserting skips}

\begin{verbatim}
\cs_new_eq:NN \skip_horizontal:N \tex_hskip:D
\cs_new:Npn \skip_horizontal:n #1
{ \skip_horizontal:N \tex_glueexpr:D #1 \scan_stop: }
\cs_generate_variant:Nn \skip_horizontal:N { c }
\end{verbatim}

(End definition for \texttt{\skip_horizontal:N} and others. These functions are documented on page 217.)
62.18 Viewing skip variables

\skip_show:N \skip_show:c\nt
\skip_show:n
\skip_log:N \skip_log:c \skip_log:n

62.19 Constant skips

\c_zero_skip \c_max_skip

62.20 Scratch skips

\l_tmpa_skip \l_tmpb_skip \g_tmpa_skip \g_tmpb_skip

62.21 Creating and initialising muskip variables

\muskip_new:N \muskip_new:c

(End definition for \skip_show:n. This function is documented on page 216.)

(End definition for \skip_show:n. This function is documented on page 216.)

(End definition for \skip_log:N and \skip_log:n. These functions are documented on page 217.)

(End definition for \c_zero_skip and \c_max_skip. These functions are documented on page 217.)

(End definition for \l_tmpa_skip and others. These variables are documented on page 217.)

(End definition for \muskip_new:N. This function is documented on page 218.)
20211 \cs_new_protected:Npn \muskip_const:Nn #1#2
20212 { \muskip_new:N #1
20213 \tex_global:D #1 = \muskip_eval:n {#2} \scan_stop:
20214 }
20215 \cs_generate_variant:Nn \muskip_const:Nn { c }

(End definition for \muskip_const:Nn. This function is documented on page 218.)

20216 \cs_new_protected:Npn \muskip_zero:N #1
20217 { #1 = \c_zero_muskip }
20218 \cs_new_protected:Npn \muskip_gzero:N #1
20219 { \tex_global:D #1 = \c_zero_muskip }
20220 \cs_generate_variant:Nn \muskip_zero:N { c }
20221 \cs_generate_variant:Nn \muskip_gzero:N { c }

(End definition for \muskip_zero:N and \muskip_gzero:N. These functions are documented on page 218.)

20222 \cs_new_protected:Npn \muskip_zero_new:N #1
20223 { \muskip_if_exist:NTF #1 { \muskip_zero:N #1 } { \muskip_new:N #1 } }
20224 \cs_new_protected:Npn \muskip_gzero_new:N #1
20225 { \muskip_if_exist:NTF #1 { \muskip_gzero:N #1 } { \muskip_new:N #1 } }
20226 \cs_generate_variant:Nn \muskip_zero_new:N { c }
20227 \cs_generate_variant:Nn \muskip_gzero_new:N { c }

(End definition for \muskip_zero_new:N and \muskip_gzero_new:N. These functions are documented on page 218.)

20229 \prg_new_eq_conditional:NNn \muskip_if_exist:N \cs_if_exist:N
20230 { TF , T , F , p }
20232 \prg_new_eq_conditional:NNn \muskip_if_exist:c \cs_if_exist:c
20233 { TF , T , F , p }

(End definition for \muskip_if_exist:NTF. This function is documented on page 218.)

62.22 Setting muskip variables

This should be pretty familiar.

(End definition for \muskip_set:Nn and \muskip_gset:Nn. These functions are documented on page 219.)
All straightforward.

Using by here deals with the (incorrect) case \muskip123.

Using \muskip expressions and variables

Evaluating a muskip expression expandably.

Accessing a ⟨muskip⟩.

Viewing muskip variables

Diagnostics.

Diagnostics. We don’t use the \TeX primitive \showthe to show muskip expressions: this gives a more unified output.
Diagnostics. Redirect output of \texttt{\textbackslash \texttt{muskip\_show\_n}} to the log.

\begin{Verbatim}
\cs_new_eq:NN \texttt{\textbackslash \texttt{muskip\_log\_N}} \__kernel_register_log:N
\cs_new_eq:NN \texttt{\textbackslash \texttt{muskip\_log\_c}} \__kernel_register_log:c
\cs_new_protected:Npn \texttt{\textbackslash \texttt{muskip\_log\_n}}
  \{ \msg\_log\_eval:Nn \texttt{\textbackslash \texttt{muskip\_eval\_n}} \}
\end{Verbatim}

(End definition for \texttt{\textbackslash \texttt{muskip\_log\_N}} and \texttt{\textbackslash \texttt{muskip\_log\_n}}. These functions are documented on page 220.)

\section{Constant muskips}

\texttt{\textbackslash \texttt{c\_zero\_muskip}} \texttt{\textbackslash \texttt{c\_max\_muskip}}

\begin{Verbatim}
\muskip\const:Nn \texttt{\textbackslash \texttt{c\_zero\_muskip}} \{ \texttt{0 \textbackslash \texttt{mu}} \}
\muskip\const:Nn \texttt{\textbackslash \texttt{c\_max\_muskip}} \{ \texttt{16383.99999 \textbackslash \texttt{mu}} \}
\end{Verbatim}

(End definition for \texttt{\textbackslash \texttt{c\_zero\_muskip}} and \texttt{\textbackslash \texttt{c\_max\_muskip}}. These functions are documented on page 220.)

\section{Scratch muskips}

\texttt{\textbackslash \texttt{l\_tmpa\_muskip}} \texttt{\textbackslash \texttt{l\_tmpb\_muskip}}
\texttt{\textbackslash \texttt{g\_tmpa\_muskip}} \texttt{\textbackslash \texttt{g\_tmpb\_muskip}}

\begin{Verbatim}
\muskip\new:N \texttt{\textbackslash \texttt{l\_tmpa\_muskip}}
\muskip\new:N \texttt{\textbackslash \texttt{l\_tmpb\_muskip}}
\muskip\new:N \texttt{\textbackslash \texttt{g\_tmpa\_muskip}}
\muskip\new:N \texttt{\textbackslash \texttt{g\_tmpb\_muskip}}
\end{Verbatim}

(End definition for \texttt{\textbackslash \texttt{l\_tmpa\_muskip}} and others. These variables are documented on page 220.)
Chapter 63

l3keys Implementation

63.1 Low-level interface

The low-level key parser’s implementation is based heavily on expkv. Compared to keyval it adds a number of additional “safety” requirements and allows to process the parsed list of key–value pairs in a variety of ways. The net result is that this code needs around one and a half the amount of time as keyval to parse the same list of keys. To optimise speed as far as reasonably practical, a number of lower-level approaches are taken rather than using the higher-level expl3 interfaces.

The general behavior of the l3keys module is to throw an error on blank key names. However to support the usage of \keyval_parse:nnn in the l3prop module we allow this error to be switched off temporarily and just ignore blank names.

This temporary macro will be used since some of the definitions will need an active comma or equals sign. Inside of this macro #1 will be the active comma and #2 will be the active equals sign.

The main function starts the first of two loops. The outer loop splits the key–value list at active commas, the inner loop will do so at other commas. The use of \s__keyval_mark here prevents loss of braces from the key argument.
\cs_if_exist:NTF \tex_expanded:D 
\endinput
{ \cs_new:Npm \keyval_parse:nnn ##1 ##2 ##3 
\endinput
{ \__kernel_exp_not:w \tex_expanded:D 
\endinput
{ \__keyval_loop_active:nnw {##1} {##2} 
\s__keyval_mark ##3 \s__keyval_tail #1 
\endinput
}
}
}
{ \cs_new:Npm \keyval_parse:nnn ##1 ##2 ##3 
\group_align_safe_begin: 
\__keyval_loop_active:nnw {##1} {##2} 
\s__keyval_mark ##3 #1 \s__keyval_tail #1 
\group_align_safe_end: 
\endinput
}
\cs_new_eq:NN \keyval_parse:NNn \keyval_parse:nnn
\endinput
(End definition for \keyval_parse:nnn and \keyval_parse:NNn. These functions are documented on page 233.)

\__keyval_loop_active:nnw First a fast test for the end of the loop is done, it’ll gobble everything up to a \s__keyval_tail. The loop ending macro will gobble everything to the last comma in this definition. If the end isn’t reached yet, start the second loop splitting at other commas, the next iteration of this first loop will be inserted by the end of \__keyval_loop_other:nnw.
\endinput
{ \cs_new:Npn \__keyval_loop_active:nnw ##1 ##2 ##3 1 
\endinput
{ \__keyval_if_recursion_tail:w ##3 
\__keyval_end_loop_active:w \s__keyval_tail 
\__keyval_loop_other:nnw {##1} {##2} #3 , \s__keyval_tail , 
\endinput
}
\endinput
(End definition for \__keyval_loop_active:nnw.)

\__keyval_split_other:w \__keyval_split_active:w These two macros allow to split at the first equals sign of category 12 or 13. At the same time they also execute branching by inserting the first token following \s__keyval_mark that followed the equals sign. Hence they also test for the presence of such an equals sign simultaneously.
\endinput
{ \cs_new:Npn \__keyval_split_other:w #1 = ##2 \s__keyval_mark ##3 1 
\endinput
{ #3 ##1 \s__keyval_stop \s__keyval_mark ##2 } 
\endinput
{ \cs_new:Npn \__keyval_split_active:w #1 #2 \s__keyval_mark ##3 1 
\endinput
{ #3 ##1 \s__keyval_stop \s__keyval_mark ##2 } 
\endinput
(End definition for \__keyval_split_other:w and \__keyval_split_active:w.)

\__keyval_loop_other:nnw The second loop uses the same test for its end as the first loop, next it splits at the first active equals sign using \__keyval_split_active:w. The \s__keyval_nil prevents
accidental brace stripping and acts as a delimiter in the next steps. First testing for an
active equals sign will reduce the number of necessary expansion steps for the expected
average use case of other equals signs and hence perform better on average.

\cs_new:Npn \__keyval_loop_other:nnw ##1 ##2 ##3 ,
\__keyval_if_recursion_tail:w ##3
\__keyval_end_loop_other:w \s__keyval_tail
\__keyval_split_active:w ##3 \s__keyval_nil
\s__keyval_mark \__keyval_split_active_auxi:w
#2 \s__keyval_mark \__keyval_clean_up_active:w
{##1} {##2}
\s__keyval_mark
\s__keyval_nil
\s__keyval_mark
\s__keyval_nil
(End definition for \__keyval_loop_other:nnw.)

\__keyval_split_active:w
\__keyval_split_active_auxi:w
\__keyval_split_active_auxii:w
\__keyval_split_active_auxiii:w
\__keyval_split_active_auxiv:w
\__keyval_split_active_auxv:w
\__keyval_split_active:w the following will only be called if there was at least
one active equals sign in the current key–value pair. Therefore this is the execution
branch for a key–value pair with an active equals sign. First it tests for other equals signs in the key name, which will
eventually throw an error via \__keyval_misplaced_equal_after_active_error:w. If
none was found we forward the key to \__keyval_split_active_auxii:w.

\cs_new:Npn \__keyval_split_active_auxi:w ##1 \s__keyval_stop
\__keyval_split_other:w ##1 \s__keyval_nil
\s__keyval_mark \__keyval_misplaced_equal_after_active_error:w
\s__keyval_stop \s__keyval_mark
#2 \s__keyval_mark \__keyval_clean_up_active:w
{ \__keyval_trim:nN {##1} \__keyval_split_active_auxiii:w ##2 \s__keyval_nil }

\__keyval_split_active_auxii:w gets the correct key name with a leading \s__-
keyval_mark as ##1. It has to sanitise the remainder of the previous test and trims
the key name which will be forwarded to \__keyval_split_active_auxiii:w.

\cs_new:Npn \__keyval_split_active_auxii:w
\__keyval_split_other:w #1 \s__keyval_nil
\s__keyval_mark \__keyval_misplaced_equal_after_active_error:w
\s__keyval_stop \s__keyval_mark
\s__keyval_mark \__keyval_clean_up_active:w
{ \__keyval_trim:nN (#1) \__keyval_split_active_auxiii:w ##2 \s__keyval_nil }

Next we test for a misplaced active equals sign in the value, if none is found \__keyval-
split_active_auxiv:w will be called.

\cs_new:Npn \__keyval_split_active_auxiii:w #1 #2 \s__keyval_nil
\__keyval_split_active:w #2 \s__keyval_nil
\s__keyval_mark \__keyval_misplaced_equal_in_split_error:w
\s__keyval_split_active_auxiv:w
{#1}

This runs the last test after sanitising the remainder of the previous one. This time test
for a misplaced equals sign of category 12 in the value. Finally the last auxiliary macro
will be called.

\cs_new:Npn \__keyval_split_active_auxiv:w
\__keyval_misplaced_equal_in_split_error:w
\s__keyval_stop \s__keyval_mark
\s__keyval_stop \s__keyval_mark

910
This last macro in this execution branch sanitises the last test, trims the value and passes it to \_\_keyval_pair:nnnn.

\cs_new:Npn \_\_keyval_split_active_auxv:w ##1 \_\_keyval_nil \_\_keyval_mark \_\_keyval_misplaced_equal_in_split_error:w = \_\_keyval_mark \_\_keyval_split_active_auxv:w

(End definition for \_\_keyval_split_active_auxi:w and others.)

\_\_keyval_clean_up_active:w

The following is the branch taken if the key–value pair doesn’t contain an active equals sign. The remainder of that test will be cleaned up by \_\_keyval_clean_up_active:w which will then split at an equals sign of category other.

\cs_new:Npn \_\_keyval_clean_up_active:w
  ##1 \_\_keyval_nil \_\_keyval_mark \_\_keyval_split_active_auxi:w \_\_keyval_stop
  \{ \_\_keyval_split_other:w ##1 \_\_keyval_nil \_\_keyval_mark \_\_keyval_split_other_auxi:w \_\_keyval_misplaced_equal_in_split_error:w = \_\_keyval_mark \_\_keyval_clean_up_other:w \}

(End definition for \_\_keyval_clean_up_active:w.)

\_\_keyval_split_other_auxi:w

\_\_keyval_split_other_auxii:w

\_\_keyval_split_other_auxiii:w

This is executed if the key–value pair doesn’t contain an active equals sign but at least one other. #1 of \_\_keyval_split_other_auxi:w will contain the complete key name, which is trimmed and forwarded to the next auxiliary macro.

\cs_new:Npn \_\_keyval_split_other_auxi:w #1 \_\_keyval_stop
  \{ \_\_keyval_split_other:w #1 \_\_keyval_nil \_\_keyval_mark \_\_keyval_split_other_auxi:w \}

We know that the value doesn’t contain misplaced active equals signs but we have to test for others. Also we need to sanitise the previous test, which is done here and not earlier to avoid superfluous argument grabbing.

\cs_new:Npn \_\_keyval_split_other_auxii:w
  #1 #2 \_\_keyval_nil = \_\_keyval_mark \_\_keyval_clean_up_other:w
  \{ \_\_keyval_split_other:w #2 \_\_keyval_nil \_\_keyval_mark \_\_keyval_misplaced_equal_in_split_error:w
    = \_\_keyval_mark \_\_keyval_split_other_auxiii:w
    \{ #1 \}
  \}

\cs_new:Npn \_\_keyval_split_other_auxiii:w sanitises the test for other equals signs, trims the value and forwards it to \_\_keyval_pair:nnnn.

\cs_new:Npn \_\_keyval_split_other_auxiii:w
  #1 \_\_keyval_nil \_\_keyval_mark \_\_keyval_misplaced_equal_in_split_error:w
  \_\_keyval_stop \_\_keyval_mark
  \{ \_\_keyval_trim:nN { #1 } \_\_keyval_pair:nnnn \}

(End definition for \_\_keyval_split_other_auxi:w, \_\_keyval_split_other_auxii:w, and \_\_keyval_split_other_auxiii:w.)
is the last branch that might exist. It is called if no
equals sign was found, hence the only possibilities left are a blank list element, which is
to be skipped, or a lonely key. If it’s no empty list element this will trim the key name
and forward it to \__keyval_key:nn.

\cs_new:Npn \__keyval_clean_up_other:w
#1 \__keyval_nil \__keyval_mark \__keyval_split_other_auxi:w \__keyval_stop
\__keyval_if_blank:w #1 \__keyval_nil \__keyval_stop \__keyval_blank_true:w
\__keyval_mark \__keyval_stop
\__keyval_trim:nN { #1 } \__keyval_key:nn
\__keyval_split_active_auxi:w #2 \__keyval_mark \__keyval_clean_up_active:w
\msg_expandable_error:nnn { keyval } { misplaced-equals-sign }
\__keyval_loop_active:nnw
\__keyval_end_loop_active:w
\__keyval_end_loop_other:w
All that’s left for the parsing loops are the macros which end the recursion. Both just
gobble the remaining tokens of the respective loop including the next recursion call.
\__keyval_end_loop_other:w also has to insert the next iteration of the active loop.

\cs_new:Npn \__keyval_misplaced_equal_after_active_error:w
\__keyval_misplaced_equal_in_split_error:w
\{ \__keyval_loop_other:nnw \}
\cs_new:Npn \__keyval_misplaced_equal_after_active_error:w
\__keyval_misplaced_equal_in_split_error:w
\{ \__keyval_loop_other:nnw \}
\__keyval_end_loop_other:w
\__keyval_end_loop_active:w
All these two macros do is gobble the remainder of the current other loop execution and
throw an error. Afterwards they have to insert the next loop iteration.

(End definition for \__keyval_clean_up_other:w.)
(End definition for \__keyval_misplaced_equal_after_active_error:w and \__keyval_misplaced_equal_in_split_error:w.)

(End definition for \__keyval_clean_up_other:w.)
The parsing loops are done, so here ends the definition of \_keyval_tmp:w, which will finally set up the macros.

\_keyval_pair:nnnn
\_keyval_key:nn

These macros will be called on the parsed keys and values of the key–value list. All arguments are completely trimmed. They test for blank key names and call the functions passed to \keyval_parse:nnn inside of \exp_not:n with the correct arguments. Afterwards they insert the next iteration of the other loop.

\_keyval_if_empty:w
\_keyval_if_blank:w
\keyval_if_recursion_tail:w

All these tests work by gobbling tokens until a certain combination is met, which makes them pretty fast. The test for a blank argument should be called with an arbitrary token following the argument. Each of these utilize the fact that the argument will contain a leading \s__keyval_mark.
These macros will be called if the tests above didn’t gobble them, they execute the branching.

```latex
\cs_new:Npn \__keyval_blank_true:w \s__keyval_mark \s__keyval_stop \__keyval_trim:nN #1 \__keyval_loop_other:nnw
\cs_new:Npn \__keyval_blank_key_error:w \s__keyval_mark \s__keyval_stop #1 \__keyval_loop_other:nnw
{ \bool_if:NTF \l__kernel_keyval_allow_blank_keys_bool
  { #1 }
  { \msg_expandable_error:nn { keyval } { blank-key-name } }
\__keyval_loop_other:nnw}
```

(End definition for \__keyval_blank_true:w and \__keyval_blank_key_error:w.)

Two messages for the low level parsing system.

```latex
\msg_new:nnn { keyval } { misplaced-equals-sign }
{ Misplaced \textasciitilde '=' \textasciitilde in \texttt{key-value-input} \msg_line_context: }
\msg_new:nnn { keyval } { blank-key-name }
{ Blank \texttt{key-name-in-key-value-input} \msg_line_context: }
\prop_gput:Nnn \g_msg_module_name_prop { keyval } { LaTeX3 }
\prop_gput:Nnn \g_msg_module_type_prop { keyval } { }
```

And an adapted version of \_\_tl_trim_spaces:nn which is a bit faster for our use case, as it can strip the braces at the end. This is pretty much the same concept, so I won’t comment on it here. The speed gain by using this instead of \_\tl_trim_spaces_apply:nn is about 10\% of the total time for \_\keyval_parse:NNn with one key and one key–value pair, so I think it’s worth it.
This is the one macro which differs from the original definition.

\cs_new:Npn \__keyval_trim_auxiii:w ##1 #1 \s__keyval_nil #2
\{ #2
  #1 \s__keyval_nil
  \__keyval_trim_auxiii:w
\}

\cs_new:Npn \__keyval_trim_auxiv:w \s__keyval_mark ##1 \s__keyval_nil \__keyval_trim_auxiii:w \s__keyval_nil \__keyval_trim_auxiii:w
\{ #2 { ##2 { ##1 } }
\}
\__keyval_tmp:w { ~ }
\group_end:

(End definition for \__keyval_trim:nN and others.)

63.2 Constants and variables

\c__keys_code_root_str
\c__keys_default_root_str
\c__keys_groups_root_str
\c__keys_inherit_root_str
\c__keys_type_root_str
\c__keys_validate_root_str

Various storage areas for the different data which make up keys.

\str_const:Nn \c__keys_code_root_str { key~code~>~ }
\str_const:Nn \c__keys_default_root_str { key~default~>~ }
\str_const:Nn \c__keys_groups_root_str { key~groups~>~ }
\str_const:Nn \c__keys_inherit_root_str { key~inherit~>~ }
\str_const:Nn \c__keys_type_root_str { key~type~>~ }
\str_const:Nn \c__keys_validate_root_str { key~validate~>~ }

(End definition for \c__keys_code_root_str and others.)

\c__keys_props_root_str

The prefix for storing properties.

\str_const:Nn \c__keys_props_root_str { key~prop~>~ }

(End definition for \c__keys_props_root_str.)

\l__keys_groups_clist

Used for storing and recovering the list of groups which apply to a key: set as a comma list but at one point we have to use this for a token list recovery.

(End definition for \l__keys_groups_clist.)

\l_keys_choice_int
\l_keys_choice_tl

Publicly accessible data on which choice is being used when several are generated as a set.

(End definition for \l_keys_choice_int and \l_keys_choice_tl. These variables are documented on page 228.)

\l__keys_key_str
\l_keys_key_tl

The name of a key itself: needed when setting keys. The tl version is deprecated but has to be handled manually.

(End definition for \l__keys_key_str and \l_keys_key_tl.)
\l__keys_module_str  The module for an entire set of keys.

\l__keys_no_value_bool  A marker is needed internally to show if only a key or a key plus a value was seen: this is recorded here.

\l__keys_only_known_bool  Used to track if only “known” keys are being set.

\l__keys_path_str  \l__keys_path_tl  The “path” of the current key is stored here: this is available to the programmer and so is public. The older version is deprecated but has to be handled manually.

\l__keys_inherit_str

\l__keys_relative_tl  The relative path for passing keys back to the user. As this can be explicitly no-value, it must be a token list.

\l__keys_property_str  The “property” begin set for a key at definition time is stored here.

\l__keys_selective_bool  \l__keys_filtered_bool  Two flags for using key groups: one to indicate that “selective” setting is active, a second to specify which type (“opt-in” or “opt-out”).

\l__keys_selective_seq  The list of key groups being filtered in or out during selective setting.
\l__keys_unused_clist  Used when setting only some keys to store those left over.

\l_keys_value_tl  The value given for a key: may be empty if no value was given.

\l__keys_tmp_bool, \l__keys_tmpa_tl, \l__keys_tmpb_tl  Scratch space.

\s__keys_nil, \s__keys_mark, \s__keys_stop  Internal scan marks.

\q__keys_no_value  Internal quarks.

\__keys_quark_if_no_value:p:N  Branching quark conditional.

63.3  The key defining mechanism

The public function for definitions is just a wrapper for the lower level mechanism, more or less. The outer function is designed to keep a track of the current module, to allow safe nesting. The module is set removing any leading / (which is not needed here).
The outer functions here record whether a value was given and then converge on a common internal mechanism. There is first a search for a property in the current key name, then a check to make sure it is known before the code hands off to the next step.

```latex
\cs_new_protected:Npn \_\_keys_define:n #1
\{\bool_set_true:N \_\_keys_no_value_bool
\_\_keys_define_aux:nn {#1} \{ \}
\}

\cs_new_protected:Npn \_\_keys_define:nn #1#2
\{\bool_set_false:N \_\_keys_no_value_bool
\_\_keys_define_aux:nn {#1} {#2}
\}

\cs_new_protected:Npn \_\_keys_define_aux:nn #1#2
\{\_\_keys_property_find:n {#1}
\cs_if_exist:cTF { \c\__keys_props_root_str \l\_\_keys_property_str }
\{ \_\_keys_define_code:n {#2} \}
\}

\cs_new_protected:Npn \_\_keys_define_code:n #1
\{\str_if_empty:NF \_\_keys_property_str
\{ \msg_error:nnxx { keys } { property-unknown }
\_\_keys_property_str \_\_keys_path_str
\}
\}

(End definition for \_\_keys_define:n, \_\_keys_define:nn, and \_\_keys_define_aux:nn.)

Searching for a property means finding the last . in the input, and storing the text before and after it. Everything is turned into strings, so there is no problem using an x-type expansion. Since \_\_keys_trim_spaces:n will turn its argument into a string anyway, this function uses \cs_set_nopar:Npx instead of \tl_set:Nx to gain some speed.

```
Two possible cases. If there is a value for the key, then just use the function. If not, then a check to make sure there is no need for a value with the property. If there should be one then complain, otherwise execute it. There is no need to check for a : as if it was missing the earlier tests would have failed.

(End definition for \__keys_property_find:n and others.)

\__keys_define_code:n
\__keys_define_code:w

63.4 Turning properties into actions

Boolean keys are really just choices, but all done by hand. The second argument here is the scope: either empty or \ g for global.

(End definition for \__keys_define_code:n and \__keys_define_code:w.)
To make a choice from a key, two steps: set the code, and set the unknown key. As
multichoices and choices are essentially the same but one function, the code is
given together.

\__keys_choice_make:
\__keys_choice_make: N
\__keys_choice_make: aux: N
Auto-generating choices means setting up the root key as a choice, then defining each choice in turn.

```
\cs_new_protected:Npn \__keys_choices_make:nn { \__keys_choices_make:Nnn \__keys_choice_make: }
\cs_new_protected:Npn \__keys_multichoices_make:nn { \__keys_choices_make:Nnn \__keys_multichoice_make: }
\cs_new_protected:Npn \__keys_choices_make:Nnn #1#2#3 { #1 \int_zero:N \l_keys_choice_int \clist_map_inline:nn {#2} { \int_incr:N \l_keys_choice_int \__keys_cmd_set:nx { \l_keys_path_str / \__keys_trim_spaces:n {##1} } { \tl_set:Nn \exp_not:N \l_keys_choice_tl {##1} \int_set:Nn \exp_not:N \l_keys_choice_int { \int_use:N \l_keys_choice_int } \exp_not:n {#3} } } {#2} }
```

(End definition for \__keys_choices_make:nn, \__keys_multichoices_make:nn, and \__keys_choices_make:Nnn.)

Creating control sequences is a bit more tricky than other cases as we need to pick up the p argument. To make the internals look clearer, the trailing n argument here is just for appearance.

```
\cs_new_protected:Npn \__keys_cs_set:NNpn #1#2#3# { \cs_set_protected:cpx { \c__keys_code_root_str \l_keys_path_str / \__keys_trim_spaces:n } ##1 \exp_not:N #2 \exp_not:n {#3} {##1} } \cs_generate_variant:Nn \__keys_cs_set:NNpn { Nc }
```

(End definition for \__keys_cs_set:NNpn.)
\__keys_default_set:n Setting a default value is easy. These are stored using \texttt{\textbackslash cs\_set:cpx} as this avoids any
worries about whether a token list exists.
\begin{verbatim}
\cs_new_protected:Npn \__keys_default_set:n #1
{\tl_if_empty:nTF {#1}
{\cs_set_eq:cN { \c__keys_default_root_str \l_keys_path_str }
\tex_undefined:D
}
{\cs_set_nopar:cpx { \c__keys_default_root_str \l_keys_path_str }
{ \exp_not:n {#1} }
\__keys_value_requirement:nn { required } { false }
}
}
\end{verbatim}
(End definition for \texttt{\__keys_default_set:n}.)

\__keys_groups_set:n Assigning a key to one or more groups uses comma lists. As the list of groups only exists
if there is anything to do, the setting is done using a scratch list. For the usual grouping
reasons we use the low-level approach to undefining a list. We also use the low-level
approach for the other case to avoid tripping up the check-declarations code.
\begin{verbatim}
\cs_new_protected:Npn \__keys_groups_set:n #1
{\clist_set:Nn \l__keys_groups_clist {#1}
\clist_if_empty:NTF \l__keys_groups_clist
{\cs_set_eq:cN { \c__keys_groups_root_str \l_keys_path_str }
\tex_undefined:D
}
{\cs_set_eq:cN { \c__keys_groups_root_str \l_keys_path_str }
\l__keys_groups_clist
}
}
\end{verbatim}
(End definition for \texttt{\__keys_groups_set:n}.)

\__keys_inherit:n Inheritance means ignoring anything already said about the key: zap the lot and set up.
\begin{verbatim}
\cs_new_protected:Npn \__keys_inherit:n #1
{\__keys_undefine:
\cs_set_nopar:cpn { \c__keys_inherit_root_str \__keys_parent:o \l_keys_path_str } {#1}
}
\end{verbatim}
(End definition for \texttt{\__keys_inherit:n}.)

\__keys_initialise:n A set up for initialisation: just run the code if it exists.
\begin{verbatim}
\cs_new_protected:Npn \__keys_initialise:n #1
{\cs_if_exist:cTF
{ \c__keys_inherit_root_str \__keys_parent:o \l_keys_path_str }
\end{verbatim}
922
\__keys_execute_inherit:

```
\str_clear:N \l__keys_inherit_str
\cs_if_exist:cT { \c__keys_code_root_str \l_keys_path_str }
{ \__keys_execute:nn \l_keys_path_str {#1} }
```

(End definition for \__keys_initialise:n.)

\__keys_legacy_if_set:nn
\__keys_legacy_if_inverse:nn
\__keys_legacy_if_inverse:nnnn

Much the same as expl3 booleans, except we assume that the switch exists.

```
\cs_new_protected:Npn \__keys_legacy_if_set:nn #1#2
{ \__keys_legacy_if_set:nnnn {#1} {#2} { true } { false } }
\cs_new_protected:Npn \__keys_legacy_if_set_inverse:nn #1#2
{ \__keys_legacy_if_set:nnnn {#1} {#2} { false } { true } }
\cs_new_protected:Npn \__keys_legacy_if_set:nnnn #1#2#3#4
{ \__keys_choice_make:
  \__keys_cmd_set:nx { \l_keys_path_str / true }
  \{ \exp_not:c { legacy_if_#2 set_ #3 :n } \} { \exp_not:n {#1} } }
\__keys_cmd_set:nx { \l_keys_path_str / false }
{ \exp_not:c { legacy_if_#2 set_ #4 :n } \} { \exp_not:n {#1} } }
\__keys_cmd_set:nn { \l_keys_path_str / unknown }
{ \msg_error:nnx { keys } { boolean-values-only }
  \l_keys_key_str
  }
\__keys_default_set:n { true }
```

(End definition for \__keys_legacy_if_set:nn, \__keys_legacy_if_inverse:nn, and \__keys_legacy-if_inverse:nnnn.)

\__keys_meta_make:n
\__keys_meta_make:nn

To create a meta-key, simply set up to pass data through.

```
\cs_new_protected:Npn \__keys_meta_make:n #1
{ \__keys_cmd_set:Vo \l_keys_path_str 
  \{ \exp_after:wN \keys_set:nn \exp_after:wN { \l__keys_module_str } {#1} \} }
\cs_new_protected:Npn \__keys_meta_make:nn #1#2
{ \__keys_cmd_set:Vn \l_keys_path_str { \keys_set:nn {#1} {#2} } }
```

(End definition for \__keys_meta_make:n and \__keys_meta_make:nn.)

\__keys_prop_put:Nn
\__keys_prop_put:cn

Much the same as other variables, but needs a dedicated auxiliary.

```
\cs_new_protected:Npn \__keys_prop_put:Nn #1#2
{ \prop_if_exist:NF #1 \{ \prop_new:N #1 \}
  \exp_after:wN \keys_set:nn \exp_after:wN \{ \l__keys_module_str \} {#1}
  \exp_after:wN \keys_set:nn \exp_after:wN \{ \l_keys_path_str \} {#2} }
```

(End definition for \__keys_prop_put:Nn and \__keys_prop_put:cn.)

923
\_\_keys_undefine: Undefining a key has to be done without \_cs_undefine:c as that function acts globally.

\_\_keys_value_requirement:nn \_\_keys_validate_forbidden: \_\_keys_validate_required:

Validating key input is done using a second function which runs before the main key code. Setting that up means setting it equal to a generic stub which does the check. This approach makes the lookup very fast at the cost of one additional csname per key that needs it. The cleanup here has to know the structure of the following code.
63.5 Creating key properties

The key property functions are all wrappers for internal functions, meaning that things stay readable and can also be altered later on.

Importantly, while key properties have “normal” argument specs, the underlying code always supplies one braced argument to these. As such, argument expansion is handled by hand rather than using the standard tools. This shows up particularly for the two-argument properties, where things would otherwise go badly wrong.
One function for this.

```latex
\cs_new_protected:cpn { \c__keys_props_root_str .bool_set:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .bool_set:c } #1
\cs_new_protected:cpn { \c__keys_props_root_str .bool_gset:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .bool_gset:c } #1

(End definition for .bool_set:N and .bool_gset:N. These functions are documented on page 223.)
```

One function for this.

```latex
\cs_new_protected:cpn { \c__keys_props_root_str .bool_set_inverse:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .bool_set_inverse:c } #1
\cs_new_protected:cpn { \c__keys_props_root_str .bool_gset_inverse:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .bool_gset_inverse:c } #1

(End definition for .bool_set_inverse:N and .bool_gset_inverse:N. These functions are documented on page 223.)
```

Making a choice is handled internally, as it is also needed by .generate_choices:n.

```latex
\cs_new_protected:cpn { \c__keys_props_root_str .choice: }
\cs_new_protected:cpn { \c__keys_props_root_str .choice_make: }

(End definition for .choice:. This function is documented on page 223.)
```

For auto-generation of a series of mutually-exclusive choices. Here, #1 consists of two separate arguments, hence the slightly odd-looking implementation.

```latex
\cs_new_protected:cpn { \c__keys_props_root_str .choices:nn } #1
\cs_new_protected:cpn { \c__keys_props_root_str .choices:Vn } #1
\cs_new_protected:cpn { \c__keys_props_root_str .choices:on } #1
\cs_new_protected:cpn { \c__keys_props_root_str .choices:xn } #1

(End definition for .choices:nn. This function is documented on page 223.)
```

Creating code is simply a case of passing through to the underlying set function.

```latex
\cs_new_protected:cpn { \c__keys_props_root_str .code:n } #1

(End definition for .code:n. This function is documented on page 224.)
```
\clist_set:N
\clist_set:c
\clist_gset:N
\clist_gset:c
\clist_gset_protected:N
\clist_gset_protected:c
\cs_set:Np
\cs_set:cp
\cs_set_protected:Np
\cs_set_protected:cp
\cs_gset:Np
\cs_gset:cp
\cs_gset_protected:Np
\cs_gset_protected:cp
\default:n
\default:V
\default:o
\default:x
\dim_set:N
\dim_set:c
\dim_gset:N
\dim_gset:c
(End definition for .clist_set:N and .clist_gset:N. These functions are documented on page 223.)
\cs_set:Np
\cs_set:cp
\cs_set_protected:Np
\cs_set_protected:cp
\cs_gset:Np
\cs_gset:cp
\cs_gset_protected:Np
\cs_gset_protected:cp
(End definition for .cs_set:Np and others. These functions are documented on page 224.)
\default:n
\default:V
\default:o
\default:x
(End definition for .default:n. This function is documented on page 224.)
\dim_set:N
\dim_set:c
\dim_gset:N
\dim_gset:c
(End definition for .dim_set:N and .dim_gset:N. These functions are documented on page 224.)
Setting a variable is very easy: just pass the data along.

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .fp_set:N } #1
\{ \__keys_variable_set_required:NnnN #1 { fp } { } n \}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .fp_set:c } #1
\{ \__keys_variable_set_required:cnnN {#1} { fp } { } n \}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .fp_gset:N } #1
\{ \__keys_variable_set_required:NnnN #1 { fp } { g } n \}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .fp_gset:c } #1
\{ \__keys_variable_set_required:cnnN {#1} { fp } { g } n \}
\end{verbatim}

(End definition for \texttt{.fp_set:N} and \texttt{.fp_gset:N}. These functions are documented on page 224.)

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .groups:n } #1
\{ \__keys_groups_set:n {#1} \}
\end{verbatim}

(End definition for \texttt{.groups:n}. This function is documented on page 224.)

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .inherit:n } #1
\{ \__keys_inherit:n {#1} \}
\end{verbatim}

(End definition for \texttt{.inherit:n}. This function is documented on page 225.)

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .initial:n } #1
\{ \__keys_initialise:n {#1} \}
\cs_new_protected:cpn { \c__keys_props_root_str .initial:V } #1
\{ \exp_args:NV \__keys_initialise:n #1 \}
\cs_new_protected:cpn { \c__keys_props_root_str .initial:o } #1
\{ \exp_args:No \__keys_initialise:n {#1} \}
\cs_new_protected:cpn { \c__keys_props_root_str .initial:x } #1
\{ \exp_args:Nx \__keys_initialise:n {#1} \}
\end{verbatim}

(End definition for \texttt{.initial:n}. This function is documented on page 225.)

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .int_set:N } #1
\{ \__keys_variable_set_required:NnnN #1 { int } { } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .int_set:c } #1
\{ \__keys_variable_set_required:cnnN {#1} { int } { } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .int_gset:N } #1
\{ \__keys_variable_set_required:NnnN #1 { int } { g } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .int_gset:c } #1
\{ \__keys_variable_set_required:cnnN {#1} { int } { g } n \}
\end{verbatim}

(End definition for \texttt{.int_set:N} and \texttt{.int_gset:N}. These functions are documented on page 225.)

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .legacy_if_set:n } #1
\{ \__keys_legacy_if_set:nn {#1} { } \}
\cs_new_protected:cpn { \c__keys_props_root_str .legacy_if_gset:n } #1
\{ \__keys_legacy_if_set:nn {#1} { g } \}
\cs_new_protected:cpn { \c__keys_props_root_str .legacy_if_set_inverse:n } #1
\{ \__keys_legacy_if_set_inverse:nn {#1} { } \}
\cs_new_protected:cpn { \c__keys_props_root_str .legacy_if_gset_inverse:n } #1
\{ \__keys_legacy_if_set_inverse:nn {#1} { g } \}
\end{verbatim}

(End definition for \texttt{.legacy_if_set:n} and \texttt{.legacy_if_gset:n}. These functions are documented on page 225.)

928
Making a meta is handled internally.

```
\cs_new_protected:cpn { \c__keys_props_root_str .meta:n } #1
\{ \__keys_meta_make:n {#1} \}
```

(End definition for .meta:n. This function is documented on page 225.)

Meta with path: potentially lots of variants, but for the moment no so many defined.

```
\cs_new_protected:cpn { \c__keys_props_root_str .meta:nn } #1
\{ \__keys_meta_make:nn #1 \}
```

(End definition for .meta:nn. This function is documented on page 225.)

The same idea as .choice: and .choices:nn, but where more than one choice is allowed.

```
\cs_new_protected:cpn { \c__keys_props_root_str .multichoice: } \{
\__keys_multichoice_make: \}
\cs_new_protected:cpn { \c__keys_props_root_str .multichoices:nn } #1
\{ \__keys_multichoices_make:nn #1 \}
\cs_new_protected:cpn { \c__keys_props_root_str .multichoices:Vn } #1
\{ \exp_args:NV \__keys_multichoices_make:nn #1 \}
\cs_new_protected:cpn { \c__keys_props_root_str .multichoices:on } #1
\{ \exp_args:No \__keys_multichoices_make:nn #1 \}
\cs_new_protected:cpn { \c__keys_props_root_str .multichoices:xn } #1
\{ \exp_args:Nx \__keys_multichoices_make:nn #1 \}
```

(End definition for .multichoice: and .multichoices:nn. These functions are documented on page 225.)

Setting a variable is very easy: just pass the data along.

```
\cs_new_protected:cpn { \c__keys_props_root_str .muskip_set:N } #1
\{ \__keys_variable_set_required:NnnN #1 { muskip } { } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .muskip_set:c } #1
\{ \__keys_variable_set_required:cnnN {#1} { muskip } { } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .muskip_gset:N } #1
\{ \__keys_variable_set_required:NnnN #1 { muskip } { g } n \}
\cs_new_protected:cpn { \c__keys_props_root_str .muskip_gset:c } #1
\{ \__keys_variable_set_required:cnnN {#1} { muskip } { g } n \}
```

(End definition for .muskip_set:N and .muskip_gset:N. These functions are documented on page 226.)

Setting a variable is very easy: just pass the data along.

```
\cs_new_protected:cpn { \c__keys_props_root_str .prop_put:N } #1
\{ \__keys_prop_put:Nn #1 { } \}
\cs_new_protected:cpn { \c__keys_props_root_str .prop_put:c } #1
\{ \__keys_prop_put:cn {#1} { } \}
\cs_new_protected:cpn { \c__keys_props_root_str .prop_gput:N } #1
\{ \__keys_prop_put:Nn #1 { g } \}
\cs_new_protected:cpn { \c__keys_props_root_str .prop_gput:c } #1
\{ \__keys_prop_put:cn {#1} { g } \}
```

(End definition for .prop_put:N and .prop_gput:N. These functions are documented on page 226.)
Setting a variable is very easy: just pass the data along.

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .skip_set:N } #1
{ \__keys_variable_set_required:NnnN {#1} { skip } { } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .skip_set:c } #1
{ \__keys_variable_set_required:cnnN {#1} { skip } { } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .skip_gset:N } #1
{ \__keys_variable_set_required:NnnN {#1} { skip } { g } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .skip_gset:c } #1
{ \__keys_variable_set_required:cnnN {#1} { skip } { g } n }
\end{verbatim}

(End definition for \texttt{.skip_set:N} and \texttt{.skip_gset:N}. These functions are documented on page 226.)

Setting a variable is very easy: just pass the data along.

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .str_set:N } #1
{ \__keys_variable_set:NnnN #1 { str } { } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .str_set:c } #1
{ \__keys_variable_set:cnnN {#1} { str } { } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .str_set_x:N } #1
{ \__keys_variable_set:NnnN #1 { str } { } x }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .str_set_x:c } #1
{ \__keys_variable_set:cnnN {#1} { str } { } x }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset:N } #1
{ \__keys_variable_set:NnnN #1 { str } { g } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset:c } #1
{ \__keys_variable_set:cnnN {#1} { str } { g } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset_x:N } #1
{ \__keys_variable_set:NnnN #1 { str } { g } x }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset_x:c } #1
{ \__keys_variable_set:cnnN {#1} { str } { g } x }
\end{verbatim}

(End definition for \texttt{.str_set:N} and others. These functions are documented on page 226.)

Setting a variable is very easy: just pass the data along.

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_set:N } #1
{ \__keys_variable_set:NnnN #1 { tl } { } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_set:c } #1
{ \__keys_variable_set:cnnN {#1} { tl } { } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_set_x:N } #1
{ \__keys_variable_set:NnnN #1 { tl } { } x }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_set_x:c } #1
{ \__keys_variable_set:cnnN {#1} { tl } { } x }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_gset:N } #1
{ \__keys_variable_set:NnnN #1 { tl } { g } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_gset:c } #1
{ \__keys_variable_set:cnnN {#1} { tl } { g } n }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_gset_x:N } #1
{ \__keys_variable_set:NnnN #1 { tl } { g } x }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .tl_gset_x:c } #1
{ \__keys_variable_set:cnnN {#1} { tl } { g } x }
\end{verbatim}

(End definition for \texttt{.tl_set:N} and others. These functions are documented on page 226.)

\texttt{.undefine}: Another simple wrapper.

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .undefine: }
{ \__keys_undefine: }
\end{verbatim}

930
These are very similar, so both call the same function.

\begin{verbatim}
\cs_new_protected:cpn { \c__keys_props_root_str .value_forbidden:n } #1
{ \__keys_value_requirement:nn { forbidden } {#1} }
\cs_new_protected:cpn { \c__keys_props_root_str .value_required:n } #1
{ \__keys_value_requirement:nn { required } {#1} }
\end{verbatim}

(End definition for \texttt{.value_forbidden:n} and \texttt{.value_required:n}. These functions are documented on page \pageref{page226}.)

### 63.6 Setting keys

A simple wrapper allowing for nesting.

\begin{verbatim}
\keys_set:nn
\keys_set:nV
\keys_set:nv
\keys_set:no
\__keys_set:nn
\__keys_set:nnn
\cs_new_protected:Npn \keys_set:nn #1#2
{ \use:x
{ \bool_set_false:N \exp_not:N \l__keys_only_known_bool
\bool_set_false:N \exp_not:N \l__keys_filtered_bool
\bool_set_false:N \exp_not:N \l__keys_selective_bool
\tl_set:Nn \exp_not:N \l__keys_relative_tl
\exp_not:N \q__keys_no_value
\__keys_set:nn \exp_not:n { {#1} {#2} }
\bool_if:NT \l__keys_only_known_bool
{ \bool_set_true:N \exp_not:N \l__keys_only_known_bool }
\bool_if:NT \l__keys_filtered_bool
{ \bool_set_true:N \exp_not:N \l__keys_filtered_bool }
\bool_if:NT \l__keys_selective_bool
{ \bool_set_true:N \exp_not:N \l__keys_selective_bool }
\tl_set:Nn \exp_not:N \l__keys_relative_tl
{ \exp_not:o \l__keys_relative_tl }
}
\cs_generate_variant:Nn \keys_set:nn { nV , nv , no }
\cs_new_protected:Npn \__keys_set:nn #1#2#3
{ \str_set:Nx \l__keys_module_str { \__keys_trim_spaces:n {#2} }
\keyval_parse:NNn \__keys_set_keyval:n \__keys_set_keyval:nn {#3}
\str_set:Nn \l__keys_module_str {#1}
\str_set:Nn \l__keys_module_str {#1}
}
\end{verbatim}

(End definition for \texttt{\keys_set:nn}, \texttt{\__keys_set:nn}, and \texttt{\__keys_set:nnn}. This function is documented on page \pageref{page229}.)

Setting known keys simply means setting the appropriate flag, then running the standard code. To allow for nested setting, any existing value of \texttt{\l__keys_unused_clist} is saved on the stack and reset afterwards. Note that for speed/simplicity reasons we use a \texttt{tl} operation to set the clist here!

\begin{verbatim}
\keys_set_known:nnN
\keys_set_known:nN
\keys_set_known:nV
\keys_set_known:nvn
\keys_set_known:non
\__keys_set_known:nnnnN
\cs_new_protected:Npn \keys_set_known:nnN #1#2#3
{ }
\end{verbatim}

931
The idea of setting keys in a selective manner again uses flags wrapped around the basic code. The comments on \texttt{\keys_set_known:nnN} also apply here. We have a bit more shuffling to do to keep everything nestable.

(End definition for \texttt{\keys_set_known:nnN} and others. These functions are documented on page 230.)
\exp_args:No \_\_keys_set_filter:nnnnnN
\_\_keys_unused_clist \{#4\} \{#1\} \{#2\} \{#3\} #5
\}
\cs_generate_variant:Nn \keys_set_filter:nnnnN \{ nnV , nnv , nno \}
\cs_new_protected:Npn \_\_keys_set_filter:nnnnN #1#2#3#4#5#6
{\clist_clear:N \l__keys_unused_clist
 \_\_keys_set_filter:nnnn \{#2\} \{#3\} \{#4\} \{#5\}
 \_\_kernel_tl_set:Nx #6 \{ \exp_not:o \_\_keys_unused_clist \}
 \tl_set:Nn \l__keys_unused_clist \{#1\}
}
\cs_new_protected:Npn \keys_set_filter:nnn #1#2#3
{\__keys_set_filter:nnnn \q__keys_no_value \{#1\} \{#2\} \{#3\}}
\cs_generate_variant:Nn \keys_set_filter:nnn \{ nnV , nnv , nno \}
\cs_new_protected:Npn \__keys_set_filter:nnnn #1#2#3#4
{\use:x
 {\bool_set_false:N \exp_not:N \_\_keys_only_known_bool
 \bool_set_true:N \exp_not:N \_\_keys_filtered_bool
 \bool_set_true:N \exp_not:N \_\_keys_selective_bool
 \tl_set:Nn \exp_not:N \l__keys_relative_tl \{ \exp_not:n \{#1\} \}
 \_\_keys_set_selective:nnn \exp_not:n \{ \{#2\} \{#3\} \{#4\} \}
 \bool_if:NT \l__keys_only_known_bool
 { \bool_set_true:N \exp_not:N \_\_keys_only_known_bool }
 \bool_if:NF \l__keys_filtered_bool
 { \bool_set_false:N \exp_not:N \_\_keys_filtered_bool }
 \bool_if:NF \l__keys_selective_bool
 { \bool_set_false:N \exp_not:N \_\_keys_selective_bool }
 \tl_set:Nn \exp_not:N \l__keys_relative_tl
 { \exp_not:o \_\_keys_relative_tl }
}
\cs_new_protected:Npn \keys_set_groups:nnn #1#2#3
{\use:x
 {\bool_set_false:N \exp_not:N \_\_keys_only_known_bool
 \bool_set_true:N \exp_not:N \_\_keys_filtered_bool
 \bool_set_true:N \exp_not:N \_\_keys_selective_bool
 \tl_set:Nn \exp_not:N \l__keys_relative_tl \{ \exp_not:n \{#1\} \}
 \_\_keys_set_selective:nnn \exp_not:n \{ \{#2\} \{#3\} \}
 \bool_if:NT \l__keys_only_known_bool
 { \bool_set_true:N \exp_not:N \_\_keys_only_known_bool }
 \bool_if:NF \l__keys_filtered_bool
 { \bool_set_true:N \exp_not:N \_\_keys_filtered_bool }
 \bool_if:NF \l__keys_selective_bool
 { \bool_set_true:N \exp_not:N \_\_keys_selective_bool }
 \tl_set:Nn \exp_not:N \l__keys_relative_tl
 { \exp_not:o \_\_keys_relative_tl }
}
\cs_generate_variant:Nn \keys_set_groups:nnn \{ nnV , nnv , nno \}
A shared system once again. First, set the current path and add a default if needed. There are then checks to see if the a value is required or forbidden. If everything passes, move on to execute the code.

\begin{verbatim}
\end{verbatim}

The key path here can be fully defined, after which there is a search for the key and module names: the user may have passed them with part of what is actually the module (for our purposes) in the key name. As that happens on a per-key basis, we use the stack approach to restore the module name without a group.

This function uses \texttt{\cs_set_nopar:Npx} internally for performance reasons, the argument \#1 is already a string in every usage, so turning it into a string again seems unnecessary.
If selective setting is active, there are a number of possible sub-cases to consider. The key name may not be known at all or if it is, it may not have any groups assigned. There is then the question of whether the selection is opt-in or opt-out.

In the case where selective setting requires a comparison of the list of groups which apply to a key with the list of those which have been set active. That requires two mappings, and again a different outcome depending on whether opt-in or opt-out is set. We cannot replace the clist mapping by \clist_if_in:NnTF because catcodes may not be the same; they cannot be normalized easily in the clist because of the remote possibility that some items need braces if they involve commas or leading/trailing spaces.
\texttt{\__keys_set_keyval:n} and others.

\texttt{\__keys_value_or_default:n} \texttt{\__keys_default_inherit:} If a value is given, return it as \texttt{#1}, otherwise send a default if available.

\begin{verbatim}
\cs_new_protected:Npn \__keys_value_or_default:n #1
\bool_if:NTF \l__keys_no_value_bool
{ \bool_if:NTF \l__keys_filtered_bool
  \__keys_store_unused:
  \__keys_execute:
}{ \bool_if:NTF \l__keys_filtered_bool
  \__keys_execute:
  \__keys_store_unused:
}
\end{verbatim}

(End definition for \texttt{\__keys_set_keyval:n} and others.)

\begin{verbatim}
\__keys_value_or_default:n \__keys_default_inherit:
\end{verbatim}
Actually executing a key is done in two parts. First, look for the key itself, then look for the unknown key with the same path. If both of these fail, complain. What exactly happens if a key is unknown depends on whether unknown keys are being skipped or if an error should be raised.

To deal with the case where there is no hit, we leave \_keys\_execute\_unknown: in the input stream and clean it up using the break function: that avoids needing a boolean.

\begin{verbatim}
\cs_new_protected:Npn \_keys\_execute:
\{ \clist_map_inline:cn
\{ \_keys\_inherit\_root\_str \_keys\_parent:o \l_keys\_path\_str \}
{ \cs_if_exist:cT \{ \_keys\_code\_root\_str #1 / \l_keys\_key\_str \}
{ \_keys\_execute:no \l_keys\_path\_str \l_keys\_value\_tl
\}
{ \cs_if_exist:cT
{ \_keys\_inherit\_root\_str \_keys\_parent:o \l_keys\_path\_str }
{ \_keys\_execute\_inherit: }
{ \_keys\_execute\_unknown: }
}
\}
\}
\_keys\_execute\_inherit:
\{ \clist_map_inline:cn
\{ \_keys\_inherit\_root\_str \_keys\_parent:o \l_keys\_path\_str \}
{ \cs_if_exist:cT \{ \_keys\_code\_root\_str #1 / \l_keys\_key\_str \}
{ \_keys\_execute:no \l_keys\_value\_tl
\}
\}
\_keys\_execute\_unknown:
\{ \bool_if:NTF \_keys\_only\_known\_bool \_keys\_store\_unused: }
{ \cs_if_exist:cT
{ \_keys\_code\_root\_str \_keys\_module\_str / unknown }
{ \_keys\_execute:no \l_keys\_value\_tl
\}
\msg_error:nnx { keys } { unknown }
\end{verbatim}
A key’s code is in the control sequence with csname \c__keys_code_root_str #1. We expand it once to get the replacement text (with argument #2) and call \use:n with this replacement as its argument. This ensures that any undefined control sequence error in the key’s code will lead to an error message of the form <argument>...⟨control sequence⟩ in which one can read the (undefined) ⟨control sequence⟩ in full, rather than an error message that starts with the potentially very long key name, which would make the (undefined) ⟨control sequence⟩ be truncated or sometimes completely hidden. See https://github.com/latex3/latex2e/issues/351.

\cs_new:Npn \__keys_execute:nn #1#2
\cs_new:Npn \__keys_execute:no #1#2
{ \__keys_execute:no {#1} { \prg_do_nothing: #2 } }

When there is no relative path, things here are easy: just save the key name and value. When we are working with a relative path, first we need to turn it into a string: that can’t happen earlier as we need to store \q__keys_no_value. Then, use a standard delimited approach to fish out the partial path.

\cs_new_protected:Npn \__keys_store_unused:
{ \__keys_quark_if_no_value:NTF \l__keys_relative_tl
\clist_put_right:Nx \l__keys_unused_clist
\l__keys_key_str
\bool_if:NF \l__keys_no_value_bool
{ = \exp_not:o \l__keys_value_tl } }

\tl_if_empty:NTF \l__keys_relative_tl
\clist_put_right:Nx \l__keys_unused_clist
\l__keys_path_str
\bool_if:NF \l__keys_no_value_bool
{ = \exp_not:o \l__keys_value_tl } }

\__keys_store_unused_aux: }

\cs_new_protected:Npn \__keys_store_unused_aux:
{
Executing a choice has two parts. First, try the choice given, then if that fails call the unknown key. That always exists, as it is created when a choice is first made. So there is no need for any escape code. For multiple choices, the same code ends up used in a mapping.

```
\cs_new:Npn \__keys_choice_find:n #1
\__keys_choice_find:nn
\__keys_multichoice_find:n
```

(End definition for \_keys_execute: and others.)
\begin{verbatim}
\clist_map_function:nN {#1} \__keys_choice_find:n \\
(End definition for \__keys_choice_find:n, \__keys_choice_find:nn, and \__keys_multichoice_find:n.)
\end{verbatim}

63.7 Utilities

Used to strip off the ending part of the key path after the last `/.

\begin{verbatim}
\cs_new:Npn \__keys_parent:o #1
{ \exp_after:wN \__keys_parent:auxi:w #1 \q_nil \__keys_parent:auxii:w / \q_nil \__keys_parent:auxiv:w }
\cs_new:Npn \__keys_parent:auxi:w #1 / #2 \q_nil #3
{ #3 { #1 } #2 \q_nil #3 }
\cs_new:Npn \__keys_parent:auxii:w #1 #2 \q_nil \__keys_parent:auxii:w
{ \__keys_parent:auxi:w #2 \q_nil \__keys_parent:auxiii:n }
\cs_new:Npn \__keys_parent:auxiii:n #1
{ / \__keys_parent:auxi:w }
\cs_new:Npn \__keys_parent:auxiv:w #1 \q_nil \__keys_parent:auxiv:w
{ }
(End definition for \__keys_parent:o and others.)
\end{verbatim}

\begin{verbatim}
\cs_set:Npn \__keys_tmp:w #1
{ \cs_new:Npn \__keys_trim_spaces:n ##1
{ \exp_after:wN \__keys_trim_spaces:auxi:w \tl_to_str:n { / ##1 } / \s__keys_nil \__keys_trim_spaces:auxi:w \s__keys_mark \__keys_trim_spaces:auxii:w \s__keys_nil \__keys_trim_spaces:auxii:w / \s__keys_mark \__keys_trim_spaces:auxiii:w \s__keys_nil \__keys_trim_spaces:auxiii:w }
\__keys_tmp:w { ~ }
\__keys_tmp:w { - }
\group_end:
\cs_new:Npn \__keys_trim_spaces:auxi:w #1 - / #2 \s__keys_nil #3
{ #3 #1 / #2 \s__keys_nil #3
\end{verbatim}

Space stripping has to allow for the fact that the key here might have several parts, and spaces need to be stripped from each part. Since the key name is turned into a string groups can’t be stripped accidentally and the precautions of \tl_trim_spaces:n aren’t necessary, in this case it is much faster to just directly strip spaces around `/.
A utility for others to see if a key exists.

\keys_if_exist_p:nn
\keys_if_exist:nn
\keys_if_choice_exist_p:nn
\keys_if_choice_exist:nn
\keys_show:nn
\keys_log:nn
\__keys_show:Nnn

To show a key, show its code using a message.

\keys_if_exist:nn
\keys_if_choice_exist:nn
\keys_show:nn
\keys_log:nn
\__keys_show:Nnn

Just an alternative view on \keys_if_exist:nn.

\keys_if_choice_exist:nn
\keys_if_choice_exist:nn
\keys_show:nn
\keys_log:nn
\__keys_show:Nnn

(End definition for \__keys_trim_spaces:n and others.)

(End definition for \keys_if_exist:nnTF. This function is documented on page 292.)

(End definition for \keys_if_choice_exist:nnTF. This function is documented on page 292.)

941
\exp_args:Nc \cs_replacement_spec:N
\c__keys_code_root_str
\__keys_trim_spaces:n { #2 / #3 }
}
}
{ }
{ }
\}
(End definition for \keys_show:nn, \keys_log:nn, and \__keys_show:llnn. These functions are documented on page 232.)

63.8 Messages
For when there is a need to complain.
\msg_new:nnnn { keys } { bad-relative-key-path }
  { The-key-’#1’-is-not-inside-the-’#2’-path. }
  { The-key-’#1’-cannot-be-expressed-relative-to-path-’#2’. }
\msg_new:nnnn { keys } { boolean-values-only }
  { Key-’#1’-accepts-boolean-values-only. }
  { The-key-’#1’-only-accepts-the-values-’true’-and-’false’. }
\msg_new:nnnn { keys } { choice-unknown }
  { Key-’#1’-accepts-only-a-fixed-set-of-choices. }
  { The-key-’#1’-only-accepts-predefined-values,-
and-’#2’-is-not-one-of-these. }
\msg_new:nnnn { keys } { unknown }
  { The-key-’#1’-is-unknown-and-is-being-ignored. }
  { The-module-’#2’-does-not-have-a-key-called-’#1’.\ Check-that-you-have-spelled-the-key-name-correctly. }
\msg_new:nnnn { keys } { nested-choice-key }
  { Attempt-to-define-’#1’-as-a-nested-choice-key. }
  { The-key-’#1’-cannot-be-defined-as-a-choice-as-the-parent-key-’#2’-is-}
  { itself-a-choice. }
\msg_new:nnnn { keys } { value-forbidden }
  { The-key-’#1’-does-not-take-a-value. }
  { The-key-’#1’-should-be-given-without-a-value.\ The-value-’#2’-was-present:-the-key-will-be-ignored. }
\msg_new:nnnn { keys } { value-required }
  { The-key-’#1’-requires-a-value. }
  { The-key-’#1’-must-have-a-value.\ No-value-was-present:-the-key-will-be-ignored. }
\msg_new:nnn { keys } { show-key }
{
  The-key-#1-
  \tl_if_empty:nTF {#2}
  { is-undefined. }
  { has-the-properties: #2 . }
}
\prop_gput:Nnn \g_msg_module_name_prop { keys } { \LaTeX3 }
\prop_gput:Nnn \g_msg_module_type_prop { keys } { }
{/package}
Chapter 64

\texttt{l3intarray} implementation

There are two implementations for this module: One \texttt{fontdimen} based one for more traditional \TeX\ engines and a Lua based one for engines with Lua support.

Both versions do not allow negative array sizes.

64.1 Lua implementation

First, let’s look at the Lua variant:
We select the Lua version if the Lua helpers were defined. This can be detected by the presence of \texttt{\_\_intarray\_gset\_count:Nw}.

64.1.1 Allocating arrays

Used to differentiate intarrays in Lua and to record an invalid index.

Used as marker for intarrays in Lua. Followed by an unbraced number identifying the array and a single space. This format is used to make it easy to scan from Lua.

lua

\texttt{luacmd('\_\_intarray:w', function() scan\_int() \end{verbatim}
\texttt{\texttt{\textbackslash intarray\_new:Nn}} \hspace{1cm} \texttt{\texttt{\textbackslash intarray\_new:cn}} \hspace{1cm} \texttt{\texttt{\textbackslash _\texttt{intarray\_new:N}}}

Declare \#1 as a tokenlist with the scanmark and a unique number. Pass the array’s size to the Lua helper. Every \texttt{intarray} must be global; it’s enough to run this check in \texttt{\texttt{\textbackslash intarray\_new:Nn}}.

\begin{verbatim}
\cs_new_protected:Npn \__intarray_new:N #1
\{\__kernel_chk_if_free_cs:N #1 \int_gincr:N \g__intarray_table_int \cs_gset_nopar:Npx #1 \{ \__intarray:w \int_use:N \g__intarray_table_int \c_space_tl \}
\}
\cs_new_protected:Npn \intarray_new:Nn #1#2
\{ \__intarray_new:N #1 \__intarray_gset_count:Nw #1 \int_eval:n {#2} \scan_stop:
\int_compare:nNnT { \intarray_count:N #1 } < 0
\{ \msg_error:nnx { kernel } { negative-array-size }
\{ \intarray_count:N #1 }
\}
\cs_generate_variant:Nn \intarray_new:Nn { c }
\end{verbatim}

\texttt{\texttt{\textbackslash _\texttt{intarray\_table}}}

Internal helper to scan an intarray token, extract the associated Lua table and return an error if the input is invalid.

\begin{verbatim}
local \_intarray_table = token\_scan\_token
local put\_next = token\_put\_next
local intarray\_marker = token\_create\_safe’\_\_intarray:w’
local use\_none = token\_create\_safe’use\_none:n’
local use\_i = token\_create\_safe’use\_n’
local expand\_after\_scan\_stop = {token\_create\_safe’exp\_after:wN’,
\hspace{1cm} token\_create\_safe’scan\_stop:’}
local comma = token\_create(string\_byte’,’)
\end{verbatim}

Before we get to the first command implemented in Lua, we first need some definitions. Since \texttt{token\_create} only works correctly if \TeX{} has seen the tokens before, we first run a short \TeX{} sequence to ensure that all relevant control sequences are known.

Before we get to the first command implemented in Lua, we first need some definitions. Since \texttt{token\_create} only works correctly if \TeX{} has seen the tokens before, we first run a short \TeX{} sequence to ensure that all relevant control sequences are known.
return tables[0]
end
local i = scan_int()
local current_table = tables[i]
if current_table then return current_table end
current_table = {}
tables[i] = current_table
return current_table
end

Since in \LaTeX this is loaded in the format, we want to preserve any intarrays which are created while format building for the actual run.

To do this, we use the `register_luadata` mechanism from l3luatex: Directly before the format get dumped, the following function gets invoked and serializes all existing tables into a string. This string gets compiled and dumped into the format and is made available at the beginning of regular runs as `get_luadata'@@'`.

if register_luadata then
  register_luadata('__intarray', function()
    local t = '{[0]={},
    for i=1, #tables do
      t = string.format('%s{%s},' .. table.concat(tables[i], ','))
    end
    return t .. '}'
  end)
end

(End definition for __intarray_table.)

\intarray_count:N
\intarray_count:c
\__intarray_gset_count:Nw
Set and get the size of an array. “Setting the size” means in this context that we add zeros until we reach the desired size.

local sprint = tex.sprint
luacmd('__intarray_gset_count:Nw', function()
  local t = __intarray_table()
  local n = scan_int()
  for i=#t+1, n do t[i] = 0 end
  end, 'protected', 'global')
luacmd('intarray_count:N', function()
  sprint(-2, #__intarray_table())
end, 'global')
\langle \lua \rangle
\langle \tex \rangle
\cs_generate_variant:Nn \intarray_count:N { c }
\langle \tex \rangle

(End definition for \intarray_count:N and \__intarray_gset_count:Nw. This function is documented on page 235.)
64.1.2 Array items

\__intarray_gset:wF
\__intarray_gset:w

The setter provided by Lua. The argument order somewhat emulates the \fontdimen:
First the array index, followed by the intarray and then the new value. This has been
chosen over a more conventional order to provide a delimiter for the numbers.

\luacmd {'__intarray_gset:wF', function()}
local i = scan_int()
local t = __intarray_table()
if t[i] then
t[i] = scan_int()
put_next(use_none)
else
tex.count.l__intarray_bad_index_int = i
scan_int()
put_next(use_i)
end
\luacmd {'__intarray_gset:w', function()}
local i = scan_int()
local t = __intarray_table()
t[i] = scan_int()
end, 'protected', 'global')

\intarray_gset:Nnn
\intarray_gset:cnn
\__kernel_intarray_gset:Nnn

The \__kernel_intarray_gset:Nnn function does not use \int_eval:n, namely its
arguments must be suitable for \int_value:w. The user version checks the position and
value are within bounds.

\cs_new_protected:Npn \__kernel_intarray_gset:Nnn #1#2#3
{ \__intarray_gset:w #2 #1 #3 \scan_stop: }
\cs_new_protected:Npn \intarray_gset:Nnn #1#2#3
{ \__intarray_gset:wF \int_eval:n {#2} #1 \int_eval:n{#3}
  \msg_error:nnxxx { kernel } { out-of-bounds }
  { \token_to_str:N #1 } { \int_use:N \l__intarray_bad_index_int } \intarray_ }
\cs_generate_variant:Nn \intarray_gset:Nnn { c }
\end{tex}

(End definition for \intarray_gset:Nnn and \__kernel_intarray_gset:Nnn. This function is docu-
mented on page 235.)

\intarray_gzero:N
\intarray_gzero:c

Set the appropriate array entry to zero. No bound checking needed.

\luacmd {'intarray_gzero:N', function()}
local t = __intarray_table()
for i=1, #t do
t[i] = 0
\end{lua}

947
Get the appropriate entry and perform bound checks. The \_\_kernel_intarray\_item:Nn function omits bound checks and omits \texttt{\inteval:n}, namely its argument must be a \TeX\ integer suitable for \texttt{\intvalue:w}.

\begin{verbatim}
   luacmd('\_\_intarray_item:wF', function()
   local i = \scan_int()
   local t = \_\_intarray_table()
   local item = t[i]
   if item then
      \put_next(\use_none)
   else
      tex.l\_\_intarray_bad_index_int = i
      \put_next(\use_i)
   end
   \put_next(\expand_after_scan_stop)
   \scan_token()
   if item then
      \sprint(-2, item)
   end
end, 'global')
\end{verbatim}

\begin{verbatim}
   luacmd('\_\_intarray_item:w', function()
   local i = scan_int()
   local t = \_\_intarray_table()
   \sprint(-2, t[i])
end, 'global')
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__kernel_intarray\_item:Nn #1#2
\cs_new:Npn \intarray\_item:Nn #1#2
\begin{verbatim}
   \__intarray\_item:wF \inteval:n \{#2\} \#1
   \{ \msg_expandable_error:nnnn { kernel } { out-of-bounds } \}
   \{ \\token_to_str:N \#1 \} \{ \\int_eval:n \l__intarray\_bad\_index_int \} \{ \intarray_\}
   0
\}
\end{verbatim}
\end{verbatim}

(End definition for \texttt{\intarray\_gzero:N}. This function is documented on page 236.)

\begin{verbatim}
\intarray\_rand\_item:N
\intarray\_rand\_item:c
\begin{verbatim}
\cs_new:Npn \intarray\_rand\_item:Nn \intarray\_rand\_item:N #1
\end{verbatim}
\end{verbatim}

(End definition for \texttt{\intarray\_item:N} and others. This function is documented on page 236.)

Importantly, \texttt{\intarray\_item:Nn} only evaluates its argument once.
64.1.3 Working with contents of integer arrays

We use the \texttt{\_kernel_intarray_gset:Nnn} which does not do bounds checking and instead automatically resizes the array. This is not implemented in Lua to ensure that the clist parsing is consistent with the clist module.

\begin{verbatim}
\cs_new_protected:Npn \intarray_const_from_clist:Nn \#1\#2 {
\__intarray_new:N \#1
\int_zero:N \l__intarray_loop_int
\clist_map_inline:nn {\#2} {
\int_incr:N \l__intarray_loop_int
\__kernel_intarray_gset:Nnn \#1 \l__intarray_loop_int \{ \int_eval:n {\#1} \} }
\}
\cs_generate_variant:Nn \intarray_const_from_clist:Nn { c }
\end{verbatim}

(End definition for \texttt{\intarray_const_from_clist:Nn}. This function is documented on page 236.)

The \texttt{\_intarray_to_clist:N} auxiliary allows to choose the delimiter and is also used in \texttt{\intarray_show:N}. Here we just pass the information to Lua and let \texttt{table.concat} do the actual work. We discard the category codes of the passed delimiter but this is not an issue since the delimiter is always just a comma or a comma and a space. In both cases \texttt{sprint(2, ...)} provides the right catcodes.

\begin{verbatim}
\cs_new:Npn \intarray_to_clist:N #1 { \__intarray_to_clist:Nn #1 { , } }
\cs_generate_variant:Nn \intarray_to_clist:N { c }
\end{verbatim}

(End definition for \texttt{\intarray_const_from_clist:Nn}, \texttt{\_intarray_to_clist:N}, and \texttt{\_intarray_to_clist:w}. This function is documented on page 301.)

Loop through part of the array.

\begin{verbatim}
\cs_new:Npn \_kernel_intarray_range_to_clist:Nnn \#1\#2\#3 {
\__intarray_range_to_clist:w \#1
\int_eval:n {\#2} \int_eval:n {\#3} -
}
\end{verbatim}

\begin{verbatim}
\end{verbatim}
local from = scan_int()
local to = scan_int()

sprint(-2, concat(t, ',', from, to))
end, 'global')

(End definition for \_\_kernel_intarray_range_to_clist:Nnn and \_\_intarray_range_to_clist:w)

Loop through part of the array. We allow additional commas at the end.

\cs_new_protected:Npn \_\_kernel_intarray_gset_range_from_clist:Nnn #1#2#3
{ \_\_intarray_gset_range:w \int_eval:w #2 #1 #3 , , \scan_stop:
}

(End definition for \_\_kernel_intarray_gset_range_from_clist:Nnn and \_\_intarray_gset_range:nNw)

In order to allow some code sharing later we provide the \_\_intarray_gset_overflow_test:nw name here. It doesn’t actually test anything since the Lua implementation accepts all integers which could be tested with \tex_ifabsnum:D.

(End definition for \_\_intarray_gset_overflow_test:nw)

64.2 Font dimension based implementation

Go to the false branch of the conditional above.
### 64.2.1 Allocating arrays

We use these primitives quite a lot in this module.

```latex
\__intarray_entry:w \__intarray_count:w
\cs_new_eq:NN \__intarray_entry:w \tex_fontdimen:D
\cs_new_eq:NN \__intarray_count:w \tex_hyphenchar:D
```

(End definition for \__intarray_entry:w and \__intarray_count:w.)

\c__intarray_sp_dim

Used to convert integers to dimensions fast.

```latex
\dim_const:Nn \c__intarray_sp_dim { 1 sp }
```

(End definition for \c__intarray_sp_dim.)

\g__intarray_font_int

Used to assign one font per array.

```latex
\int_new:N \g__intarray_font_int
```

(End definition for \g__intarray_font_int.)

\intarray_new:Nn\intarray_new:cn\__intarray_new:N

Declare \#1 to be a font (arbitrarily \texttt{cmr10} at a never-used size). Store the array’s size as the \texttt{hyphenchar} of that font and make sure enough \texttt{fontdimen} are allocated, by setting the last one. Then clear any \texttt{fontdimen} that \texttt{cmr10} starts with. It seems LuaTeX’s \texttt{cmr10} has an extra \texttt{fontdimen} parameter number 8 compared to other engines (for a math font we would replace 8 by 22 or some such). Every \texttt{intarray} must be global; it’s enough to run this check in \texttt{intarray_new:Nn}.

```latex
\cs_new_protected:Npn \__intarray_new:N #1
  { \__kernel_chk_if_free_cs:N #1 \int_gincr:N \g__intarray_font_int \tex_global:D \tex_font:D #1\c__intarray_sp_dim \scan_stop: \int_step_inline:nn { 8 } { \__kernel_intarray_gset:Nnn #1 {##1} \c_zero_int } }
\cs_new_protected:Npn \intarray_new:Nn #1#2
  { \__intarray_new:N #1 \__intarray_count:w #1 = \int_eval:n {#2} \scan_stop: \int_compare:nNnT { \intarray_count:N #1 } < 0 { \msg_error:nnx { kernel } { negative-array-size } \{ \intarray_count:N #1 } > 0 { \int_compare:nNnT { \intarray_count:N #1 } > 0 } { \intarray_count:N #1 } = 0 } \intarray_new:Nn \intarray_new:Nn \intarray_new:cn \__intarray_new:N
```

(End definition for \intarray_new:Nn and \__intarray_new:N. This function is documented on page 235.)

\intarray_count:N\intarray_count:cn

Size of an array.

```latex
\cs_new:Npn \intarray_count:N \#1 \{ \int_value:w \__intarray_count:w \#1 \}
\cs_generate_variant:Nn \intarray_count:N { c }
```

(End definition for \intarray_count:N. This function is documented on page 235.)
64.2.2 Array items

\_\_intarray\_signed\_max\_dim:n

Used when an item to be stored is larger than \c\_max\_dim in absolute value; it is replaced by ±\c\_max\_dim.
\begin{verbatim}
c_new:Npn \_\_intarray\_signed\_max\_dim:n #1
{ \int_value:w \int_compare:nNnT {#1} < 0 { - } \c\_max\_dim }
\end{verbatim}

(End definition for \_\_intarray\_signed\_max\_dim:n.)

\_\_intarray\_bounds:NNnTF
\_\_intarray\_bounds\_error:NNnw

The functions \texttt{\textbackslash intarray\_gset:Nnn} and \texttt{\textbackslash intarray\_item:Nn} share bounds checking. The T branch is used if #3 is within bounds of the array #2.
\begin{verbatim}
c_new:Npn \_\_intarray\_bounds:NNnTF #1#2#3
{ \if_int_compare:w 1 > #3 \exp_stop_f:
\_\_intarray\_bounds\_error:NNnw #1 #2 {#3}
\else:
\_\_intarray\_bounds\_error:NNnw #1 #2 \_\_intarray\_count:N #2
\fi:
\use_i:nn }
c_new:Npn \_\_intarray\_bounds\_error:NNnw #1#2#3#4 \use_i:nn #5#6
{ #4
#1 \{ kernel \} \{ out-of-bounds \}
{ \token_to_str:N #2 } \{ \_\_intarray\_count:N #2 \}
#6
}
\end{verbatim}

(End definition for \_\_intarray\_bounds:NNnTF and \_\_intarray\_bounds\_error:NNnw.)

\texttt{\textbackslash intarray\_gset:Nnn}
\texttt{\textbackslash kernel\_intarray\_gset:Nnn}
\texttt{\textbackslash intarray\_gset:Nnn}
\texttt{\textbackslash intarray\_gset\_overflow:Nnn}

Set the appropriate \texttt{\textbackslash fontdimen}. The \_\_kernel\_\_intarray\_gset:Nnn function does not use \texttt{\textbackslash int\_eval:n}, namely its arguments must be suitable for \texttt{\textbackslash int\_value:w}. The user version checks the position and value are within bounds.
\begin{verbatim}
c_new_protected:Npn \_\_kernel\_intarray\_gset:Nnn #1#2#3
{ \_\_intarray\_entry:w #2 \_\_intarray\_sp\_dim \_\_intarray\_gset\_overflow:list }\_\_intarray\_gset:Nnn #1#2#3
\cs_generate_variant:Nn \intarray\_gset:Nnn { c }
c_new_protected:Npn \_\_kernel\_intarray\_gset\_overflow:Nnn { c }
{ \_\_intarray\_bounds:NNnTF \_\_kernel\_intarray\_gset\_overflow\_test:nw {#3} \_\_kernel\_intarray\_gset:Nnn #1 \_\_intarray\_overflows:NNn #3 \_\_kernel\_intarray\_gset\_overflow\_test:nw {#3} \_\_kernel\_intarray\_gset\_overflow\_test:nw {#3} }
\end{verbatim}

952
\cs_if_exist:NTF \tex_ifabsnum:D
{
  \cs_new_protected:Npn \__intarray_gset_overflow_test:nw #1
  {
  \tex_ifabsnum:D #1 > \c_max_dim
  \exp_after:wN \__intarray_gset_overflow:NNnn
  \fi:
  }
}

\cs_new_protected:Npn \__intarray_gset_overflow_test:nw #1
{
  \if_int_compare:w \int_abs:n {#1} > \c_max_dim
  \exp_after:wN \__intarray_gset_overflow:NNnn
  \fi:
}

\cs_new_protected:Npn \__intarray_gset_overflow:NNnn #1#2#3#4
{
  \msg_error:nnxxxx { kernel } { overflow }
  { \token_to_str:N #2 } {#3} {#4} { \__intarray_signed_max_dim:n {#4} }
#1 #2 {#3} { \__intarray_signed_max_dim:n {#4} }
}

(End definition for \intarray_gset:Nnn and others. This function is documented on page 235.)

\begin{verbatim}
\intarray_gzero:N
\intarray_gzero:c
\end{verbatim}

Set the appropriate \fontdimen to zero. No bound checking needed. The \prg_repli\textunderscore cates:nn possibly uses quite a lot of memory, but this is somewhat comparable to the size of the array, and it is much faster than an \int_step_inline:nn loop.

\begin{verbatim}
\intarray_item:Nn
\intarray_item:cn
\__kernel_intarray_item:Nn
\intarray_item:Nn
\end{verbatim}

Get the appropriate \fontdimen and perform bound checks. The \__kernel_intarray_item:Nn function omits bound checks and omits \int_eval:n, namely its argument must be a \TeX integer suitable for \int_value:w.

(End definition for \intarray_gzero:N. This function is documented on page 236.)

\begin{verbatim}
\intarray_item:Nn
\intarray_item:cn
\__kernel_intarray_item:Nn
\intarray_item:Nn
\end{verbatim}
\cs_new:Npn \__intarray_item:Nw #1 #2 ;
{ \__intarray_bounds:NNnTF \msg_expandable_error:nnfff #1 {#2}
{ \__kernel_intarray_item:Nn #1 {#2} } 0 }

(End definition for \intarray_item:Nn, \__kernel_intarray_item:Nn, and \__intarray_item:Nn. This function is documented on page 236.)

\intarray_rand_item:N
\intarray_rand_item:c

Importantly, \intarray_item:Nn only evaluates its argument once.
\cs_new:Npn \intarray_rand_item:N #1
{ \intarray_item:Nn #1 { \int_rand:n { \intarray_count:N #1 } } }
\cs_generate_variant:Nn \intarray_rand_item:N { c }

(End definition for \intarray_rand_item:N. This function is documented on page 236.)

64.2.3 Working with contents of integer arrays

Similar to \intarray_new:Nn (which we don’t use because when debugging is enabled that function checks the variable name starts with g_). We make use of the fact that \LaTeX allows allocation of successive \fontdimen as long as no other font has been declared: no need to count the comma list items first. We need the code in \intarray_gset:Nnn that checks the item value is not too big, namely \__intarray_gset_overflow_test:nw, but not the code that checks bounds. At the end, set the size of the intarray.

\cs_new_protected:Npn \intarray_const_from_clist:Nn #1 #2
{ \__intarray_new:N #1 \int_zero:N \l__intarray_loop_int
\clist_map_inline:nn {#2} { \exp_args:Nf \__intarray_const_from_clist:nN {\int_eval:n {##1}} #1 } \__intarray_count:w #1 \l__intarray_loop_int }
\cs_generate_variant:Nn \intarray_const_from_clist:Nn { c }
\cs_new_protected:Npn \__intarray_const_from_clist:nN #1 #2
{ \int_incr:N \l__intarray_loop_int \__intarray_gset_overflow_test:nw {#1} \__kernel_intarray_gset:Nnn #2 \l__intarray_loop_int {#1} }

(End definition for \intarray_const_from_clist:Nn and \__intarray_const_from_clist:nN. This function is documented on page 236.)

\intarray_to_clist:N
\intarray_to_clist:c
\__intarray_to_clist:N
\__intarray_to_clist:w

Loop through the array, putting a comma before each item. Remove the leading comma with f-expansion. We also use the auxiliary in \intarray_show:N with argument comma, space.
\cs_new:Npn \intarray_to_clist:N #1 \{ \__intarray_to_clist:Nn #1 \{ , \} \}
\cs_generate_variant:Nn \intarray_to_clist:N { c }
\cs_new:Npn \__intarray_to_clist:Nn #1 #2
{ \int_compare:nNnF \{ \intarray_count:N #1 \} = \c_zero int 
\exp_last_unbraced:Nf \use_none:n

954
\section*{\texttt{\_\_intarray\_to\_clist}}\texttt{:w}

\begin{verbatim}
\cs_new:Npn \\_\_intarray\_to\_clist:w #1 \#2 \prg_break_point: \\
\}
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \\_\_intarray\_range\_to\_clist:Nnn #1#2#3 \\
{ \exp_last_unbraced:Nf \use_none:n \\
\exp_after:wN \\_\_intarray\_range\_to\_clist:ww \\
\int_value:w \int_eval:w #2 \exp_after:wN ; \\
\int_value:w \int_eval:w #3 ; \\
\#1 \prg_break_point:
}
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \\_\_intarray\_range\_to\_clist:ww #1 ; #2 ; #3 \\
{ \if_int_compare:w #1 > #2 \exp_stop_f: \\
\prg_break:n \\
\fi: \\
, \\_\_kernel\_intarray\_item:Nn #3 \{\#1\}
\exp_after:wN \\_\_intarray\_range\_to\_clist:ww \\
\int_value:w \int_eval:w \#1 + \c_one_int ; #2 \#3
}
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \\_\_kernel\_intarray\_gset\_range_from\_clist:Nnn #1#2#3 \\
{ \int_set:Nn \l__intarray\_loop\_int {#2} \\
\\_\_intarray\_gset\_range:Nw \#1 \#3 , , \prg_break_point:
}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \\_\_intarray\_gset\_range:Nw #1 #2 , \\
{ \if_catcode:w \scan_stop: \tl_to_str:n \#2 \scan_stop: \\
\prg_break:n \\
\fi: \\
\\_\_kernel\_intarray\_gset:Nn \#1 \l__intarray\_loop\_int \{\#2\} \\
\int_incr:N \l__intarray\_loop\_int \\
\\_\_intarray\_gset\_range:Nw \#1
}
\end{verbatim}

\section*{\texttt{\_\_intarray\_range\_to\_clist:w}}

Loop through part of the array.

\begin{verbatim}
\cs_new:Npn \\_\_kernel\_intarray\_range\_to\_clist:ww #1 \#2 \#3 \\
{ \exp_last_unbraced:Nf \use_none:n \\
\exp_after:wN \\_\_intarray\_range\_to\_clist:w \\
\int_value:w \int_eval:w \#2 \exp_after:wN ; \\
\int_value:w \int_eval:w \#3 ; \\
\#1 \prg_break_point:
}
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \\_\_kernel\_intarray\_range\_to\_clist:Nnn #1#2#3 \\
{ \exp_last_unbraced:Nf \use_none:n \\
\exp_after:wN \\_\_intarray\_range\_to\_clist:w \\
\int_value:w \int_eval:w #2 \exp_after:wN ; \\
\int_value:w \int_eval:w #3 ; \\
\#1 \prg_break_point:
}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \\_\_kernel\_intarray\_gset\_range:ww #1 #2 , \\
{ \if_catcode:w \scan_stop: \tl_to_str:n \#2 \scan_stop: \\
\prg_break:n \\
\fi: \\
\\_\_kernel\_intarray\_gset:Nn \#1 \l__intarray\_loop\_int \{\#2\} \\
\int_incr:N \l__intarray\_loop\_int \\
\\_\_intarray\_gset\_range:Nw \#1
}
\end{verbatim}

(End definition for \texttt{\_\_intarray\_to\_clist:N}, \texttt{\_\_intarray\_to\_clist:Hn}, and \texttt{\_\_intarray\_to\_clist:w}.
\textit{This function is documented on page 301.})

\section*{\texttt{\_\_kernel\_intarray\_range\_to\_clist:Nn}}

\section*{\texttt{\_\_intarray\_range\_to\_clist:ww}}

\section*{\texttt{\_\_kernel\_intarray\_gset\_range:ww}}

(End definition for \texttt{\_\_kernel\_intarray\_range\_to\_clist:NNn} and \texttt{\_\_kernel\_intarray\_range\_to\_clist:ww}.)

Loop through part of the array.

\begin{verbatim}
\cs_new_protected:Npn \\_\_kernel\_intarray\_gset\_range_from\_clist:Nnn #1#2#3 \\
{ \int_set:Nn \l__intarray\_loop\_int \{#2\} \\
\\_\_intarray\_gset\_range:Nw \#1 \#3 , , \prg_break_point:
}
\end{verbatim}

\section*{\texttt{\_\_intarray\_gset\_range:Nw}}

(End definition for \texttt{\_\_kernel\_intarray\_gset\_range:NNn} and \texttt{\_\_kernel\_intarray\_gset\_range:ww}.)

Loop through part of the array.

(End definition for \texttt{\_\_kernel\_intarray\_range\_to\_clist:Nn} and \texttt{\_\_kernel\_intarray\_range\_to\_clist:ww}.)

955
End definition for \_\_kernel_intarray_gset_range_from_clist:Nnn and \_\_intarray_gset_range:Nw.

64.3 Common parts

<table>
<thead>
<tr>
<th>\intarray_show:N</th>
<th>Convert the list to a comma list (with spaces after each comma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\intarray_show:c</td>
<td>\cs_new_protected:Npn \intarray_show:N { __intarray_show:NN \msg_show:nnxxxx }</td>
</tr>
<tr>
<td>\intarray_log:N</td>
<td>\cs_generate_variant:Nn \intarray_show:N { c }</td>
</tr>
<tr>
<td>\intarray_log:c</td>
<td>\cs_new_protected:Npn \intarray_log:N { __intarray_show:NN \msg_log:nnxxxx }</td>
</tr>
<tr>
<td></td>
<td>\cs_generate_variant:Nn \intarray_log:N { c }</td>
</tr>
<tr>
<td></td>
<td>\cs_new_protected:Npn __intarray_show:NN #1#2 { __kernel_chkDefined:NT #2 { \token_toStr:N #2 } { \intarray_count:N #2 } { __intarray_to_clist:Nn #2 { , ~ } } }</td>
</tr>
<tr>
<td></td>
<td>(End definition for \intarray_show:N and \intarray_log:N. These functions are documented on page 236.)</td>
</tr>
</tbody>
</table>

64.3.1 Random arrays

| \intarray_gset_rand:Nn | We only perform the bounds checks once. This is done by two \_\_\_intarray_gset_-overflow_test:nw, with an appropriate empty argument to avoid a spurious “at position #1” part in the error message. Then calculate the number of choices: this is at most $2^{30} - 1 - (-(2^{30} - 1)) + 1 = 2^{31} - 1$, which just barely does not overflow. For small ranges use \_\_\_kernel_randint:n (making sure to subtract 1 before adding the random number to the $\langle min \rangle$, to avoid overflow when $\langle min \rangle$ or $\langle max \rangle$ are $\pm \cmax_int$), otherwise \_\_\_kernel_randint:nn. Finally, if there are no random numbers do not define any of the auxiliaries. |
| \intarray_gset_rand:cn | \cs_new_protected:Npn \intarray_gset_rand:Nn #1 { \intarray_gset_rand:Nnn #1 { 1 } } |
| \intarray_gset_rand:Nnn | \cs_generate_variant:Nn \intarray_gset_rand:Nn { c } |
| \intarray_gset_rand:nn | \sys_ifRandExist:TF |
| \intarray_gset_rand:NNn | \cs_new_protected:Npn \intarray_gset_rand:NNn #1\#2\#3 { \__intarray_gset_rand:Nnn \#1 \#2 \int_compare:nNnTF {\#2} {\#3} { \int_eval:n {\#2} } { \int_eval:n {\#3} } } |
| \intarray_gset_rand:Nff | \cs_new_protected:Npn \intarray_gset_rand:Nnn \#1\#2\#3 { \msg_expandable_error:nnnn { kernel } { randint-backward-range } {\#2} {\#3} } |
| \intarray_gset_all_same:Nn | (End definition for \_\_\_intarray_gset_rand:Nn and \_\_\_intarray_gset_range:Nw, \_\_\_intarray_gset_range_from_clist:Nnn and \_\_\_intarray_gset_range:Nw.) |
(End definition for \texttt{\intarray_gset_rand:Nn} and others. These functions are documented on page 301.)
Chapter 65

l3fp implementation

Nothing to see here: everything is in the subfiles!
Chapter 66

\texttt{l3fp-aux} implementation

66.1 Access to primitives

Largely for performance reasons, we need to directly access primitives rather than use \texttt{\int_eval:n}. This happens \textit{a lot}, so we use private names. The same is true for \texttt{\romannumeral}, although it is used much less widely.

\begin{verbatim}
\cs_new_eq:NN \__fp_int_eval:w \tex_numexpr:D
\cs_new_eq:NN \__fp_int_eval_end: \scan_stop:
\cs_new_eq:NN \__fp_int_to_roman:w \tex_romannumeral:D
\end{verbatim}

(End definition for \__fp_int_eval:w, \__fp_int_eval_end:, and \__fp_int_to_roman:w.)

66.2 Internal representation

Internally, a floating point number \(x\) is a token list containing

\begin{verbatim}
\s__fp \__fp_chk:w (case) (sign) (body)
\end{verbatim}

Let us explain each piece separately.

Internal floating point numbers are used in expressions, and in this context are subject to \texttt{f-expansion}. They must leave a recognizable mark after \texttt{f-expansion}, to prevent the floating point number from being re-parsed. Thus, \texttt{\s__fp} is simply another name for \texttt{\relax}.

When used directly without an accessor function, floating points should produce an error: this is the role of \texttt{\__fp_chk:w}. We could make floating point variables be protected to prevent them from expanding under \texttt{x-expansion}, but it seems more convenient to treat them as a subcase of token list variables.

The (decimal part of the) IEEE-754-2008 standard requires the format to be able to represent special floating point numbers besides the usual positive and negative cases. We distinguish the various possibilities by their \textit{(case)}, which is a single digit:

0 zeros: \(+0\) and \(-0\),

1 “normal” numbers (positive and negative),
Table 3: Internal representation of floating point numbers.

<table>
<thead>
<tr>
<th>Representation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 \s__fp_... ;</td>
<td>Positive zero.</td>
</tr>
<tr>
<td>0 2 \s__fp_... ;</td>
<td>Negative zero.</td>
</tr>
<tr>
<td>1 0 ⟨exponent⟩ {⟨X₁⟩} {⟨X₂⟩} {⟨X₃⟩} {⟨X₄⟩} ;</td>
<td>Positive floating point.</td>
</tr>
<tr>
<td>1 2 ⟨exponent⟩ {⟨X₁⟩} {⟨X₂⟩} {⟨X₃⟩} {⟨X₄⟩} ;</td>
<td>Negative floating point.</td>
</tr>
<tr>
<td>2 0 \s__fp_... ;</td>
<td>Positive infinity.</td>
</tr>
<tr>
<td>2 2 \s__fp_... ;</td>
<td>Negative infinity.</td>
</tr>
<tr>
<td>3 1 \s__fp_... ;</td>
<td>Quiet \text{n}an.</td>
</tr>
<tr>
<td>3 1 \s__fp_... ;</td>
<td>Signalling \text{n}an.</td>
</tr>
</tbody>
</table>

2 infinities: \(+\text{inf}\) and \(-\text{inf}\),

3 quiet and signalling \text{n}an.

The \langle\text{sign}\rangle is 0 (positive) or 2 (negative), except in the case of \text{n}an, which have \langle\text{sign}\rangle = 1. This ensures that changing the \langle\text{sign}\rangle digit to \(2 – \langle\text{sign}\rangle\) is exactly equivalent to changing the sign of the number.

Special floating point numbers have the form
\s__fp \s__fp_chk:w \langle\text{case}\rangle \langle\text{sign}\rangle \s__fp_... ;
where \s__fp_... is a scan mark carrying information about how the number was formed (useful for debugging).

Normal floating point numbers ((\text{case} = 1)) have the form
\s__fp \s__fp_chk:w 1 \langle\text{sign}\rangle {⟨exponent⟩} {⟨X₁⟩} {⟨X₂⟩} {⟨X₃⟩} {⟨X₄⟩} ;
Here, the \langle\text{exponent}\rangle is an integer, between \(-10000\) and \(10000\). The body consists in four blocks of exactly 4 digits, \(0000 \leq ⟨Xᵢ⟩ \leq 9999\), and the floating point is
\((-1)^{\langle\text{sign}\rangle/2} ⟨X₁⟩⟨X₂⟩⟨X₃⟩⟨X₄⟩ \cdot 10^{⟨\text{exponent}⟩−16}\)
where we have concatenated the 16 digits. Currently, floating point numbers are normalized such that the \langle\text{exponent}\rangle is minimal, in other words, \(1000 \leq ⟨X₁⟩ \leq 9999\).
Calculations are done in base \(10000\), i.e. one myriad.

### 66.3 Using arguments and semicolons

\_\_fp\_use\_none\_stop\_f:n This function removes an argument (typically a digit) and replaces it by \exp_stop_f:, a marker which stops f-type expansion.
\_\_fp\_use\_s:n \_\_fp\_use\_s:nn Those functions place a semicolon after one or two arguments (typically digits).
Those functions select specific arguments among a set of arguments delimited by a semicolon.

\cs_new:Npn \__fp_use_none_until_s:w #1; { }  
\cs_new:Npn \__fp_use_i_until_s:nw #1#2; {#1}  
\cs_new:Npn \__fp_use_ii_until_s:nnw #1#2#3; {#2}

(End definition for \__fp_use_none_until_s:w, \__fp_use_i_until_s:nw, and \__fp_use_ii_until_s:nnw.)

\__fp_reverse_args:Nww

Many internal functions take arguments delimited by semicolons, and it is occasionally useful to swap two such arguments.
\cs_new:Npn \__fp_reverse_args:Nww #1 #2; #3; { #1 #3; #2; }

(End definition for \__fp_reverse_args:Nww.)

\__fp_rrot:www

Rotate three arguments delimited by semicolons. This is the inverse (or the square) of the Forth primitive \texttt{ROT}, hence the name.
\cs_new:Npn \__fp_rrot:www #1; #2; #3; { #2; #3; #1; }

(End definition for \__fp_rrot:www.)

\__fp_use_i:ww \__fp_use_i:www

Many internal functions take arguments delimited by semicolons, and it is occasionally useful to remove one or two such arguments.
\cs_new:Npn \__fp_use_i:ww #1; #2; { #1; }  
\cs_new:Npn \__fp_use_i:www #1; #2; #3; { #1; }

(End definition for \__fp_use_i:ww and \__fp_use_i:www.)

66.4 Constants, and structure of floating points

\__fp_misused:n

This receives a floating point object (floating point number or tuple) and generates an error stating that it was misused. This is called when for instance an \texttt{fp} variable is left in the input stream and its contents reach \TeX{}’s stomach.
\cs_new_protected:Npn \__fp_misused:n #1  
\msg_error:nnx { fp } { misused } { \fp_to_tl:n {#1} } 

(End definition for \__fp_misused:n.)

\__fp

Floating points numbers all start with \texttt{\s__fp \__fp_chk:w}, where \texttt{\s__fp} is equal to the \TeX{} primitive \texttt{relax}, and \texttt{\__fp_chk:w} is protected. The rest of the floating point number is made of characters (or \texttt{relax}). This ensures that nothing expands under \texttt{f}-expansion, nor under \texttt{x}-expansion. However, when typeset, \texttt{\s__fp} does nothing, and \texttt{\__fp_chk:w} is expanded. We define \texttt{\__fp_chk:w} to produce an error.
\scan_new:N \s__fp  
\cs_new_protected:Npn \__fp_misused:n #1  
\msg_error:nnx { \texttt{fp} } { \texttt{misused} } { \texttt{fp_to_tl:n \{#1\}} } 

(End definition for \texttt{\s__fp} and \texttt{\__fp_chk:w}.)

\s__fp_expr_mark \s__fp_expr_stop

Aliases of \texttt{\tex_relax:D}, used to terminate expressions.
\cs_new:N \s__fp_expr_mark  
\cs_new:N \s__fp_expr_stop 

(End definition for \texttt{\s__fp_expr_mark} and \texttt{\s__fp_expr_stop}.)
\s__fp_mark
\s__fp_stop

Generic scan marks used throughout the module.

\_\_fp_use_i_delimit_by_s_stop:nw

Functions to gobble up to a scan mark.

\s__fp_invalid
\s__fp_underflow
\s__fp_overflow
\s__fp_exact

A couple of scan marks used to indicate where special floating point numbers come from.

\c_zero_fp
\c_minus_zero_fp
\c_inf_fp
\c_minus_inf_fp
\c_nan_fp

The special floating points. We define the floating points here as “exact”.

\c__fp_prec_int
\c__fp_half_prec_int
\c__fp_block_int

The number of digits of floating points.

\c__fp_myriad_int

Blocks have 4 digits so this integer is useful.

\c__fp_minus_min_exponent_int
\c__fp_max_exponent_int

Normal floating point numbers have an exponent between \(-\text{\textbar\textbar}_{\text{min}}\) and \text{\textbar\textbar}_{\text{max}} inclusive. Larger numbers are rounded to \pm\infty. Smaller numbers are rounded to \pm0. It would be more natural to define a \text{\textbar\textbar}_{\text{min}} with the opposite sign but that would waste one \TeX count.

\c__fp_max_exp_exponent_int

If a number’s exponent is larger than that, its exponential overflows/underflows.

(End definition for \s__fp_mark and \s__fp_stop.)

(End definition for \_\_fp_use_i_delimit_by_s_stop:nw.)

(End definition for \s__fp_invalid and others.)

(End definition for \c_zero_fp and others. These variables are documented on page 245.)

(End definition for \c__fp_prec_int, \c__fp_half_prec_int, and \c__fp_block_int.)

(End definition for \c__fp_myriad_int.)

(End definition for \c__fp_minus_min_exponent_int and \c__fp_max_exponent_int.)

(End definition for \c__fp_max_exp_exponent_int.)
A floating point number that is bigger than all normal floating point numbers. This replaces infinities when converting to formats that do not support infinities.

\[ \texttt{\_fp\_overflowing\_fp} \]

In case of overflow or underflow, we have to output a zero or infinity with a given sign.

\[ \texttt{\_fp\_zero\_fp:N\_fp\_inf\_fp:N} \]

For normal numbers, the function expands to the exponent, otherwise to 0. This is used in l3str-format.

\[ \texttt{\_fp\_neg\_sign:N\_fp\_kind:w} \]

Expands to 0 for zeros, 1 for normal floating point numbers, 2 for infinities, 3 for NaN, 4 for tuples.
\section{Overflow, underflow, and exact zero}

Expects the sign and the exponent in some order, then the significand (which we don't touch). Outputs the corresponding floating point number, possibly underflowed to ±0 or overflowed to ±\infty. The functions \texttt{\_fp\_underflow:w} and \texttt{\_fp\_overflow:w} are defined in l3fp-traps.

\begin{verbatim}
\cs_new:Npn \_fp\_sanitize:Nw #1 #2; 
{ 
\if_case:w 
\if_int_compare:w #2 > \c__fp\_max\_exponent\_int 1 - \else:
\if_int_compare:w #2 < \c__fp\_min\_min\_exponent\_int 2 - \else:
\if_meaning:w 1 #1 3 - \fi: \fi: \fi: 0 - 
\or: \exp\_after:wN \_fp\_overflow:w 
\or: \exp\_after:wN \_fp\_underflow:w 
\or: \exp\_after:wN \_fp\_sanitize\_zero:w 
\fi:
\s__fp \_fp\chk:w 1 #1 (#2) }
\cs_new:Npn \_fp\_sanitize:wN #1; #2 { \_fp\_sanitize:Nw #2 #1; }
\cs_new:Npn \_fp\_sanitize\_zero:w \s__fp \_fp\chk:w #1 #2 { \c\zero\_fp }
\end{verbatim}

(End definition for \_fp\_sanitize:Nw, \_fp\_sanitize:wN, and \_fp\_sanitize\_zero:w.)

\section{Expanding after a floating point number}

\begin{verbatim}
\_fp\_exp\_after\_o:w \_fp\_exp\_after\_f:nw \_fp\_exp\_after\_o:w \_fp\_exp\_after\_f:nw (floating point)
\_fp\_exp\_after\_f:nw (tokens) (floating point)
\_fp\_exp\_after\_o:w \_fp\_exp\_after\_f:nw \{\textit{tokens}\} \_fp\_exp\_after\_o:w \_fp\_exp\_after\_f:nw \{\textit{tokens}\}
\end{verbatim}

Places \textit{tokens} (empty in the case of \texttt{\_fp\_exp\_after\_o:w}) between the \textit{(floating point)} and the following tokens, then hits those tokens with o or f-expansion, and leaves the floating point number unchanged.

We first distinguish normal floating points, which have a significand, from the much simpler special floating points.

\begin{verbatim}
\cs_new:Npn \_fp\_exp\_after\_o:w \_fp\_exp\_after\_f:nw \s__fp \_fp\chk:w #1 
{ 
\if_meaning:w 1 #1 \exp\_after:wN \_fp\_exp\_after\_normal:nNNw 
\else: \exp\_after:wN \_fp\_exp\_after\_special:nNNw 
\fi: \} 
\cs_new:Npn \_fp\_exp\_after\_f:nw #1 \s__fp \_fp\chk:w #2 
{ \if_meaning:w 1 #2 \exp\_after:wN \_fp\_exp\_after\_normal:nNNw 
\else: \exp\_after:wN \_fp\_exp\_after\_special:nNNw 
\fi: \} 
\exp:w \exp\_end\_continue\_f:w #1 }
\end{verbatim}

964
\_\_fp\_exp\_after\_special:nNNw

\_\_fp\_exp\_after\_special:nNNw \{(after)\} \{(case)\} \{(sign)\} \{(scan mark)\};

Special floating point numbers are easy to jump over since they contain few tokens.

\cs_new:Npn \__fp_exp_after_special:nNNw #1#2#3#4;
\exp_after:wN \s__fp
\exp_after:wN \__fp_chk:w
\exp_after:wN #2
\exp_after:wN #3
\exp_after:wN #4
\exp_after:wN #1

\cs_new:Npn \__fp_exp_after_normal:nNNw #1 1 #2 #3 #4#5#6#7;
\exp_after:wN \__fp_exp_after_normal:Nwwwww
\exp_after:wN #2
\int_value:w #3 \exp_after:wN ;
\int_value:w 1 #4 \exp_after:wN ;
\int_value:w 1 #5 \exp_after:wN ;
\int_value:w 1 #6 \exp_after:wN ;
\int_value:w 1 #7 \exp_after:wN ; #1

66.7 Other floating point types

Floating point tuples take the form \s\_fp\_tuple \__fp\_tuple\_chk:w \{\fp 1\} \{\fp 2\} \ldots\}; where each \{\fp\} is a floating point number or tuple, hence ends with ; itself. When a tuple is typeset, \__fp\_tuple\_chk:w produces an error, just like usual floating point numbers. Tuples may have zero or one element.

\scan_new:N \s\_fp\_tuple
\cs_new_protected:Npn \\__fp\_tuple\_chk:w #1 ;
\tl_const:Nn \c\_fp\_empty\_tuple_fp
\{ \__fp\_tuple \__fp\_tuple\_chk:w \} ;

(End definition for \s\_fp\_tuple, \__fp\_tuple\_chk:w, and \c\_fp\_empty\_tuple_fp.)
Count the number of items in a tuple of floating points by counting semicolons. The technique is very similar to \texttt{\_\_fp\_tuple\_count:n}, but with the loop built-in. Checking for the end of the loop is done with the \texttt{\use\_none:n #1} construction.

\begin{verbatim}
\cs_new:Npn \__fp\_tuple\_count:w \s__fp\_tuple \__fp\_tuple\_chk:w #1 ;
{ \int_value:w \__fp\_int\_eval:w 0 \__fp\_tuple\_count_loop:Nw #1 { ? \prg\_break: } ; \prg\_break\_point: \__fp\_int\_eval\_end: }
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp\_tuple\_count_loop:Nw #1#2;
{ \use\_none:n #1 + 1 \__fp\_tuple\_count_loop:Nw }
\end{verbatim}

(End definition for \texttt{\_\_fp\_tuple\_count:w}, \texttt{\_\_fp\_array\_count:n}, and \texttt{\_\_fp\_tuple\_count\_loop:Nw}).

Used as \texttt{\_\_fp\_if\_type\_fp:NTwFw ⟨marker⟩ \{⟨true code⟩⟩ \s__fp \{⟨false code⟩⟩ \s__fp\_stop}, this test whether the ⟨marker⟩ is \s__fp or not and runs the appropriate ⟨code⟩. The very unusual syntax is for optimization purposes as that function is used for all floating point operations.

\begin{verbatim}
\cs_new:Npn \__fp\_if\_type\_fp:NTwFw #1 \s__fp #2 #3 \s__fp\_stop {#2}
\end{verbatim}

(End definition for \texttt{\_\_fp\_if\_type\_fp:NTwFw}).

True if all items are floating point numbers. Used for \texttt{min}.

\begin{verbatim}
\cs_new:Npn \__fp\_array\_if\_all\_fp:nTF \__fp\_array\_if\_all\_fp\_loop:w
{ \__fp\_array\_if\_all\_fp\_loop:w #1 { \s__fp \prg\_break: } ; \prg\_break\_point: \use\_i:nn }
\end{verbatim}

(End definition for \texttt{\_\_fp\_array\_if\_all\_fp:nTF} and \texttt{\_\_fp\_array\_if\_all\_fp\_loop:w}).

Used as \texttt{\_\_fp\_type\_from\_scan:N ⟨token⟩}. Grabs the pieces of the stringified ⟨token⟩ which lies after the first \s__fp. If the ⟨token⟩ does not contain that string, the result is _?_.

\begin{verbatim}
\cs_new:Npn \__fp\_type\_from\_scan:N \__fp\_type\_from\_scan\_other:N \__fp\_type\_from\_scan:w
{ \__fp\_if\_type\_fp:NTwFw \s__fp \{ \__fp\_type\_from\_scan\_other:N #1 \} \s__fp\_stop }
\end{verbatim}

\begin{verbatim}
\cs_new:Npx \__fp\_type\_from\_scan\_other:N #1
{ \exp\_not:N \exp\_after:wN \exp\_not:N \__fp\_type\_from\_scan:w }
\end{verbatim}
Arguments are \langle type marker \rangle \langle function \rangle \langle recovery \rangle. This gives the function obtained by placing the type after @@. If the function is not defined then \langle recovery \rangle \langle function \rangle is used instead; however that test is not run when the \langle type marker \rangle is \s__fp.

The Nnw function simply dispatches to the appropriate \__fp_exp_after\_any\_f:nw with "..." (either empty or \langle type \rangle) extracted from #1, which should start with \s__fp. If it doesn’t start with \s__fp the function \__fp_exp_after\_f:nw defined in l3fp-parse gives an error; another special \langle type \rangle is stop, useful for loops, see below. The nw function has an important optimization for floating points numbers; it also fetches its type marker #2 from the floating point.
The loop works by using the \texttt{\texttt{n}} argument of \_\_fp\_exp\_after\_any\_f:nw to place the loop macro after the next item in the tuple and expand it.

\begin{verbatim}
\_\_fp\_exp\_after\_array\_f:w
\_\_fp\_exp\_after\_tuple\_o:w
\end{verbatim}

\texttt{\_\_fp\_exp\_after\_tuple\_f:nw}

\texttt{\_\_fp\_exp\_after\_array\_f:w}

The idea is that adding \texttt{10^9} to the number ensures that it has exactly 9 digits, and can then easily find which digits correspond to what position in the number. Of course, this can be modified for any number of digits less or equal to 9 (we are limited by \TeX’s integers). This method is very heavily relied upon in l3fp-basics.

More specifically, the auxiliary inserts \texttt{+ \texttt{\#1\#2\#3\#4\#5}; \{\texttt{\#6}\}}, which allows us to compute several blocks of 4 digits in a nested manner, performing carries on the fly. Say we want to compute \texttt{1 2345 \times 6677 8899}. With simplified names, we would do

\begin{verbatim}
\end{verbatim}

\texttt{968}
The \exp_after:wN \post_processing:w \__fp_int_value:w \__fp_int_eval:w, which starts a first computation, whose initial value is $-5\,0000$ (the “leading shift”). In that computation appears an \exp_after:wN \pack:NNNNw \__fp_int_value:w \__fp_int_eval:w with starting value $4\,9995\,0000$ (the “middle shift”). That, in turn, expands \exp_after:wN \pack:NNNNw which triggers the third computation. The third computation’s value is $5\,0000\,0000 + 12345 \times 8899$, which has 9 digits. Adding $5 \times 10^8$ to the product allowed us to know how many digits to expect as long as the numbers to multiply are not too big; it also works to some extent with negative results. The pack function puts the last 4 of those 9 digits into a brace group, moves the semi-colon delimiter, and inserts a $+$, which combines the carry with the previous computation. The shifts nicely combine into $5\,0000\,0000\times 10^4 + 4\,9995\,0000 = 5\,0000\,0000$. As long as the operands are in some range, the result of this second computation has 9 digits. The corresponding pack function, expanded after the result is computed, braces the last 4 digits, and leaves $+ \langle 5 \text{ digits} \rangle$ for the initial computation. The “leading shift” cancels the combination of the other shifts, and the \post_processing:w takes care of packing the last few digits.

Admittedly, this is quite intricate. It is probably the key in making \l3fp as fast as other pure \TeX floating point units despite its increased precision. In fact, this is used so much that we provide different sets of packing functions and shifts, depending on ranges of input.

This set of shifts allows for computations involving results in the range $[-4\times 10^8, 5\times 10^8 - 1]$. Shifted values all have exactly 9 digits.

\begin{verbatim}
\__fp_pack:NNNNw \c__fp_leading_shift_int \c__fp_middle_shift_int \c__fp_leading_shift_int
\cs_new:Npn \__fp_pack:NNNNw #1#2#3#4#5 #6; { + #1#2#3#4#5 ; {#6} }
\end{verbatim}

This set of shifts allows for computations involving results in the range $[-5\times 10^9, 6\times 10^8 - 1]$ (actually a bit more). Shifted values all have exactly 10 digits. Note that the upper bound is due to \TeX’s limit of $2^{31} - 1$ on integers. The shifts are chosen to be roughly the mid-point of $10^9$ and $2^{31}$, the two bounds on 10-digit integers in \TeX.

\begin{verbatim}
\__fp_pack_big:NNNNNNw \c__fp_big_leading_shift_int \c__fp_big_middle_shift_int \c__fp_big_leading_shift_int
\cs_new:Npn \__fp_pack_big:NNNNNNw #1#2#3#4#5#6#7 #8; { + #1#2#3#4#5#6#7 ; {#8} }
\end{verbatim}

(End definition for \__fp_pack:NNNNw and others.)

This set of shifts allows for computations involving results in the range $[-5\times 10^9, 6\times 10^8 - 1]$ (actually a bit more). Shifted values all have exactly 10 digits. Note that the upper bound is due to \TeX’s limit of $2^{31} - 1$ on integers. The shifts are chosen to be roughly the mid-point of $10^9$ and $2^{31}$, the two bounds on 10-digit integers in \TeX.

\begin{verbatim}
\__fp_pack_big:NNNNNNw \c__fp_big_leading_shift_int \c__fp_big_middle_shift_int \c__fp_big_leading_shift_int
\cs_new:Npn \__fp_pack_big:NNNNNNw #1#2#3#4#5#6 #7; { + #1#2#3#4#5#6 ; {#7} }
\end{verbatim}

(End definition for \__fp_pack_big:NNNNNNw and others.)
This set of shifts allows for computations with results in the range \([-1 \cdot 10^9, 147483647]\); the end-point is \(2^{31} - 1 - 2 \cdot 10^9 \approx 1.47 \cdot 10^{10}\). Shifted values all have exactly 10 digits.

\begin{verbatim}
\int_const:Nn \c__fp_Bigg_leading_shift_int { - 200000 }
\int_const:Nn \c__fp_Bigg_middle_shift_int { 20000 * 9999 }
\int_const:Nn \c__fp_Bigg_trailing_shift_int { 20000 * 10000 }
\cs_new:Npn \__fp_pack_Bigg:NNNNNNw \#1\#2\#3\#4\#5\#6 \#7; { + \#1\#2\#3\#4\#5\#6 ; {\#7} }
\end{verbatim}

(End definition for \__fp_pack_Bigg:NNNNNNw and others.)

\begin{verbatim}
\cs_new:Npn \__fp_pack_twice_four:wNNNNNNNN #1; #2#3#4#5 #6#7#8#9 { #1 {#2#3#4#5} {#6#7#8#9} ; }
\end{verbatim}

(End definition for \__fp_pack_twice_four:wNNNNNNNN.)

\begin{verbatim}
\cs_new:Npn \__fp_pack_eight:wNNNNNNNN #1; #2#3#4#5 #6#7#8#9 { #1 {#2#3#4#5#6#7#8#9} ; }
\end{verbatim}

(End definition for \__fp_pack_eight:wNNNNNNNN.)

Addition and multiplication of significands are done in two steps: first compute a (more or less) exact result, then round and pack digits in the final (braced) form. These functions take care of the packing, with special attention given to the case where rounding has caused a carry. Since rounding can only shift the final digit by 1, a carry always produces an exact power of 10. Thus, \__fp_basics_pack_high_carry:w is always followed by four times \{0000\}.

This is used in l3fp-basics and l3fp-extended.

\begin{verbatim}
\cs_new:Npn \__fp_basics_pack_low:NNNNNw \#1 \#2#3#4#5 \#6; { + \#1 - 1 ; {\#2#3#4#5} {\#6} ; }
\cs_new:Npn \__fp_basics_pack_high:NNNNNw \#1 \#2#3#4#5 \#6; { \if_meaning:w 2 \#1 \__fp_basics_pack_high_carry:w \fi: ; {\#2#3#4#5} {\#6} }
\cs_new:Npn \__fp_basics_pack_high_carry:w \fi: ; \{ \if: \fi: + 1 ; \{0000\} \}
\end{verbatim}

(End definition for \__fp_basics_pack_low:NNNNNw, \__fp_basics_pack_high:NNNNNw, and \__fp_basics_pack_high_carry:w.)
This is used in \texttt{l3fp-basics} for additions and divisions. Their syntax is confusing, hence the name.

\begin{verbatim}
\cs_new:Npn \__fp_basics_pack_weird_low:NNNNw #1 #2#3#4 #5; 
{ \if_meaning:w 2 #1 + 1 \fi: \__fp_int_eval_end: #2#3#4; {#5} ; }
\cs_new:Npn \__fp_basics_pack_weird_high:NNNNNNNNw 1 #1#2#3#4 #5#6#7#8 #9; { ; {#1#2#3#4} {#5#6#7#8} {#9} }
\end{verbatim}

(End definition for \__fp_basics_pack_weird_low:NNNNw and \__fp_basics_pack_weird_high:NNNNNNNNw.)

\subsection{Decimate (dividing by a power of 10)}

\begin{verbatim}
\__fp_decimate:nNnnn \{\langle shift\rangle\} \{f_1\}
{\langle X_1\rangle} {\langle X_2\rangle} {\langle X_3\rangle} {\langle X_4\rangle}
\end{verbatim}

Each \(X_i\) consists in 4 digits exactly, and 1000 \(\leq \langle X_1 \rangle < 9999\). The first argument determines by how much we shift the digits. \(f_1\) is called as follows:

\[
\langle f_1 \rangle \langle\text{rounding}\rangle \{\langle X'_1 \rangle\} \{\langle X'_2 \rangle\} \{\text{extra-digits} \}
\]

where \(0 \leq \langle X'_i \rangle < 10^8 - 1\) are 8 digit integers, forming the truncation of our number. In other words,

\[
\left(\sum_{i=1}^{4} \langle X_i \rangle \cdot 10^{-4i} \cdot 10^{-\langle shift \rangle}\right) - \left(\langle X'_1 \rangle \cdot 10^{-8} + \langle X'_2 \rangle \cdot 10^{-16}\right) = 0 \cdot \langle \text{extra-digits} \rangle \cdot 10^{-16} \in [0, 10^{-16}).
\]

To round properly later, we need to remember some information about the difference. The \langle\text{rounding}\rangle digit is \(0\) if and only if the difference is exactly \(0\), and \(5\) if and only if the difference is exactly \(0.5 \cdot 10^{-16}\). Otherwise, it is the (non-\(0\), non-\(5\)) digit closest to \(10^{17}\) times the difference. In particular, if the shift is \(17\) or more, all the digits are dropped, \langle\text{rounding}\rangle is \(1\) (not \(0\)), and \(\langle X'_1 \rangle\) and \(\langle X'_2 \rangle\) are both zero.

If the shift is \(1\), the \langle\text{rounding}\rangle digit is simply the only digit that was pushed out of the brace groups (this is important for subtraction). It would be more natural for the \langle\text{rounding}\rangle digit to be placed after the \(\langle X'_i \rangle\), but the choice we make involves less reshuffling.

Note that this function treats negative \langle\text{shift}\rangle as \(0\).

\begin{verbatim}
\cs_new:Npn \__fp_decimate:nNnnn \{#1\} \{f_1\}
\cs:w \__fp_decimate_ \if_int_compare:w \__fp_int_eval:w #1 > \c__fp_prec_int tiny \else: \__fp_int_to_roman:w \__fp_int_eval:w #1 \fi:
\cs_end:
\end{verbatim}
Each of the auxiliaries see the function $f_1$, followed by 4 blocks of 4 digits.

*(End definition for \_fp_decimate:_Nnnnn.)*

If the \(<shift>\) is zero, or too big, life is very easy.

\begin{verbatim}
\cs_new:Npn \_fp_decimate_:Nnnnn #1 #2#3#4#5
\{ #1 0 {#2#3} {#4#5} ; \}
\cs_new:Npn \_fp_decimate_tiny:Nnnnn #1 #2#3#4#5
\{ #1 1 { 0000 0000 } { 0000 0000 } 0 #2#3#4#5 ; \}
\end{verbatim}

*(End definition for \_fp_decimate_:Nnnnn and \_fp_decimate_tiny:Nnnnn.)*

Shifting happens in two steps: compute the \(<rounding>\) digit, and repack digits into two blocks of 8. The sixteen functions are very similar, and defined through \_fp_tmp:w. The arguments are as follows: #1 indicates which function is being defined; after one step of expansion, #2 yields the “extra digits” which are then converted by \_fp_round_digit:Nw to the \(<rounding>\) digit (note the + separating blocks of digits to avoid overflowing TeX’s integers). This triggers the r-expansion of \_fp_decimate_pack:nnnnnnnnw,\(^\text{10}\) responsible for building two blocks of 8 digits, and removing the rest. For this to work, #3 alternates between braced and unbraced blocks of 4 digits, in such a way that the 5 first and 5 next token groups yield the correct blocks of 8 digits.

\begin{verbatim}
\cs_new:Npn \_fp_tmp:w #1 #2 #3
\{ \cs_new:cpn { __fp_decimate_ #1 :Nnnnn } ##1 ##2##3##4##5
\{ \exp_after:wN ##1 \int_value:w \exp_after:wN \_fp_round_digit:Nw #2 ; \_fp_decimate_pack:nnnnnnnnnww #3 ; \}
\}
\__fp_tmp:w {i} \{\use_none:nnn #50}{ 0{#2}#3{#4}#5 }
\__fp_tmp:w {ii} \{\use_none:nn #5}{ 00{#2}#3{#4}#5 }
\__fp_tmp:w {iii} \{\use_none:n #5}{ 000{#2}#3{#4}#5 }
\__fp_tmp:w {iv} \{ #5}{ (0000)2(#3)4 #5 }
\__fp_tmp:w {v} \{ \use_none:nnn #4#5}{ 0(0000)2(#3)4 #5 }
\__fp_tmp:w {vi} \{ \use_none:nn #4#5}{ 00(0000)2(#3)4 #5 }
\__fp_tmp:w {vii} \{ \use_none:n #4#5}{ 000(0000)2(#3)4 #5 }
\__fp_tmp:w {viii}{ #4#5}{ (0000)(0000)(#2)3 #4 #5 }
\__fp_tmp:w {ix} \{ \use_none:nnn #3#5}{ (00000000)(#2)3 #4 #5 }
\__fp_tmp:w {x} \{ \use_none:n #3#5}{ (00000000)(#2)3 #4 #5 }
\__fp_tmp:w {xi} \{ \use_none:nnn #3#5}{ (00000000)(#2)3 #4 #5 }
\__fp_tmp:w {xii}{ #3#5}{ (00000000)(#2)3 #4 #5 }
\__fp_tmp:w {xiii} \{ \use_none:nnn #2#5}{ (00000000)(#2)3 #4 #5 }
\__fp_tmp:w {xiv} \{ \use_none:n #2#5}{ (00000000)(#2)3 #4 #5 }
\__fp_tmp:w {xv} \{ \use_none:nnn #2#5}{ (00000000)(#2)3 #4 #5 }
\__fp_tmp:w {xvi}{ #2#5}{ (00000000)(#2)3 #4 #5}
\end{verbatim}

*(End definition for \_fp_decimate_auxi:Nnnnn and others.)*

\(^{10}\)No, the argument spec is not a mistake: the function calls an auxiliary to do half of the job.
The computation of the \textit{(rounding)} digit leaves an unfinished \texttt{\int_value:w}, which expands the following functions. This allows us to repack nicely the digits we keep. Those digits come as an alternation of unbraced and braced blocks of 4 digits, such that the first 5 groups of token consist in 4 single digits, and one brace group (in some order), and the next 5 have the same structure. This is followed by some digits and a semicolon.

\begin{verbatim}
\cs_new:Npn \__fp_decimate_pack:nnnnnnnnnnw #1#2#3#4#5
\__fp_decimate_pack:nnnnnnnnnnw { #1#2#3#4#5 } \\\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_decimate_pack:nnnnnnnnnnw #1 #2#3#4#5#6
\{ #1 \{ #2#3#4#5#6 \} \}
\end{verbatim}

\textit{(End definition for \__fp_decimate_pack:nnnnnnnnnnw.)}

66.10 \textbf{Functions for use within primitive conditional branches}

The functions described in this section are not pretty and can easily be misused. When correctly used, each of them removes one \texttt{\fi:} as part of its parameter text, and puts one back as part of its replacement text.

Many computation functions in \texttt{l3fp} must perform tests on the type of floating points that they receive. This is often done in an \texttt{\if_case:w} statement or another conditional statement, and only a few cases lead to actual computations: most of the special cases are treated using a few standard functions which we define now. A typical use context for those functions would be

\begin{verbatim}
\if_case:w (integer) \exp_stop_f:
  \__fp_case_return_o:Nw \texttt{\fp var}
\or: \__fp_case_use:nw \{ \texttt{some computation} \}
\or: \__fp_case_return_same_o:w
\or: \__fp_case_return:nw \{ \texttt{something} \}
\fi:
\langle \texttt{junk} \rangle
\langle \texttt{floating point} \rangle
\end{verbatim}

In this example, the case 0 returns the floating point \langle \texttt{fp var} \rangle, expanding once after that floating point. Case 1 does \langle \texttt{some computation} \rangle using the \langle \texttt{floating point} \rangle (presumably compute the operation requested by the user in that non-trivial case). Case 2 returns the \langle \texttt{floating point} \rangle without modifying it, removing the \langle \texttt{junk} \rangle and expanding once after. Case 3 closes the conditional, removes the \langle \texttt{junk} \rangle and the \langle \texttt{floating point} \rangle, and expands \langle \texttt{something} \rangle next. In other cases, the \langle \texttt{\langle junk\rangle} \rangle is expanded, performing some other operation on the \langle \texttt{floating point} \rangle. We provide similar functions with two trailing \langle \texttt{floating points} \rangle.

\begin{verbatim}
\__fp_case_use:nw
\end{verbatim}

This function ends a \TeX conditional, removes junk until the next floating point, and places its first argument before that floating point, to perform some operation on the floating point.

\begin{verbatim}
\cs_new:Npn \__fp_case_use:nw \texttt{\#1} \#2 \fi: \#3 \exp_stop_f \{ \fi: \#1 \texttt{\#2} \}
\end{verbatim}

\textit{(End definition for \__fp_case_use:nw.)}
\_fp_case_return:nw  This function ends a TeX conditional, removes junk and a floating point, and places its first argument in the input stream. A quirk is that we don’t define this function requiring a floating point to follow, simply anything ending in a semicolon. This, in turn, means that the \textlangle junk\textrangle may not contain semicolons.

```latex
\cs_new:Npn \_fp_case_return:nw #1 \fi: #3 ; { \fi: #1 }
```

(End definition for \_fp_case_return:nw.)

\_fp_case_return_o:Nw  This function ends a TeX conditional, removes junk and a floating point, and returns its first argument (an \textlangle fp var\textrangle) then expands once after it.

```latex
\cs_new:Npn \_fp_case_return_o:Nw #1 \fi: #3 \s__fp #4 ; { \fi: \exp_after:wN #1 }
```

(End definition for \_fp_case_return_o:Nw.)

\_fp_case_return_same_o:w  This function ends a TeX conditional, removes junk, and returns the following floating point, expanding once after it.

```latex
\cs_new:Npn \_fp_case_return_same_o:w #1 \fi: #2 \s__fp #3 ; { \fi: \__fp_exp_after:o:w \s__fp #3 ; }
```

(End definition for \_fp_case_return_same_o:w.)

\_fp_case_return_o:Nww  Same as \_fp_case_return_o:Nw but with two trailing floating points.

```latex
\cs_new:Npn \_fp_case_return_o:Nww #1 \fi: #3 \s__fp #4 ; #5 ; { \fi: \exp_after:wN #1 }
```

(End definition for \_fp_case_return_o:Nww.)

\_fp_case_return_i_o:ww \_fp_case_return_ii_o:ww  Similar to \_fp_case_return_same_o:w, but this returns the first or second of two trailing floating point numbers, expanding once after the result.

```latex
\cs_new:Npn \_fp_case_return_i_o:ww #1 \fi: #2 \s__fp #3 ; \s__fp #4 ; { \fi: \__fp_exp_after:o:w #3 ; }
\cs_new:Npn \_fp_case_return_ii_o:ww #1 \fi: #2 \s__fp #3 ; { \fi: \__fp_exp_after:o:w }
```

(End definition for \_fp_case_return_i_o:ww and \_fp_case_return_ii_o:ww.)

66.11 Integer floating points

\_fp_int:p:w \_fp_int:wTF  Tests if the floating point argument is an integer. For normal floating point numbers, this holds if the rounding digit resulting from \_fp_decimate:nNnnnn is 0.

```latex
\prg_new_conditional:Npnn \_fp_int:w \s__fp \__fp_chk:w #1 #2 #3 #4; { TF \ , \ T \ , \ F \ , \ p }
\prg_return_true:
\prg_return_false:
```

\_fp_int_p:w
66.12 Small integer floating points

Tests if the floating point argument is an integer or \( \pm \infty \). If so, it is clipped to an integer in the range \([-10^8, 10^8]\) and fed as a braced argument to the \langle true code \rangle. Otherwise, the \langle false code \rangle is performed.

First filter special cases: zeros and infinities are integers, \texttt{nan} is not. For normal numbers, decimate. If the rounding digit is not 0 run the \langle false code \rangle. If it is, then the integer is \#2 \#3; use \#3 if \#2 vanishes and otherwise \( 10^8 \).

\begin{verbatim}
cs_new:Npn \_fp_small_int:wTF \s__fp \_fp_chk:w #1#2 { \if_case:w #1 \exp_stop_f: \_fp_case_return:nw { \_fp_small_int_true:wTF 0 ; } \or: \exp_after:wN \_fp_small_int_normal:NnwTF \or: \_fp_case_return:nw \exp_after:wN \_fp_small_int_true:wTF \int_value:w \if_meaning:w 2 #2 - \fi: 1 0000 0000 ; } \else: \_fp_case_return:nw \use_ii:nn \fi: #2 } \cs_new:Npn \_fp_small_int_true:wTF #1; #2#3 { #2 {#1} } \cs_new:Npn \_fp_small_int_normal:NnwTF #1#2#3; { \_fp_decimate:nNnnnn { \c__fp_prec_int - #2 } \_fp_small_int_test:NnnwNw #3 #1 } \cs_new:Npn \_fp_small_int_test:NnnwNw #1#2#3#4; #5 { \if_meaning:w 0 #1 \exp_after:wN \_fp_small_int_true:wTF \int_value:w \if_meaning:w 2 #5 - \fi: \_if_int_compare:w #2 > \c_zero_int 1 0000 0000 \else: \_if_int_compare:w #3 \c_zero_int \fi: \_fp_after:wN ; \else: \_fp_after:wN \use_ii:nn \fi: #5 } \if_meaning:w 0 \c_zero_int \else: \_fp_after:wN \use_ii:nn \fi: #5 } \end{verbatim}

(End definition for \_fp_small_int:wTF and others.)
66.13 Fast string comparison

\__fp_str_if_eq:nn

A private version of the low-level string comparison function.

\cs_new_eq:NN \__fp_str_if_eq:nn \tex_strcmp:D

(End definition for \__fp_str_if_eq:nn.)

66.14 Name of a function from its l3fp-parse name

\__fp_func_to_name:N \__fp_func_to_name_aux:w

The goal is to convert for instance \__fp_sin_o:w to sin. This is used in error messages hence does not need to be fast.

\cs_new:Npn \__fp_func_to_name:N #1
\exp_last_unbraced:Nf \__fp_func_to_name_aux:w { \cs_to_str:N #1 } X
\cs_set_protected:Npn \__fp_tmp:w #1 #2
{ \cs_new:Npn \__fp_func_to_name_aux:w ##1 #1 ##2 #2 ##3 X {##2} }
\exp_args:Nff \__fp_tmp:w { \tl_to_str:n { __fp_ } }
{ \tl_to_str:n { _o: } }

(End definition for \__fp_func_to_name:N and \__fp_func_to_name_aux:w.)

66.15 Messages

Using a floating point directly is an error.

\msg_new:nnnn { fp } { misused }{ A floating-point-with-value-’#1’-was-misused. }
\{ To-obtain-the-value-of-a-floating-point-variable,-use-
\’\token_to_str:N \fp_to_decimal:N’,,-
\’\token_to_str:N \fp_to_tl:N’,,-or-other-
conversion-functions. }
\prop_gput:Nnn \g_msg_module_name_prop { fp } { LaTeX3 }
\prop_gput:Nnn \g_msg_module_type_prop { fp } { }

{/package}
Chapter 67

13fp-traps Implementation

Exceptions should be accessed by an n-type argument, among

- invalid_operation
- division_by_zero
- overflow
- underflow
- inexact (actually never used).

### 67.1 Flags

Flags to denote exceptions.

```plaintext
flag fp_invalid_operation
flag fp_division_by_zero
flag fp_overflow
flag fp_underflow

(flag_new:n { fp_invalid_operation })
(flag_new:n { fp_division_by_zero })
(flag_new:n { fp_overflow })
(flag_new:n { fp_underflow })
```

(End definition for flag fp_invalid_operation and others. These variables are documented on page 247.)

### 67.2 Traps

Exceptions can be trapped to obtain custom behaviour. When an invalid operation or a division by zero is trapped, the trap receives as arguments the result as an n-type floating point number, the function name (multiple letters for prefix operations, or a single symbol for infix operations), and the operand(s). When an overflow or underflow is trapped, the trap receives the resulting overly large or small floating point number if it is not too big, otherwise it receives +∞. Currently, the inexact exception is entirely ignored.

The behaviour when an exception occurs is controlled by the definitions of the functions

```plaintext
\_fp_invalid_operation
```

977
• \_fp_invalid_operation_o:Nww,
• \_fp_invalid_operation_tl_o:ff,
• \_fp_division_by_zero_o:Nww,
• \_fp_division_by_zero_o:NNww,
• \_fp_overflow:w,
• \_fp_underflow:w.

Rather than changing them directly, we provide a user interface as \fp_trap:nn \{⟨exception⟩\} \{⟨way of trapping⟩\}, where the ⟨way of trapping⟩ is one of error, flag, or none.

We also provide \_fp_invalid_operation_o:nw, defined in terms of \_fp_invalid_operation:nw.

\fp_trap:nn

We provide three types of trapping for invalid operations: either produce an error and raise the relevant flag; or only raise the flag; or don’t even raise the flag. In most cases, the function produces as a result its first argument, possibly with post-expansion.

\_fp_trap_invalid_operation_set_error:
\_fp_trap_invalid_operation_set_flag:
\_fp_trap_invalid_operation_set_none:
\_fp_trap_invalid_operation_set:N
\flag_raise_if_clear:n \{ fp_invalid_operation \}
\#1
\}
\exp_args:Nno \use:n
\{ \cs_set:Npn \__fp_invalid_operation_o:Nww ##1##2; ##3; \}
\{
#1
\__fp_error:nffn \{ invalid-ii \}
\{ \fp_to_tl:n \{ ##2; \} \} \{ \fp_to_tl:n \{ ##3; \} \} \{##1\}
\flag_raise_if_clear:n \{ fp_invalid_operation \}
\exp_after:wN \c_nan_fp
\}
\exp_args:Nno \use:n
\{ \cs_set:Npn \__fp_invalid_operation_tl_o:ff ##1##2 \}
\{
#1
\__fp_error:nffn \{ invalid \} \{##1\} \{##2\} \{ \}
\flag_raise_if_clear:n \{ fp_invalid_operation \}
\exp_after:wN \c_nan_fp
\}
\}

(End definition for \__fp_trap_invalid_operation_set_error: and others.)

\__fp_trap_division_by_zero_set_error:
\__fp_trap_division_by_zero_set_flag:
\__fp_trap_division_by_zero_set_none:
\__fp_trap_division_by_zero_set:N

We provide three types of trapping for invalid operations and division by zero: either produce an error and raise the relevant flag; or only raise the flag; or don’t even raise the flag. In all cases, the function must produce a result, namely its first argument, ±∞ or NaN.

\cs_new_protected:Npn \__fp_trap_division_by_zero_set_error:
{ \__fp_trap_division_by_zero_set:N \prg_do_nothing: }
\cs_new_protected:Npn \__fp_trap_division_by_zero_set_flag:
{ \__fp_trap_division_by_zero_set:N \use_none:nnnnn }
\cs_new_protected:Npn \__fp_trap_division_by_zero_set_none:
{ \__fp_trap_division_by_zero_set:N \use_none:nnnnnnn }
\cs_new_protected:Npn \__fp_trap_division_by_zero_set:N #1
{ \exp_args:Nno \use:n
\{ \cs_set:Npn \__fp_division_by_zero_o:Nnw ##1##2##3; \}
\{
#1
\__fp_error:nffn \{ zero-div \} \{##1\} \{ ##2; \} \{ \}
\flag_raise_if_clear:n \{ fp_division_by_zero \}
\exp_after:wN \c_nan_fp
\}
}
\exp_args:Nno \use:n
\{ \cs_set:Npn \__fp_division_by_zero_o:Nww #1##2##3; \}
\{
#1
\__fp_error:nffn \{ zero-div-ii \}
\{ ##2; \} \{ ##3; \} \{ \}
\flag_raise_if_clear:n \{ fp_division_by_zero \}
\exp_after:wN \c_nan_fp
\}
\}
\}

979
Just as for invalid operations and division by zero, the three different behaviours are obtained by feeding \prg_do_nothing:, \use_none:nnnn or \use_none:nnnnnnn to an auxiliary, with a further auxiliary common to overflow and underflow functions. In most cases, the argument of the \__fp_overflow:w and \__fp_underflow:w functions will be an (almost) normal number (with an exponent outside the allowed range), and the error message thus displays that number together with the result to which it overflowed or underflowed. For extreme cases such as 10 ** 1e9999, the exponent would be too large for \TeX{}, and \__fp_overflow:w receives ±\infty (\__fp_underflow:w would receive ±0); then we cannot do better than simply say an overflow or underflow occurred.

(End definition for \__fp_trap_division_by_zero_set_error: and others.)

Initialize the control sequences (to log properly their existence). Then set invalid operations to trigger an error, and division by zero, overflow, and underflow to act silently on their flag.

(End definition for \__fp_trap_division_by_zero_set_error: and others.)
Convenient short-hands for returning \c_nan_fp for a unary or binary operation, and expanding after.

\begin{align*}
\cs_new:Npn \__fp_invalid_operation_o:nw &= \__fp_invalid_operation:nnw { \exp_after:wN \c_nan_fp } \\
\cs_generate_variant:Nn \__fp_invalid_operation_o:nw { f }
\end{align*}

(End definition for \__fp_invalid_operation_o:nw.)

### 67.3 Errors

### 67.4 Messages

Some messages.

\begin{align*}
\cs_new:nnnn \__fp_error:nnnn &= \__fp_error:nnfn \__fp_error:nffn \__fp_error:nfff \\
\cs_generate_variant:Nn \__fp_error:nnnn { nnf, nff, nfff }
\end{align*}

(End definition for \__fp_error:nnnn.)
* - none

}\msg_new:nnn { fp } { flow }
{-#3 - occurred. #}
\msg_new:nnn { fp } { flow-to }
{-#1 - #3 ed - to - #2 . #}
\msg_new:nnn { fp } { zero-div }
{-#Division-by-zero-in- #1 (#2) #}
\msg_new:nnn { fp } { zero-div-ii }
{-#Division-by-zero-in- (#1) #3 (#2) #}
\msg_new:nnn { fp } { invalid }
{-#Invalid-operation- #1 (#2) #}
\msg_new:nnn { fp } { invalid-ii }
{-#Invalid-operation- (#1) #3 (#2) #}
\msg_new:nnn { fp } { unknown-type }
{-#Unknown-type-for- '#1' #}
\msg_new:nnn { fp } { zero-div }

/*package*/
Chapter 68

\texttt{l3fp-round} implementation

\begin{verbatim}
\__fp_parse_word_trunc:N = \__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_zero:NNN
\__fp_parse_word_floor:N = \__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_ninf:NNN
\__fp_parse_word_ceil:N = \__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_pinf:NNN
\__fp_parse_word_round:N #1#2 = \__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_nearest:NNN #1 #2
\__fp_parse_round:Nw #1 #2 \__fp_round_to_nearest:NNN #3 #4 = [#2 #1 #3]
\end{verbatim}

(End definition for \_\_fp_parse_word_trunc:N, \_\_fp_parse_word_floor:N, and \_\_fp_parse_word_ceil:N.)

(End definition for \_\_fp_parse_word_round:N and \_\_fp_parse_round:Nw.)

68.1 Rounding tools

\texttt{c\_fp_five_int} This is used as the half-point for which numbers are rounded up/down.

\begin{verbatim}
\int_const:Nn \c\_fp_five_int { 5 }
\end{verbatim}

(End definition for \_c\_fp_five_int.)

Floating point operations often yield a result that cannot be exactly represented in a significand with 16 digits. In that case, we need to round the exact result to a representable number. The IEEE standard defines four rounding modes:
• Round to nearest: round to the representable floating point number whose absolute difference with the exact result is the smallest. If the exact result lies exactly at the mid-point between two consecutive representable floating point numbers, round to the floating point number whose last digit is even.

• Round towards negative infinity: round to the greatest floating point number not larger than the exact result.

• Round towards zero: round to a floating point number with the same sign as the exact result, with the largest absolute value not larger than the absolute value of the exact result.

• Round towards positive infinity: round to the least floating point number not smaller than the exact result.

This is not fully implemented in l3fp yet, and transcendental functions fall back on the “round to nearest” mode. All rounding for basic algebra is done through the functions defined in this module, which can be redefined to change their rounding behaviour (but there is not interface for that yet).

The rounding tools available in this module are many variations on a base function \_\_fp_round:NNN, which expands to 0\exp_stop_f: or 1\exp_stop_f: depending on whether the final result should be rounded up or down.

• \_\_fp_round:NNN \langle sign \rangle \langle digit1 \rangle \langle digit2 \rangle can expand to 0\exp_stop_f: or 1\exp_stop_f:.

• \_\_fp_round_s:NNN\langle sign \rangle \langle digit1 \rangle \langle digit2 \rangle <more digits> can expand to 0\exp_stop_f:; or 1\exp_stop_f:;.

• \_\_fp_round_neg:NNN \langle sign \rangle \langle digit1 \rangle \langle digit2 \rangle can expand to 0\exp_stop_f: or 1\exp_stop_f:.

See implementation comments for details on the syntax.


If rounding the number \langle final sign \rangle \langle digit1 \rangle \langle digit2 \rangle to an integer rounds it towards zero (truncates it), this function expands to 0\exp_stop_f:, and otherwise to 1\exp_stop_f:. Typically used within the scope of an \_\_fp_int_eval:w, to add 1 if needed, and thereby round correctly. The result depends on the rounding mode.

It is very important that \langle final sign \rangle be the final sign of the result. Otherwise, the result would be incorrect in the case of rounding towards \(-\infty\) or towards \(+\infty\). Also recall that \langle final sign \rangle is 0 for positive, and 2 for negative.

By default, the functions below return 0\exp_stop_f:, but this is superseded by \_\_fp_round_return_one:, which instead returns 1\exp_stop_f:, expanding everything and removing 0\exp_stop_f: in the process. In the case of rounding towards \(\pm\infty\) or towards 0, this is not really useful, but it prepares us for the “round to nearest, ties to even” mode.

The “round to nearest” mode is the default. If the \langle digit2 \rangle is larger than 5, then round up. If it is less than 5, round down. If it is exactly 5, then round such that \langle digit1 \rangle plus the result is even. In other words, round up if \langle digit1 \rangle is odd.

The “round to nearest” mode has three variants, which differ in how ties are rounded: down towards \(-\infty\), truncated towards 0, or up towards \(+\infty\).
\cs_new:Npn \_fp_round_to_ninf:NNN #1 #2 #3 
{
  \if_meaning:w 2 #1
    \if_int_compare:w #3 > \c_zero_int
      \_fp_round_return_one:
    \fi:
  \fi:
  \c_zero_int
}
\cs_new:Npn \_fp_round_to_zero:NNN #1 #2 #3 { \c_zero_int }
\cs_new:Npn \_fp_round_to_pinf:NNN #1 #2 #3 
{
  \if_meaning:w 0 #1
    \if_int_compare:w #3 > \c_zero_int
      \_fp_round_return_one:
    \fi:
  \fi:
  \c_zero_int
}
\cs_new:Npn \_fp_round_to_nearest:NNN #1 #2 #3 
{
  \if_int_compare:w #3 > \c__fp_five_int
    \_fp_round_return_one:
  \else:
    \if_meaning:w 5 #3
      \if_int_odd:w #2 \exp_stop_f:
        \_fp_round_return_one:
      \fi:
    \fi:
    \c_zero_int
  \fi:
}
\cs_new:Npn \_fp_round_to_nearest_ninf:NNN #1 #2 #3 
{
  \if_int_compare:w #3 > \c__fp_five_int
    \_fp_round_return_one:
  \else:
    \if_meaning:w 5 #3
      \if_meaning:w 2 #1
        \_fp_round_return_one:
      \fi:
    \fi:
    \c_zero_int
  \fi:
}
\cs_new:Npn \_fp_round_to_nearest_zero:NNN #1 #2 #3 
{
  \if_int_compare:w #3 > \c__fp_five_int
    \_fp_round_return_one:
  \fi:
  \c_zero_int
}
\cs_new:Npn \_fp_round_to_nearest_pinf:NNN #1 #2 #3 
{

\if_int_compare:w #3 > \c__fp_five_int
  \__fp_round_return_one:
\else:
  \if_meaning:w 5 #3
    \if_meaning:w 0 #1
      \__fp_round_return_one:
    \fi:
  \fi:
  \c_zero_int
\fi:
\fi:
\cs_new_eq:NN \__fp_round:NNN \__fp_round_to_nearest:NNN
(End definition for \__fp_round:NNN and others.)
\__fp_round_s:NNNw \__fp_round_s:NNNw ⟨final sign⟩ ⟨digit⟩ ⟨more digits⟩;
Similar to \__fp_round:NNN, but with an extra semicolon, this function expands
to 0\exp_stop_f; if rounding ⟨final sign⟩⟨digit⟩⟨more digits⟩ to an integer truncates,
and to 1\exp_stop_f; otherwise. The ⟨more digits⟩ part must be a digit, followed
by something that does not overflow a \int_use:N \__fp_int_eval:w construction. The
only relevant information about this piece is whether it is zero or not.
\cs_new:Npn \__fp_round_s:NNNw #1 #2 #3 #4;
\exp_after:wN \__fp_round:NNN
\exp_after:wN #1
\exp_after:wN #2
\int_value:w \__fp_int_eval:w
\if_int_odd:w \if_meaning:w 0 #1 1 \else: \if_meaning:w 5 #1 1 \fi: \exp_stop_f:
\if_int_compare:w \__fp_int_eval:w #4 > \c_zero_int
  1 + \fi:
\fi:
\fi:
\c_zero_int
\fi:
\fi:
\cs_new:Npn \__fp_round_digit:Nw #1 #2;
\cs_new:Npn \__fp_round_digit:Nw ⟨digit⟩ ⟨intexpr⟩;
This function should always be called within an \int_value:w or \__fp_int_eval:w expansion; it may add an extra \__fp_int_eval:w, which means that the integer or
integer expression should not be ended with a synonym of \relax, but with a semi-colon
for instance.
\cs_new:Npn \__fp_round_digit:Nw #1 #2;
\cs_new:Npn \__fp_round_digit:Nw ⟨digit⟩ ⟨intexpr⟩;
(End definition for \__fp_round_digit:Nw.)
\__fp_round_digit:Nw
This expands to 0\exp_stop_f: or 1\exp_stop_f: after doing the following test.
Starting from a number of the form (final sign)0.(15 digits)(digit) with exactly 15 (non-
all-zero) digits before (digit), subtract from it (final sign)0.0...0(digit), where there
are 16 zeros. If in the current rounding mode the result should be rounded down, then
this function returns 0\exp_stop_f:. Otherwise, i.e., if the result is rounded back to
the first operand, then this function returns 0\exp_stop_f:.

It turns out that this negative “round to nearest” is identical to the positive one.
And this is the default mode.

First check that all arguments are floating point numbers. The \_fpparse\_function\_all\_fp\_o:fnw
functions expect one or two arguments (the second is 0 by default), and the round function
also accepts a third argument (nan by default), which changes #1 from \_fp\_round\_to\_nearest:NNN to one of its analogues.

First check that all arguments are floating point numbers. The \_fpparse\_function\_all\_fp\_o:fnw
functions expect one or two arguments (the second is 0 by default), and the round function
also accepts a third argument (nan by default), which changes #1 from \_fp\_round\_to\_nearest:NNN to one of its analogues.

68.2 The round function
\__fp_round_no_arg_o:Nw

\cs_new:Npn \__fp_round_no_arg_o:Nw #1
\{\cs_if_eq:NNTF #1 \__fp_round_to_nearest:NNN
\{ \__fp_error:nffn \{ num-args \} \{ \__fp_round_name_from_cs:N #1 \} \{ 2 \} \}
\exp_after:wN \c_nan_fp \}
\}

(End definition for \__fp_round_no_arg_o:Nw.)

\__fp_round:Nwww

Having three arguments is only allowed for \texttt{round}, not \texttt{trunc}, \texttt{ceil}, \texttt{floor}, so check for that case. If all is well, construct one of \__fp_round_to_nearest:NNN, \__fp_round_to_nearest_zero:NNN, \__fp_round_to_nearest_ninf:NNN, \__fp_round_to_nearest_-pinf:NNN and act accordingly.

\cs_new:Npn \__fp_round:Nwww #1#2 ; #3 ; \s__fp \__fp_chk:w #4#5#6 ; #7 @
\{\cs_if_eq:NNTF #1 \__fp_round_to_nearest:NNN
\{ \__fp_error:nffn \{ num-args \} \{ \__fp_round_name_from_cs:N #1 \} \{ 2 \} \}
\exp_after:wN \c_nan_fp \}
\}

(End definition for \__fp_round_no_arg_o:Nw.)
If the number of digits to round to is an integer or infinity all is good; if it is \texttt{nan} then just produce a \texttt{nan}; otherwise invalid as we have something like \texttt{round(1,3.14)} where the number of digits is not an integer.

\begin{verbatim}
\__fp_round:Nww \__fp_round:Nwn \__fp_round_normal:NwNNnn \__fp_round_normal:NNwNNnn \__fp_round_normal_end:wwNnn \__fp_round_special:NwwNnn \__fp_round_special_aux:Nw
\end{verbatim}
\cs_new:Npn \__fp_round_normal:NnnwNNnn \#1\#2\#3\#4; \#5\#6
\{\exp_after:wN \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\}
\cs_new:Npn \__fp_round_normal:NnnwNNnn \#1\#2\#3\#4; \#5\#6
\{\exp_after:wN \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\}
\cs_new:Npn \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\{\exp_after:wN \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\}
\cs_new:Npn \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\{\exp_after:wN \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\}
\cs_new:Npn \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\{\exp_after:wN \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\}
\cs_new:Npn \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\{\exp_after:wN \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\}
\cs_new:Npn \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\{\exp_after:wN \__fp_round_normal:NNwNnn \#1 \#2 \#3\#4; \#5\#6
\}
\cs_new:Npn \__fp_round_special_aux:Nw #1#2;
{
\exp_after:wN \__fp_exp_after_o:w \exp:w \exp_end_continue_f:w
\__fp_sanitize:Nw #1#2; {1000}{0000}{0000}{0000};
}

(End definition for \__fp_round:Nww and others.)

(/package)
Chapter 69

l3fp-parse implementation

69.1 Work plan

The task at hand is non-trivial, and some previous failed attempts show that the code leads to unreadable logs, so we had better get it (almost) right the first time. Let us first describe our goal, then discuss the design precisely before writing any code.

In this file at least, a floating point object is a floating point number or tuple. This can be extended to anything that starts with \s__fp or \s__fp_(type) and ends with ; with some internal structure that depends on the (type).

\_\_fp_parse:n \{\fexpr\}

Evaluates the floating point expression and leaves the result in the input stream as a floating point object. This function forms the basis of almost all public l3fp functions. During evaluation, each token is fully f-expanded.

\_\_fp_parse_o:n does the same but expands once after its result.

\TeXhackers note: Registers (integers, toks, etc.) are automatically unpacked, without requiring a function such as \int_use:N. Invalid tokens remaining after f-expansion lead to unrecoverable low-level \TeX errors.

(End definition for \_\_fp_parse:n.)

Floating point expressions are composed of numbers, given in various forms, infix operators, such as +, **, or , (which joins two numbers into a list), and prefix operators, such as the unary -, functions, or opening parentheses. Here is a list of precedences which control the order of evaluation (some distinctions are irrelevant for the order of evaluation, but serve as signals), from the tightest binding to the loosest binding.

16 Function calls.

13/14 Binary ** and ~ (right to left).

12 Unary +, -, ! (right to left).

11 Juxtaposition (implicit *) with no parenthesis.
10 Binary * and /.
9 Binary + and –.
7 Comparisons.
6 Logical and, denoted by &&.
5 Logical or, denoted by ||.
4 Ternary operator ?; piece ?.
3 Ternary operator ?; piece :.
2 Commas.
1 Place where a comma is allowed and generates a tuple.

0 Start and end of the expression.

\int_const:Nn \c__fp_prec_func_int { 16 }
\int_const:Nn \c__fp_prec_hatii_int { 14 }
\int_const:Nn \c__fp_prec_hat_int { 13 }
\int_const:Nn \c__fp_prec_not_int { 12 }
\int_const:Nn \c__fp_prec_juxt_int { 11 }
\int_const:Nn \c__fp_prec_times_int { 10 }
\int_const:Nn \c__fp_prec_plus_int { 9 }
\int_const:Nn \c__fp_prec_comp_int { 7 }
\int_const:Nn \c__fp_prec_and_int { 6 }
\int_const:Nn \c__fp_prec_or_int { 5 }
\int_const:Nn \c__fp_prec_quest_int { 4 }
\int_const:Nn \c__fp_prec_colon_int { 3 }
\int_const:Nn \c__fp_prec_comma_int { 2 }
\int_const:Nn \c__fp_prec_tuple_int { 1 }
\int_const:Nn \c__fp_prec_end_int { 0 }

(End definition for \c__fp_prec_func_int and others.)

69.1.1 Storing results
The main question in parsing expressions expandably is to decide where to put the intermediate results computed for various subexpressions.

One option is to store the values at the start of the expression, and carry them together as the first argument of each macro. However, we want to f-expand tokens one by one in the expression (as \int_eval:n does), and with this approach, expanding the next unread token forces us to jump with \exp_after:wN over every value computed earlier in the expression. With this approach, the run-time grows at least quadratically in the length of the expression, if not as its cube (inserting the \exp_after:wN is tricky and slow).

A second option is to place those values at the end of the expression. Then expanding the next unread token is straightforward, but this still hits a performance issue: for long expressions we would be reaching all the way to the end of the expression at every step of the calculation. The run-time is again quadratic.

A variation of the above attempts to place the intermediate results which appear when computing a parenthesized expression near the closing parenthesis. This still lets
us expand tokens as we go, and avoids performance problems as long as there are enough parentheses. However, it would be better to avoid requiring the closing parenthesis to be present as soon as the corresponding opening parenthesis is read: the closing parenthesis may still be hidden in a macro yet to be expanded.

Hence, we need to go for some fine expansion control: the result is stored before the start!

Let us illustrate this idea in a simple model: adding positive integers which may be resulting from the expansion of macros, or may be values of registers. Assume that one number, say, 12345, has already been found, and that we want to parse the next number. The current status of the code may look as follows.

```latex
\exp_after:wN \add:ww \int_value:w 12345 \exp_after:wN ;
\exp:w \operand:w \langle stuff \rangle
```

One step of expansion expands `\exp_after:wN`, which triggers the primitive `\int_value:w`, which reads the five digits we have already found, 12345. This integer is unfinished, causing the second `\exp_after:wN` to expand, and to trigger the construction `\exp:w`, which expands `\operand:w`, defined to read what follows and make a number out of it, then leave `\exp_end:`, the number, and a semicolon in the input stream. Once `\operand:w` is done expanding, we obtain essentially

```latex
\exp_after:wN \add:ww \int_value:w 12345 ;
\exp:w \exp_end: 333444 ;
```

where in fact `\exp_after:wN` has already been expanded, `\int_value:w` has already seen 12345, and `\exp:w` is still looking for a number. It finds `\exp_end:`, hence expands to nothing. Now, `\int_value:w` sees the `;`, which cannot be part of a number. The expansion stops, and we are left with

```latex
\add:ww 12345 ; 333444 ;
```

which can safely perform the addition by grabbing two arguments delimited by `;`.

If we were to continue parsing the expression, then the following number should also be cleaned up before the next use of a binary operation such as `\add:ww`. Just like `\int_value:w 12345 \exp_after:wN ;` expanded what follows once, we need `\add:ww` to do the calculation, and in the process to expand the following once. This is also true in our real application: all the functions of the form `\_fp_\ldots_o:ww` expand what follows once. This comes at the cost of leaving tokens in the input stack, and we need to be careful not to waste this memory. All of our discussion above is nice but simplistic, as operations should not simply be performed in the order they appear.

69.1.2 Precedence and infix operators

The various operators we will encounter have different precedences, which influence the order of calculations: $1 + 2 \times 3 = 1 + 2 \times 3$ because $\times$ has a higher precedence than $+$. The true analog of our macro `\operand:w` must thus take care of that. When looking for an operand, it needs to perform calculations until reaching an operator which has lower precedence than the one which called `\operand:w`. This means that `\operand:w` must know what the previous binary operator is, or rather, its precedence: we thus rename it `\operand:Nw`. Let us describe as an example how we plan to do the calculation $41 - 2^3 + 4 + 5$. More precisely we describe how to perform the first operation in this expression. Here, we abuse notations: the first argument of `\operand:Nw` should be an integer
constant (\_fp_prec_plus_int, \ldots) equal to the precedence of the given operator, not directly the operator itself.

- Clean up 41 and find -. We call \operand:Nw - to find the second operand.

- Clean up 2 and find ".

- Compare the precedences of - and ". Since the latter is higher, we need to compute the exponentiation. For this, find the second operand with a nested call to \operand:Nw ^.

- Clean up 3 and find *.

- Compare the precedences of - and *. Since the former is higher, \operand:Nw ^ has found the second operand of the exponentiation, which is computed: $2^3 = 8$.

- We now have 41-8*4+5, and \operand:Nw - is still looking for a second operand for the subtraction. Is it 8?

- Compare the precedences of - and *. Since the latter is higher, we are not done with 8. Call \operand:Nw * to find the second operand of the multiplication.

- Clean up 4, and find +.

- Compare the precedences of * and +. Since the former is higher, \operand:Nw * has found the second operand of the multiplication, which is computed: $8 * 4 = 32$.

- We now have 41-32+5, and \operand:Nw - is still looking for a second operand for the subtraction. Is it 32?

- Compare the precedences of - and +. Since they are equal, \operand:Nw - has found the second operand for the subtraction, which is computed: $41 - 32 = 9$.

- We now have 9+5.

The procedure above stops short of performing all computations, but adding a surrounding call to \operand:Nw with a very low precedence ensures that all computations are performed before \operand:Nw is done. Adding a trailing marker with the same very low precedence prevents the surrounding \operand:Nw from going beyond the marker.

The pattern above to find an operand for a given operator, is to find one number and the next operator, then compare precedences to know if the next computation should be done. If it should, then perform it after finding its second operand, and look at the next operator, then compare precedences to know if the next computation should be done. This continues until we find that the next computation should not be done. Then, we stop.

We are now ready to get a bit more technical and describe which of the \l3fp-parse functions correspond to each step above.

First, \_fp_parse_operand:Nw is the \operand:Nw function above, with small modifications due to expansion issues discussed later. We denote by \(\text{precedence}\) the argument of \_fp_parse_operand:Nw, that is, the precedence of the binary operator whose operand we are trying to find. The basic action is to read numbers from the input stream. This is done by \_fp_parse_one:Nw. A first approximation of this function is that it reads one \emph{number}, performing no computation, and finds the following binary \emph{operator}. Then it expands to
expanding the `infix` auxiliary before leaving the above in the input stream.

We now explain the `infix` auxiliaries. We need some flexibility in how we treat the case of equal precedences: most often, the first operation encountered should be performed, such as `1-2-3` being computed as `(1-2)-3`, but `2^3^4` should be evaluated as `2^(3^4)` instead. For this reason, and to support the equivalence between `**` and `^` more easily, each binary operator is converted to a control sequence `\__fp_parse_infix_\langle operator\rangle:N` when it is encountered for the first time. Instead of passing both precedences to a test function to do the comparison steps above, we pass the \langle precedence\rangle (of the earlier operator) to the `infix` auxiliary for the following \langle operator\rangle, to know whether to perform the computation of the \langle operator\rangle. If it should not be performed, the `infix` auxiliary expands to

\begin{verbatim}
@ \use_none:n \__fp_parse_infix_\langle operator\rangle:N
\end{verbatim}

and otherwise it calls `\__fp_parse_operand:Nw` with the precedence of the \langle operator\rangle to find its second operand \langle number\rangle and the next \langle operator2\rangle, and expands to

\begin{verbatim}
\exp_after:wN \__fp_parse_continue:NwN \langle operator\rangle \langle number2\rangle
\exp_after:wN \__fp_parse_infix_\langle operator2\rangle:N
\end{verbatim}

The `infix` function is responsible for comparing precedences, but cannot directly call the computation functions, because the first operand \langle number\rangle is before the `infix` function in the input stream. This is why we stop the expansion here and give control to another function to close the loop.

A definition of `\__fp_parse_operand:Nw` with some of the expansion control removed is

\begin{verbatim}
\exp_after:wN \__fp_parse_continue:NwN \langle precedence\rangle
\exp_after:wN \__fp_parse_one:Nw \langle precedence\rangle
\exp:w \exp_end_continue_f:w
\__fp_parse_one:Nw \langle precedence\rangle
\end{verbatim}

This expands `\__fp_parse_one:Nw` completely, which finds a number, wraps the next \langle operator\rangle into an `infix` function, feeds this function the \langle precedence\rangle, and expands it, yielding either

\begin{verbatim}
\__fp_parse_contiue:NwN \langle precedence\rangle \langle number\rangle @ \use_none:n \__fp_parse_infix_\langle operator\rangle:N
\end{verbatim}

or

\begin{verbatim}
\__fp_parse_contiue:NwN \langle precedence\rangle \langle number\rangle @ \__fp_parse_apply_binary:NwN \langle operator\rangle \langle number2\rangle
\__fp_parse_infix_\langle operator2\rangle:N
\end{verbatim}

The definition of `\__fp_parse_contiue:NwN` is then very simple:

\begin{verbatim}
\cs_new:Npn \__fp_parse_contiue:NwN { \#3 \#1 \#2 @ }
\end{verbatim}
In the first case, \texttt{#3} is \texttt{\use_none:n}, yielding
\begin{verbatim}
\use_none:n (precedence) (number) @
\_\_fp_parse_infix\_\_(operator):N
\end{verbatim}
then \texttt{(number) @ \_\_fp_parse_infix\_\_(operator):N}. In the second case, \texttt{#3} is \texttt{\_\_fp\_parse_apply_binary:NwNwN}, whose role is to compute \texttt{⟨number⟩ ⟨operator⟩ ⟨number2⟩} and to prepare for the next comparison of precedences: first we get
\begin{verbatim}
\_\_fp_parse_apply_binary:NwNwN
⟨precedence⟩ ⟨number⟩ @
⟨operator⟩ ⟨number2⟩
@ \_\_fp_parse_infix\_\_(operator2):N
\end{verbatim}
then
\begin{verbatim}
\exp_after:wN \_\_fp_parse_continue:NwN
\exp_after:wN \exp_after:wN
\_\_fp\_(operator)\_o:ww (number) (number2)
\exp:w \exp_end_continue_f:w
\_\_fp_parse_infix\_\_(operator2):N (precedence)
\end{verbatim}
where \texttt{\_\_fp\_(operator)\_o:ww} computes \texttt{⟨number⟩ ⟨operator⟩ ⟨number2⟩} and expands after the result, thus triggers the comparison of the precedence of the \texttt{⟨operator2⟩} and the \texttt{⟨precedence⟩}, continuing the loop.

We have introduced the most important functions here, and the next few paragraphs we describe various subtleties.

\subsection{Prefix operators, parentheses, and functions}

Prefix operators (unary \texttt{-}, \texttt{+}, \texttt{!}) and parentheses are taken care of by the same mechanism, and functions (\texttt{sin}, \texttt{exp}, etc.) as well. Finding the argument of the unary \texttt{-}, for instance, is very similar to grabbing the second operand of a binary infix operator, with a subtle precedence explained below. Once that operand is found, the operator can be applied to it (for the unary \texttt{-}, this simply flips the sign). A left parenthesis is just a prefix operator with a very low precedence equal to that of the closing parenthesis (which is treated as an infix operator, since it normally appears just after numbers), so that all computations are performed until the closing parenthesis. The prefix operator associated to the left parenthesis does not alter its argument, but it removes the closing parenthesis (with some checks).

Prefix operators are the reason why we only summarily described the function \texttt{\_\_fp\_parse_one:Nw} earlier. This function is responsible for reading in the input stream the first possible \texttt{⟨number⟩} and the next infix \texttt{⟨operator⟩}. If what follows \texttt{\_\_fp\_parse_one:Nw ⟨precedence⟩} is a prefix operator, then we must find the operand of this prefix operator through a nested call to \texttt{\_\_fp\_parse_operand:Nw} with the appropriate precedence, then apply the operator to the operand found to yield the result of \texttt{\_\_fp\_parse_one:Nw}. So far, all is simple.

The unary operators \texttt{+, -, !} complicate things a little bit: \texttt{-3**2} should be \texttt{\(\neg(3^2) = -9\)}, and not \texttt{\((-3)^2 = 9\)}. This would easily be done by giving \texttt{-} a lower precedence, equal to that of the infix \texttt{+} and \texttt{-}. Unfortunately, this fails in cases such as \texttt{3**-2*4}, yielding \texttt{3^{\text{-}\text{2} \times 4}} instead of the correct \texttt{3^{-2} \times 4}. A second attempt would be to call \texttt{\_\_fp\_parse_operand:Nw} with the \texttt{⟨precedence⟩} of the previous operator, but \texttt{0>-2+3} is then
parsed as $0 > -(2+3)$: the addition is performed because it binds more tightly than the comparison which precedes $-$. The correct approach is for a unary $-$ to perform operations whose precedence is greater than both that of the previous operation, and that of the unary $-$ itself. The unary $-$ is given a precedence higher than multiplication and division. This does not lead to any surprising result, since $-(x/y) = (-x)/y$ and similarly for multiplication, and it reduces the number of nested calls to $\_fp\_parse\_operand:Nw$.

Functions are implemented as prefix operators with very high precedence, so that their argument is the first number that can possibly be built.

Note that contrarily to the *infix* functions discussed earlier, the *prefix* functions do perform tests on the previous (*precedence*) to decide whether to find an argument or not, since we know that we need a number, and must never stop there.

### 69.1.4 Numbers and reading tokens one by one

So far, we have glossed over one important point: what is a “number”? A number is typically given in the form $\langle\text{significand}\rangle e\langle\text{exponent}\rangle$, where the $\langle\text{significand}\rangle$ is any non-empty string composed of decimal digits and at most one decimal separator (a period), the exponent “$e\langle\text{exponent}\rangle$” is optional and is composed of an exponent mark $e$ followed by a possibly empty string of signs $+$ or $-$ and a non-empty string of decimal digits. The $\langle\text{significand}\rangle$ can also be an integer, dimension, skip, or muskip variable, in which case dimensions are converted from points (or mu units) to floating points, and the $\langle\text{exponent}\rangle$ can also be an integer variable. Numbers can also be given as floating point variables, or as named constants such as `nan`, `inf` or `pi`. We may add more types in the future.

When $\_\_fp\_parse\_one:Nw$ is looking for a “number”, here is what happens.

- If the next token is a control sequence with the meaning of $\scan_stop$; it can be: $\s__fp$, in which case our job is done, as what follows is an internal floating point number, or $\s__fp\_expr\_mark$, in which case the expression has come to an early end, as we are still looking for a number here, or something else, in which case we consider the control sequence to be a bad variable resulting from c-expansion.

- If the next token is a control sequence with a different meaning, we assume that it is a register, unpack it with $\tex\_the:D$, and use its value (in pt for dimensions and skips, mu for muskips) as the $\langle\text{significand}\rangle$ of a number: we look for an exponent.

- If the next token is a digit, we remove any leading zeros, then read a significand larger than 1 if the next character is a digit, read a significand smaller than 1 if the next character is a period, or we have found a significand equal to 0 otherwise, and look for an exponent.

- If the next token is a letter, we collect more letters until the first non-letter: the resulting word may denote a function such as `asin`, a constant such as `pi` or be unknown. In the first case, we call $\_\_fp\_parse\_operand:Nw$ to find the argument of the function, then apply the function, before declaring that we are done. Otherwise, we are done, either with the value of the constant, or with the value `nan` for unknown words.

- If the next token is anything else, we check whether it is a known prefix operator, in which case $\_\_fp\_parse\_operand:Nw$ finds its operand. If it is not known, then either a number is missing (if the token is a known infix operator) or the token is simply invalid in floating point expressions.
Once a number is found, \texttt{__fp\_parse\_one:Nw} also finds an infix operator. This goes as follows.

- If the next token is a control sequence, it could be the special marker \texttt{s\_\_fp\_expr\_mark}, and otherwise it is a case of juxtaposing numbers, such as \texttt{2\_c\_zero\_int}, with an implied multiplication.

- If the next token is a letter, it is also a case of juxtaposition, as letters cannot be proper infix operators.

- Otherwise (including in the case of digits), if the token is a known infix operator, the appropriate \texttt{\_\_fp\_infix\_⟨operator⟩:N} function is built, and if it does not exist, we complain. In particular, the juxtaposition \texttt{\_c\_zero\_int 2} is disallowed.

In the above, we need to test whether a character token \texttt{#1} is a digit:

\begin{verbatim}
\if_int_compare:w 9 < 1 \token_to_str:N #1 \exp_stop_f:
  is a digit
\else:
  not a digit
\fi:
\end{verbatim}

To exclude 0, replace 9 by 10. The use of \texttt{\token_to_str:N} ensures that a digit with any catcode is detected. To test if a character token is a letter, we need to work with its character code, testing if \texttt{‘#1} lies in [65, 90] (uppercase letters) or [97, 112] (lowercase letters)

\begin{verbatim}
\if_int_compare:w \__fp\_int\_eval:w \ ‘#1 \if_int_compare:w ‘#1 > ‘Z - 32 \fi: ) / 26 = 3 \exp_stop_f:
  is a letter
\else:
  not a letter
\fi:
\end{verbatim}

At all steps, we try to accept all category codes: when \texttt{#1} is kept to be used later, it is almost always converted to category code other through \texttt{\token_to_str:N}. More precisely, catcodes \{3, 6, 7, 8, 11, 12\} should work without trouble, but not \{1, 2, 4, 10, 13\}, and of course \{0, 5, 9\} cannot become tokens.

Floating point expressions should behave as much as possible like \TeX-based integer expressions and dimension expressions. In particular, \texttt{f-expansion} should be performed as the expression is read, token by token, forcing the expansion of protected macros, and ignoring spaces. One advantage of expanding at every step is that restricted expandable functions can then be used in floating point expressions just as they can be in other kinds of expressions. Problematically, spaces stop \texttt{f-expansion}: for instance, the macro \texttt{\_X} below would not be expanded if we simply performed \texttt{f-expansion}.

\begin{verbatim}
\DeclareDocumentCommand {\test} {m} { \fp_eval:n {#1} }
\ExplSyntaxOff
\test { 1 + \_X }
\end{verbatim}

Of course, spaces typically do not appear in a code setting, but may very easily come in document-level input, from which some expressions may come. To avoid this problem, at every step, we do essentially what \texttt{\use:f} would do: take an argument, put it back
in the input stream, then f-expand it. This is not a complete solution, since a macro’s expansion could contain leading spaces which would stop the f-expansion before further macro calls are performed. However, in practice it should be enough: in particular, floating point numbers are correctly expanded to the underlying \_\_f\\_p ... structure. The f-expansion is performed by \_\_f\\_p_parse_expand:w.

\_\_f\\_p_parse_operand:Nw
\exp:w \_\_f\\_p_parse_operand:Nw (preceence) \_\_f\\_p_parse_expand:w

Reads the “...”, performing every computation with a precedence higher than ⟨preceence⟩, then expands to

⟨result⟩ @ \_\_f\\_p_parse_infix_(operation):N ...

where the ⟨operation⟩ is the first operation with a lower precedence, possibly end, and the “...” start just after the ⟨operation⟩.

(End definition for \_\_f\\_p_parse_operand:Nw.)

\_\_f\\_p_parse_infix:+:N
\_\_f\\_p_parse_infix:+:N (preceence) ...

If + has a precedence higher than the ⟨preceence⟩, cleans up a second ⟨operand⟩ and finds the ⟨operation2⟩ which follows, and expands to

@ \_\_f\\_p_parse_apply_binary:NwNwN + ⟨operand⟩ @ \_\_f\\_p_parse_infix_(operation2):N ...

Otherwise expands to

@ \use_none:n \_\_f\\_p_parse_infix_+:+:N ...

A similar function exists for each infix operator.

(End definition for \_\_f\\_p_parse_infix_+:+:N.)

\_\_f\\_p_parse_one:Nw
\_\_f\\_p_parse_one:Nw (preceence) ...

Cleans up one or two operands depending on how the precedence of the next operation compares to the ⟨preceence⟩. If the following ⟨operation⟩ has a precedence higher than ⟨preceence⟩, expands to

⟨operand1⟩ @ \_\_f\\_p_parse_apply_binary:NwNwN ⟨operation⟩ ⟨operand2⟩ @ \_\_f\\_p_parse_infix_(operation2):N ...

and otherwise expands to

⟨operand⟩ @ \use_none:n \_\_f\\_p_parse_infix_(operation):N ...

(End definition for \_\_f\\_p_parse_one:Nw.)

69.2 Main auxiliary functions
69.3 Helpers

\_\_fp_parse.expand:w \exp:w \_\_fp_parse.expand:w \langle \text{tokens} \rangle

This function must always come within a \exp:w expansion. The \langle \text{tokens} \rangle should be the part of the expression that we have not yet read. This requires in particular closing all conditionals properly before expanding.

\cs_new:Npn \_\_fp_parse.expand:w \#1 \{ \exp_end_continue_f:w \#1 \}

(End definition for \_\_fp_parse.expand:w.)

\_fp.parse.return.semicolon:w

This very odd function swaps its position with the following \fi: and removes \_\_fp_parse.expand:w normally responsible for expansion. That turns out to be useful.

\cs_new:Npn \_\_fp_parse.return.semicolon:w \#1 \fi: \_\_fp_parse.expand:w \{ \fi: ; \#1 \}

(End definition for \_\_fp_parse.return.semicolon:w.)


These functions must be called within an \int_value:w or \_\_fp_int_eval:w construction. The first token which follows must be f-expanded prior to calling those functions. The functions read tokens one by one, and output digits into the input stream, until meeting a non-digit, or up to a number of digits equal to their index. The full expansion is

\langle \text{digits} \rangle ; \langle \text{filling 0} \rangle ; \langle \text{length} \rangle

where \langle \text{filling 0} \rangle is a string of zeros such that \langle \text{digits} \rangle \langle \text{filling 0} \rangle has the length given by the index of the function, and \langle \text{length} \rangle is the number of zeros in the \langle \text{filling 0} \rangle string. Each function puts a digit into the input stream and calls the next function, until we find a non-digit. We are careful to pass the tested tokens through \token_to_str:N to normalize their category code.

\cs_set_protected:Npn \_\_fp_tmp:w \#1 \#2 \#3

\{ \cs_new:cpn { \_\_fp_parse_digits_#1 :N } ##1 \}

\{ \if_int_compare:w 9 < 1 \token_to_str:N ##1 \exp_stop_f: \token_to_str:N ##1 \exp_after:wN #2 \exp:w \else: \_\_fp_parse.return.semicolon:w \#3 \#1 \fi: \_\_fp_parse.expand:w \}

(End definition for \_\_fp_parse_digits.vii:N and others.)
69.4 Parsing one number

This function finds one number, and packs the symbol which follows in an \_\_fp_parse_infix... csname. #1 is the previous (precedence), and #2 the first token of the operand. We distinguish four cases: #2 is equal to \scan_stop: in meaning, #2 is a different control sequence, #2 is a digit, and #2 is something else (this last case is split further later). Despite the earlier f-expansion, #2 may still be expandable if it was protected by \exp_not:N, as may happen with the \LaTeX2ε command \protect. Using a well placed \reverse_if:N, this case is sent to \_\_fp_parse_one_fp:NN which deals with it robustly.

\cs_new:Npn \_\_fp_parse_one:Nw \#1 \#2
\begin{verbatim}
  \if_catcode:w \scan_stop: \exp_not:N \#2
    \exp_after:wN \if_meaning:w \exp_not:N \#2 \#2 \else:
      \exp_after:wN \reverse_if:N \fi:
    \if_meaning:w \scan_stop: \#2
      \exp_after:wN \exp_after:wN \exp_after:wN \_\_fp_parse_one_fp:NN
    \else:
      \exp_after:wN \exp_after:wN \exp_after:wN \_\_fp_parse_one_register:NN
    \fi:
  \else:
    \if_int_compare:w 9 < 1 \token_to_str:N \#2 \exp_stop_f:
      \exp_after:wN \exp_after:wN \_\_fp_parse_one_digit:NN
    \else:
      \exp_after:wN \exp_after:wN \_\_fp_parse_one_other:NN
    \fi:
    \#1 \#2
\end{verbatim}

(End definition for \_\_fp_parse_one:Nw.)

\_\_fp_parse_one_fp:NN
\_\_fp_exp_after_expr_mark_f:nw
\_\_fp_exp_after_?. f:nw

This function receives a (precedence) and a control sequence equal to \scan_stop: in meaning. There are three cases.

- \_\_fp starts a floating point number, and we call \_\_fp_exp_after_f:nw, which f-expands after the floating point.
- \_\_fp_expr_mark is a premature end, we call \_\_fp_exp_after_expr_mark-_f:nw, which triggers an fp-early-end error.
- For a control sequence not containing \_\_fp, we call \_\_fp_exp_after_?. f:nw, causing a bad-variable error.

This scheme is extensible: additional types can be added by starting the variables with a scan mark of the form \_\_fp{⟨type⟩} and defining \_\_fp_exp_after_{⟨type⟩}_f:nw. In all cases, we make sure that the second argument of \_\_fp_parse_infix:NN is correctly expanded. A special case only enabled in \LaTeX2ε is that if \protect is encountered then
the error message mentions the control sequence which follows it rather than \texttt{\protect} itself. The test for \LaTeX \texttt{2e} uses \texttt{@unexpandable@protect} rather than \texttt{\protect} because \texttt{\protect} is often \texttt{\scan_stop}: hence “does not exist”.  
\begin{verbatim}
\cs_new:Npn \__fp_parse_one_fp:Nn \#1
\{  
  \__fp_exp_after_any_f:nw
  {  
    \exp_after:wN \__fp_parse_infix:Nn
    \exp_after:wN \#1 \exp:w \__fp_parse_expand:w
  }
\}
\cs_new:Npn \__fp_exp_after_expr_mark_f:nw \#1
\{  
  \int_case:nnF { \exp_after:wN \use_i:nnn \use_none:nnn \#1 }
  {  
    \c__fp_prec_comma_int { }
    \c__fp_prec_tuple_int { }
    \c__fp_prec_end_int 
    {  
      \exp_after:wN \c__fp_empty_tuple_fp
      \exp:w \exp_end_continue_f:w
    }
  }
  {  
    \msg_expandable_error:nnn { fp } { early-end }
    \exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w \#1
  }
\}
\cs_new:cpn { __fp_exp_after_?_f:nw } \#1\#2
\{  
  \msg_expandable_error:nnn { kernel } { bad-variable }
  \#2 \}
\exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w \#1
\cs_set_protected:Npn \__fp_tmp:w \#1
\{  
  \cs_if_exist:NT \#1
  {  
    \cs_gset:cpn { __fp_exp_after_?_f:nw } ##1##2
    {  
      \exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w ##1
    \str_if_eq:nnTF {##2} { \protect }
    {  
      \cs_if_eq:NNTF \#1 #1 \{ \use_i:nn \} \{ \use:n \}
      {  
        \msg_expandable_error:nnn { fp }
        \{ robust-cmd \}
      }
    }
  }
  {  
    \msg_expandable_error:nnn { kernel }
    \{ bad-variable \} \#2
  }
\}
\end{verbatim}
This is called whenever \#2 is a control sequence other than \scan_stop: in meaning. We special-case \wd, \ht, \dp (see later) and otherwise assume that it is a register, but care-fuly unpack it with \tex_the:D within braces. First, we find the exponent following \#2. Then we unpack \#2 with \tex_the:D, and the auxii auxiliary distinguishes integer regis-ters from dimensions/skips from muskips, according to the presence of a period and/or of pt. For integers, simply convert \langle value \rangle e \langle exponent \rangle to a floating point number with \__fp_parse:n (this is somewhat wasteful). For other registers, the decimal rounding provided by \TeX does not accurately represent the binary value that it manipulates, so we extract this binary value as a number of scaled points with \int_value:w \dim_to_decimal_in_sp:n \{ \langle decimal value \rangle pt \}, and use an auxiliary of \dim_to_fp:n, which performs the multiplication by \(2^{-16}\), correctly rounded.
The \texttt{\textbackslash wd}, \texttt{\textbackslash dp}, \texttt{\textbackslash ht} primitives expect an integer argument. We abuse the exponent parser to find the integer argument: simply include the exponent marker \texttt{\textbackslash e}. Once that “exponent” is found, use \texttt{\textbackslash tex_the:D} to find the box dimension and then copy what we did for dimensions.

A digit marks the beginning of an explicit floating point number. Once the number is found, we catch the case of overflow and underflow with \texttt{\textbackslash __fp_sanitize:W},
then \texttt{\_\_fp\_parse\_infix\_after\_operand:NwN} expands \texttt{\_\_fp\_parse\_infix:NN} after the number we find, to wrap the following infix operator as required. Finding the number itself begins by removing leading zeros: further steps are described later.

\begin{verbatim}
\cs_new:Npn \__fp_parse_one_digit:NN #1 { \exp_after:wN \__fp_parse_infix_after_operand:NwN \exp_after:wN #1 \exp:w \exp_end_continue_f:w \exp_after:wN \__fp_sanitize:wN \int_value:w \__fp_int_eval:w 0 \__fp_parse_trim_zeros:N }
\end{verbatim}

\textit{(End definition for \texttt{\_\_fp\_parse\_one\_digit:NN}.)}

\texttt{\_\_fp\_parse\_one\_other:NN} For this function, \texttt{#2} is a character token which is not a digit. If it is an ASCII letter, \texttt{\_\_fp\_parse\_letters:N} beyond this one and give the result to \texttt{\_\_fp\_parse\_\langle\textit{operator}\rangle:Nw}. Otherwise, the character is assumed to be a prefix operator, and we build \texttt{\_\_fp\_parse\_prefix\_\langle\textit{operator}\rangle:Nw}.

\begin{verbatim}
\cs_new:Npn \__fp_parse_one_other:NN #1 #2 { \if_int_compare:w \__fp_int_eval:w ( '#2 \if_int_compare:w '#2 > 'Z - 32 \fi: ) / 26 = 3 \exp_stop_f: \exp_after:wN \__fp_parse_word:Nw \exp_after:wN #1 \exp_after:wN #2 \exp:w \exp_after:wN \__fp_parse_letters:N \exp:w \else: \exp_after:wN \__fp_parse_prefix:NNN \exp_after:wN #1 \exp_after:wN #2 \cs:w \__fp_parse_prefix_ \token_to_str:N #2 :Nw \exp_after:wN \cs_end: \exp:w \fi: \__fp_parse_expand:w }
\end{verbatim}

\textit{(End definition for \texttt{\_\_fp\_parse\_one\_other:NN}.)}

\texttt{\_\_fp\_parse\_word:Nw} \texttt{\_\_fp\_parse\_letters:N} Finding letters is a simple recursion. Once \texttt{\_\_fp\_parse\_letters:N} has done its job, we try to build a control sequence from the word \texttt{#2}. If it is a known word, then the corresponding action is taken, and otherwise, we complain about an unknown word, yield \texttt{\c_nan_fp}, and look for the following infix operator. Note that the unknown word could be a mistyped function as well as a mistyped constant, so there is no way to tell whether to look for arguments; we do not. The standard requires “inf” and “infinity” and “nan” to be recognized regardless of case, but we probably don’t want to allow every \texttt{l3fp} word to have an arbitrary mixture of lower and upper case, so we test and use a differently-named control sequence.
\cs_new:Npn \__fp_parse_word:Nw #1\#2;
{\cs_if_exist_use:cF \__fp_parse_word_#2:N }
\cs_if_exist_use:cF \__fp_parse_caseless_\str_foldcase:n \#2 \#2 }
\msg_expandable_error:nnn { fp } { unknown-fp-word } \#2 \exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w
\__fp_parse_infix:NN
}
#1
\cs_new:Npn \__fp_parse_letters:N #1
{\exp_end_continue_f:w \if_int_compare:w \if_catcode:w \scan_stop: \exp_not:N #1 0 \else:\__fp_int_eval:w ( '#1 \if_int_compare:w #1 > 'Z - 32 \fi: ) / 26 \fi: = 3 \exp_stop_f:
\exp:w \exp_after:wN \__fp_parse_letters:N \exp:w
\__fp_parse_return_semicolon:w #1 \fi:
\__fp_parse_expand:w
\if_meaning:w \scan_stop: #3 \exp_after:wN \__fp_parse_prefix_unknown:NNN \exp:w \fi:
\__fp_parse_prefix:NNN #1 #2 #3
\cs_new:Npn \__fp_parse_prefix:NNN #1 #2 #3
{ #3 #1 \if_int_compare:w \if_catcode:w \scan_stop: \exp_not:N #1 0 \else:\__fp_int_eval:w ( '#1 \if_int_compare:w #1 > 'Z - 32 \fi: ) / 26 \fi: = 3 \exp_stop_f:
\exp:w \exp_after:wN \__fp_parse_letters:N \exp:w
\__fp_parse_return_semicolon:w #1 \fi:
\__fp_parse_expand:w
\if_meaning:w \scan_stop: #3 \exp_after:wN \__fp_parse_prefix_unknown:NNN \exp:w \fi:
\__fp_parse_prefix_unknown:NNN #1 #2 #3
\cs_new:Npn \__fp_parse_prefix_unknown:NNN #1 #2 #3

(End definition for \__fp_parse_word:Nw and \__fp_parse_letters:N.)

\__fp_parse_prefix:NNN \__fp_parse_prefix_unknown:NNN

For this function, \#1 is the previous \langle precedence \rangle, \#2 is the operator just seen, and \#3 is a control sequence which implements the operator if it is a known operator. If this control sequence is \scan_stop:, then the operator is in fact unknown. Either the expression is missing a number there (if the operator is valid as an infix operator), and we put \texttt{nan}, wrapping the infix operator in a 	exttt{csname} as appropriate, or the character is simply invalid in floating point expressions, and we continue looking for a number, starting again from \__fp_parse_one:Nw.
69.4.1 Numbers: trimming leading zeros

Numbers are parsed as follows: first we trim leading zeros, then if the next character is a digit, start reading a significand $\geq 1$ with the set of functions $\__fp_parse_large...$; if it is a period, the significand is $< 1$; and otherwise it is zero. In the second case, trim additional zeros after the period, counting them for an exponent shift $\langle exp_1 \rangle < 0$, then read the significand with the set of functions $\__fp_parse_small...$ Once the significand is read, read the exponent if $e$ is present.

This function expects an already expanded token. It removes any leading zero, then distinguishes three cases: if the first non-zero token is a digit, then call $\__fp_parse_large: N$ (the significand is $\geq 1$); if it is $\ldots$, then continue trimming zeros with $\__fp_parse_strim_zeros: N$; otherwise, our number is exactly zero, and we call $\__fp_parse_zero:$ to take care of that case.

\cs_if_exist:cTF { __fp_parse_infix_ \token_to_str:N \#1 :N } \begin{verbatim}
\__fp_parse_infix:NN #3 #1 
\msg_expandable_error:nnn { fp } { missing-number } {#1} 
\exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w 
\__fp_parse_infix:NN #3 #1 
\msg_expandable_error:nnn { fp } { unknown-symbol } {#1} 
\__fp_parse_one:Nw #3 
\end{verbatim}

(End definition for $\__fp_parse_prefix:NNN$ and $\__fp_parse_prefix_unknown:NNN$.)
23262 \fi:
23263  #1
23264 }

(End definition for \_\_fp_parse_trim_zeros:N and \_\_fp_parse_trim_end:w.)

\_\_fp_parse_strim_zeros:N \_\_fp_parse_strim_end:w
If we have removed all digits until a period (or if the body started with a period), then
enter the “small_trim” loop which outputs −1 for each removed 0. Those −1 are added
to an integer expression waiting for the exponent. If the first non-zero token is a digit,
call \_\_fp_parse_small:N (our significand is smaller than 1), and otherwise, the number
is an exact zero. The name strim stands for “small trim”.

23265 \cs_new:Npn \_\_fp_parse_strim_zeros:N #1
23266 {  
23267 \if:w 0 \exp_not:N #1
23268 - 1
23269 \exp_after:wN \_\_fp_parse_strim_zeros:N \exp:w
23270 \else:
23271 \_\_fp_parse_strim_end:w #1
23272 \fi:
23273 \_\_fp_parse_expand:w
23274 }
23275 \cs_new:Npn \_\_fp_parse_strim_end:w #1 \fi: \_\_fp_parse_expand:w
23276 {  
23277 \fi:
23278 \if_int_compare:w 9 < 1 \token_to_str:N #1 \exp_stop_f:
23279 \exp_after:wN \_\_fp_parse_small:N
23280 \else:
23281 \exp_after:wN \_\_fp_parse_zero:
23282 \fi:
23283 #1
23284 }

(End definition for \_\_fp_parse_strim_zeros:N and \_\_fp_parse_strim_end:w.)

\_\_fp_parse_zero:
After reading a significand of 0, find any exponent, then put a sign of 1 for \_\_fp_-
sanitize:wN, which removes everything and leaves an exact zero.

23285 \cs_new:Npn \_\_fp_parse_zero:
23286 {  
23287 \exp_after:wN ; \exp_after:wN 1
23288 \int_value:w \_\_fp_parse_exponent:N
23289 }

(End definition for \_\_fp_parse_zero:)  

69.4.2 Number: small significand
\_\_fp_parse_small:N
This function is called after we have passed the decimal separator and removed all leading
zeros from the significand. It is followed by a non-zero digit (with any catcode). The
goal is to read up to 16 digits. But we can’t do that all at once, because \texttt{int.value:w}
(which allows us to collect digits and continue expanding) can only go up to 9 digits.
Hence we grab digits in two steps of 8 digits. Since \#1 is a digit, read seven more digits
using \_\_fp_parse_digits_vii:N. The small_leading auxiliary leaves those digits in
the \texttt{int.value:w}, and grabs some more, or stops if there are no more digits. Then the
pack-leading auxiliary puts the various parts in the appropriate order for the processing further up.

\cs_new:Npn \__fp_parse_small:N #1
\exp_after:wN \__fp_parse_pack_leading:NNNNNww
\int_value:w \__fp_int_eval:w 1 \token_to_str:N #1
\exp_after:wN \__fp_parse_small_leading:wwNN
\int_value:w 1
\exp_after:wN \__fp_parse_digits_vii:N
\exp:w \__fp_parse_expand:w
\endinput

(End definition for \__fp_parse_small:N.)

\__fp_parse_small_leading:wwNN 1 \__fp_parse_small_leading:wwNN 1 (digits) ; (zeros) ; (number of zeros)

We leave \langle digits \rangle (\langle zeros \rangle) in the input stream: the functions used to grab digits are such that this constitutes digits 1 through 8 of the significand. Then prepare to pack 8 more digits, with an exponent shift of zero (this shift is used in the case of a large significand). If \#4 \langle next token \rangle is a digit, leave it behind for the packing function, and read 6 more digits to reach a total of 15 digits: further digits are involved in the rounding. Otherwise put 8 zeros in to complete the significand, then look for an exponent.

\cs_new:Npn \__fp_parse_small_leading:wwNN 1 #1 ; #2; #3 #4
\exp_after:wN \__fp_parse_pack_trailing:NNNNNww
\exp_after:wN 0
\int_value:w \__fp_int_eval:w 1
\if_int_compare:w 9 < 1 \token_to_str:N #4 \exp_stop_f:
\token_to_str:N #4
\exp_after:wN \__fp_parse_small_trailing:wwNN
\int_value:w 1
\exp_after:wN \__fp_parse_digits_vi:N
\exp:w
\else:
0000 0000 \__fp_parse_exponent:Nw #4
\fi:
\__fp_parse_expand:w
\endinput

(End definition for \__fp_parse_small_leading:wwNN.)

\__fp_parse_small_trailing:wwNN 1 \__fp_parse_small_trailing:wwNN 1 (digits) ; (zeros) ; (number of zeros)
\langle next token \rangle

Leave digits 10 to 15 (arguments \#1 and \#2) in the input stream. If the \langle next token \rangle is a digit, it is the 16th digit, we keep it, then the small_round auxiliary considers this digit and all further digits to perform the rounding: the function expands to nothing, to +0 or to +1. Otherwise, there is no 16-th digit, so we put a 0, and look for an exponent.
Those functions are expanded after all the digits are found, we took care of the rounding, as well as the exponent. The last argument is the exponent. The previous five arguments are 8 digits which we pack in groups of 4, and the argument before that is 1, except in the rare case where rounding lead to a carry, in which case the argument is 2. The trailing function has an exponent shift as its first argument, which we add to the exponent found in the e... syntax. If the trailing digits cause a carry, the integer expression for the leading digits is incremented (+1 in the code below). If the leading digits propagate this carry all the way up, the function \_fp_parse_pack_carry:w increments the exponent, and changes the significand from 0000... to 1000.... this is simple because such a carry can only occur to give rise to a power of 10.

\cs_new:Npn \_fp_parse_pack_trailing:NNNNNNww #1 #2 #3#4#5#6 #7; #8 ;
\cs_new:Npn \_fp_parse_pack_leading:NNNNNww #1 #2#3#4#5 #6; #7;,
\cs_new:Npn \_fp_parse_pack_carry:w \fi: ; 0 #1
\cs_new:Npn \_fp_parse_large:N #1
\{ \exp_after:wN \_fp_parse_large_leading:wwNN \int_value:w 1 \token_to_str:N #1
\exp:w \_fp_parse_digits_vii:N \exp:w \_fp_parse_expand:w
\}
\{ \exp_after:wN \_fp_parse_large_leading:wwNN \int_value:w 1 \token_to_str:N #1
\exp:w \_fp_parse_digits_vii:N \exp:w \_fp_parse_expand:w
\}
\{ \exp_after:wN \_fp_parse_large_leading:wwNN \int_value:w 1 \token_to_str:N #1
\exp:w \_fp_parse_digits_vii:N \exp:w \_fp_parse_expand:w
\}
\{ \exp_after:wN \_fp_parse_large_leading:wwNN \int_value:w 1 \token_to_str:N #1
\exp:w \_fp_parse_digits_vii:N \exp:w \_fp_parse_expand:w
\}
\{ \exp_after:wN \_fp_parse_large_leading:wwNN \int_value:w 1 \token_to_str:N #1
\exp:w \_fp_parse_digits_vii:N \exp:w \_fp_parse_expand:w
\}
\{ \exp_after:wN \_fp_parse_large_leading:wwNN \int_value:w 1 \token_to_str:N #1
\exp:w \_fp_parse_digits_vii:N \exp:w \_fp_parse_expand:w
\}

69.4.3 Number: large significand

Parsing a significand larger than 1 is a little bit more difficult than parsing small significands. We need to count the number of digits before the decimal separator, and add that to the final exponent. We also need to test for the presence of a dot each time we run out of digits, and branch to the appropriate parse_small function in those cases.

\_fp_parse_large:N
This function is followed by the first non-zero digit of a “large” significand (\geq 1). It is called within an integer expression for the exponent. Grab up to 7 more digits, for a total of 8 digits.

\cs_new:Npn \_fp_parse_large:N #1
\{ 
\exp_after:wN \_fp_parse_large_leading:wwNN
\int_value:w 1 \token_to_str:N #1
\exp:w \_fp_parse_digits_vii:N
\exp:w \_fp_parse_expand:w
\}

1011
We shift the exponent by the number of digits in #1, namely the target number, 8, minus the \(\langle\text{number of zeros}\rangle\) (number of digits missing). Then prepare to pack the 8 first digits. If the \(\langle\text{next token}\rangle\) is a digit, read up to 6 more digits (digits 10 to 15). If it is a period, try to grab the end of our 8 first digits, branching to the small functions since the number of digit does not affect the exponent anymore. Finally, if this is the end of the significand, insert the \(\langle\text{zeros}\rangle\) to complete the 8 first digits, insert 8 more, and look for an exponent.

\begin{verbatim}
\cs_new:Npn \__fp_parse_large_leading:wwNN 1 #1 ; #2; #3 #4
\{
+ \c__fp_half_prec_int - #3
\int_value:w \__fp_int_eval:w 1 #1
\if_int_compare:w 9 < 1 \token_to_str:N #4 \exp_stop_f:
\exp_after:wN \__fp_parse_pack_leading:NNNNNww
\int_value:w 1 \token_to_str:N #4
\exp_after:wN \__fp_parse_large_trailing:wwNN
\exp_after:wN \__fp_parse_digits_vi:N
\exp:w
\else:
\if:w . \exp_not:N #4
\exp_after:wN \__fp_parse_small_leading:wwNN
\int_value:w 1 \token_to_str:N #4
\cs:w
\__fp_parse_digits_
\__fp_int_to_roman:w #3
: N \exp_after:wN
\cs_end:
\exp:w
\else:
#2
\exp_after:wN \__fp_parse_pack_trailing:NNNNNww
\exp_after:wN 0
\int_value:w 1 0000 0000
\__fp_parse_exponent:Nw #4
\fi:
\fi:
\__fp_parse_expand:w
\}
\end{verbatim}

\begin{verbatim}
\__fp_parse_large_trailing:wwNN
\__fp_parse_large_trailing:wwNN 1 \langle\text{digits}\rangle ; \langle\text{zeros}\rangle ; \langle\text{number of zeros}\rangle
\end{verbatim}

We keep the \(\langle\text{digits}\rangle\) and this 16-th digit, and find how this should be rounded using \(\__fp_parse_large_round:NN\). Otherwise, the exponent shift is the number of \(\langle\text{digits}\rangle\), 7 minus the \(\langle\text{number of zeros}\rangle\), and we test for a decimal point. This case happens in 123451234512345.67 with exactly 15 digits before the decimal separator. Then branch to the appropriate small auxiliary, grabbing a few more digits to complement the digits.
we already grabbed. Finally, if this is truly the end of the significand, look for an exponent after using the \textlangle zeros\rangle and providing a 16-th digit of 0.

\begin{Verbatim}
\cs_new:Npn \_\_fp_parse_large_trailing:wwNN 1 \#1 ; \#2; \#3 \#4
\{
\if_int_compare:w 9 < 1 \token_to_str:N \#4 \exp_stop_f:
\exp_after:wN \_\_fp_parse_pack_trailing:NNNNNNww
\exp_after:wN \c__fp_half_prec_int
\int_value:w \_\_fp_int_eval:w 1 \#1 \token_to_str:N \#4
\exp_after:wN \_\_fp_parse_large_round:NN
\exp_after:wN \#4
\exp:w
\else:
\exp_after:wN \_\_fp_parse_pack_trailing:NNNNNNww
\int_value:w \_\_fp_int_eval:w 7 - \#3 \exp_stop_f:
\if:w . \exp_not:N \#4
\exp_after:wN \_\_fp_parse_small_trailing:wwNN
\int_value:w 1
\cs:w
\_\_fp_parse_digits_
\_\_fp_int_to_roman:w \#3
:N \exp_after:wN
\cs: end:w
\exp:w
\else:
\#2 0 \_\_fp_parse_exponent:Nw \#4
\fi:
\_\_fp_parse_expand:w
\}
\end{Verbatim}

(End definition for \_\_fp_parse_large_trailing:wwNN.)

\subsection*{69.4.4 Number: beyond 16 digits, rounding}

This loop is called when rounding a number (whether the mantissa is small or large). It should appear in an integer expression. This function reads digits one by one, until reaching a non-digit, and adds 1 to the integer expression for each digit. If all digits found are 0, the function ends the expression by ;0, otherwise by ;1. This is done by switching the loop to \texttt{round\_up} at the first non-zero digit, thus we avoid to test whether digits are 0 or not once we see a first non-zero digit.
After the loop \__fp_parse_round_loop:N, this function fetches an exponent with \__fp_parse_exponent:N, and combines it with the number of digits counted by \__fp_parse_round_loop:N. At the same time, the result 0 or 1 is added to the surrounding integer expression.

\cs_new:Npn \__fp_parse_round_after:wN #1; #2
\int_value:w \__fp_int_eval:w #1 + \__fp_parse_exponent:N
\exp_after:wN \__fp_parse_round_after:wN
\int_value:w \__fp_int_eval:w 0 * \__fp_int_eval:w 0
\exp_after:wN \__fp_parse_round_loop:N
\exp:w
\else:
\__fp_parse_exponent:Nw #2
\fi:
\__fp_parse_round_after:wN
\begin{align*}
\_fp_parse_round_after:wN & \text{\( \star \)} \\
\_fp_parse_round_after:wN & \text{\( \star \)} \\
\_fp_parse_small_round:NN & \text{\( \star \)} \\
\_fp_parse_round_after:wN & \text{\( \star \)}
\end{align*}

Here, \#1 is the digit that we are currently rounding (we only care whether it is even or odd). If \#2 is not a digit, then fetch an exponent and expand to ∥\text{exponent}∥ only. Otherwise, we expand to +0 or +1, then ∥\text{exponent}∥. To decide which, call ∥\_fp-round_s:NNw∥ to know whether to round up, giving it as arguments a sign 0 (all explicit numbers are positive), the digit \#1 to round, the first following digit \#2, and either +0 or +1 depending on whether the following digits are all zero or not. This last argument is obtained by ∥\_fp-parse_round_loop:N∥, whose number of digits we discard by multiplying it by 0. The exponent which follows the number is also fetched by ∥\_fp-parse_round_after:wN∥.
Large numbers are harder to round, as there may be a period in the way. Again, \#1 is
the digit that we are currently rounding (we only care whether it is even or odd). If there
are no more digits (\#2 is not a digit), then we must test for a period: if there is one,
then switch to the rounding function for small significands, otherwise fetch an exponent.
If there are more digits (\#2 is a digit), then round, checking with \_\_fp_parse_round-\loop:N if all further digits vanish, or some are non-zero. This loop is not enough, as it is
stopped by a period. After the loop, the aux function tests for a period: if it is present,
then we must continue looking for digits, this time discarding the number of digits we
find.

\cs_new:Npn \_\_fp_parse_large_round:NN #1#2
\{\fi:
\_\_fp_parse_exp:w
\}

(End definition for \_\_fp_parse_small_round:NN and \_\_fp_parse_round_after:wn.)
69.4.5 Number: finding the exponent

Expansion is a little bit tricky here, in part because we accept input where multiplication is implicit.

\__fp_parse:n { 3.2 erf(0.1) }
\__fp_parse:n { 3.2 e\l_my_int }
\__fp_parse:n { 3.2 \c_pi_fp }

The first case indicates that just looking one character ahead for an “e” is not enough, since we would mistake the function erf for an exponent of “rf”. An alternative would be to look two tokens ahead and check if what follows is a sign or a digit, considering in that case that we must be finding an exponent. But taking care of the second case requires that we unpack registers after e. However, blindly expanding the two tokens ahead completely would break the third example (unpacking is even worse). Indeed, in the course of reading 3.2, \c_pi_fp is expanded to \s__fp \__fp_chk:w 1 0 {-1} {3141} \cdots; and \s__fp stops the expansion. Expanding two tokens ahead would then force the expansion of \__fp_chk:w (despite it being protected), and that function tries to produce an error.

What can we do? Really, the reason why this last case breaks is that just as \TeX{} does, we should read ahead as little as possible. Here, the only case where there may be an exponent is if the first token ahead is e. Then we expand (and possibly unpack) the second token.

\__fp_parse_exponent:N
\__fp_parse_exponent_aux:NN

This function should be called within an \int_value:w expansion (or within an integer expression). It leaves digits of the exponent behind it in the input stream, and terminates the expansion with a semicolon. If there is no e (or E), leave an exponent of 0. If there is an e or E, expand the next token to run some tests on it. The first rough test is that if the character code of #1 is greater than that of 9 (largest code valid for an exponent, less than any code valid for an identifier), there was in fact no exponent; otherwise, we search for the sign of the exponent.
\if:w e \if:w E \exp_not:N #1 e \else: \exp_not:N #1 \fi:
\exp_after:wN \__fp_parse_exponent_aux:NN
\exp_after:wN #1
\exp:w
\else:
  0 \__fp_parse_return_semicolon:w #1
\fi:
\__fp_parse_expand:w
\cs_new:Npn \__fp_parse_exponent_aux:NN #1#2
{\if_int_compare:w \if_catcode:w \scan_stop: #2 \exp_not:N \exp_after:wN ; \exp_after:wN #1 \else: '#2 \fi: > '9 \exp_stop_f:
  0 \exp_after:wN ; \exp_after:wN #1
\else:
  \exp_after:wN \__fp_parse_exponent_sign:N
\fi:
#2}
(End definition for \__fp_parse_exponent:N and \__fp_parse_exponent_aux:NN.)
\__fp_parse_exponent_sign:N
Read signs one by one (if there is any).
\cs_new:Npn \__fp_parse_exponent_sign:N #1
{\if:w + \if:w - \exp_not:N #1 + \token_to_str:N #1 \exp_after:wN \__fp_parse_exponent_sign:N
\exp:w \exp_after:wN \__fp_parse_exponent_digits:N
\exp:w \exp_after:wN \__fp_parse_exponent_body:N \exp_after:wN \__fp_parse_exponent_keep:NTF #1
{ \__fp_parse_return_semicolon:w #1 }
{ \exp:w
  \exp:w
}
\fi:
\__fp_parse_expand:w
(End definition for \__fp_parse_exponent_sign:N.)
\__fp_parse_exponent_body:N
An exponent can be an explicit integer (most common case), or various other things
(most of which are invalid).
\cs_new:Npn \__fp_parse_exponent_body:N #1
{\if_int_compare:w 9 < 1 \token_to_str:N #1 \exp_stop_f:
 \token_to_str:N #1
\exp_after:wN \__fp_parse_exponent_digits:N
\exp:w
\else:
 \__fp_parse_exponent_keep:NTF #1
{ \__fp_parse_return_semicolon:w #1 }
{ \exp:w
  \exp:w
}
\fi:
\__fp_parse_expand:w
\_\_fp\_parse\_exponent\_digits:N

Read digits one by one, and leave them behind in the input stream. When finding a non-digit, stop, and insert a semicolon. Note that we do not check for overflow of the exponent, hence there can be a \TeX\ error. It is mostly harmless, except when parsing 0e9876543210, which should be a valid representation of 0, but is not.

\cs_new:Npn \__fp_parse_exponent_digits:N #1
\begin{verbatim}
\if_int_compare:w 9 < 1 \token_to_str:N #1 \exp_stop_f: \exp:w \\
\token_to_str:N #1 \exp_after:wN \__fp_parse_exponent_digits:N \else:w \__fp_parse_return_semicolon:w #1 \fi:w \__fp_parse_expand:w
\end{verbatim}

(End definition for \__fp_parse_exponent_digits:N.)

\__fp_parse_exponent_keep:N

This is the last building block for parsing exponents. The argument \#1 is already fully expanded, and neither + nor - nor a digit. It can be:

- \s__fp, marking the start of an internal floating point, invalid here;
- another control sequence equal to \relax, probably a bad variable;
- a register: in this case we make sure that it is an integer register, not a dimension;
- a character other than +, - or digits, again, an error.

\prg_new_conditional:Npnn \__fp_parse_exponent_keep:N #1 { TF }
\begin{verbatim}
\if_catcode:w \scan_stop: \exp_not:N #1 \if_meaning:w \scan_stop: #1 \if:w 0 \__fp_str_if_eq:nn { \s__fp } { \exp_not:N #1 } 0 \msg_expandable_error:nnn { fp } { after-e } { floating-point- } \prg_return_true:w \else:w \msg_expandable_error:nnn { kernel } { bad-variable } \#1 \prg_return_false:w \fi:w \else:w \if:w 0 \__fp_str_if_eq:nn { \int_value:w #1 } { \tex_the:D #1 } \int_value:w #1 \else:w 0 \msg_expandable_error:nnn { fp } { after-e } { dimension-#1 } \fi:w \prg_return_false:w
\end{verbatim}

1018
69.5 Constants, functions and prefix operators

69.5.1 Prefix operators

A unary + does nothing: we should continue looking for a number.
\begin{verbatim}
\cs_new_eq:Ncn { \__fp_parse_prefix_+:Nw } \__fp_parse_one:Nw
\end{verbatim}

Here, \#1 is a precedence, \#2 is some extra data used by some functions, \#3 is \textit{e.g.}, \__fp_sin_o:w, and expands once after the calculation, \#4 is the operand, and \#5 is a \__fp_parse_infix_...:N function. We feed the data \#2, and the argument \#4, to the function \#3, which expands \exp:w thus the infix function \#5.
\begin{verbatim}
\cs_new:Npn \__fp_parse_apply_function:NNNwN #1#2#3#4@#5
\__fp_parse_apply_unary_chk:NwNw #4 @ ; . \s__fp_stop
\__fp_parse_apply_unary_type:NNN #3 #2 #4 @ \exp:w \exp_end_continue_f:w #5 #1
\end{verbatim}

In contrast to \__fp_parse_apply_function:NNNwN, this checks that the operand \#4 is a single argument (namely there is a single ;). We use the fact that any floating point starts with a “safe” token like \s__fp. If there is no argument produce the fp-no-arg error; if there are at least two produce fp-multi-arg. For the error message extract the mathematical function name (such as sin) from the expl3 function that computes it, such as \__fp_sin_o:w.

In addition, since there is a single argument we can dispatch on type and check that the resulting function exists. This catches things like \texttt{sin((1,2))} where it does not make sense to take the sine of a tuple.
\begin{verbatim}
\cs_new:Npn \__fp_parse_apply_unary:NNNwN #1#2\s__fp_stop
{ \__fp_parse_apply_unary_chk:NwNw \__fp_parse_apply_unary_type:NNN
\__fp_parse_apply_unary_error:Nw \exp:w \exp_end_continue_f:w #5 #1 }
\end{verbatim}

(End definition for \__fp_parse_apply_function:NNNwN.)
The unary - and boolean not are harder: we parse the operand using a precedence equal to the maximum of the previous precedence \#1 and the precedence \__fp_prec_not_int of the unary operator, then call the appropriate \__fp⟨operation⟩_o:w function, where the \langle operation⟩ is set_sign or not.

Numbers which start with a decimal separator (a period) end up here. Of course, we do not look for an operand, but for the rest of the number. This function is very similar to \__fp_parse_one_digit:NN but calls \__fp_parse_strim_zeros:N to trim zeros after the decimal point, rather than the trim_zeros function for zeros before the decimal point.
The left parenthesis is treated as a unary prefix operator because it appears in exactly the same settings. If the previous precedence is \texttt{\textbackslash c\_fp\_prec\_func\_int} we are parsing arguments of a function and commas should not build tuples; otherwise commas should build tuples. We distinguish these cases by precedence: \texttt{\textbackslash c\_fp\_prec\_comma\_int} for the case of arguments, \texttt{\textbackslash c\_fp\_prec\_tuple\_int} for the case of tuples. Once the operand is found, the \texttt{lparen\_after} auxiliary makes sure that there was a closing parenthesis (otherwise it complains), and leaves in the input stream an operand, fetching the following infix operator.

The right parenthesis can appear as a prefix in two similar cases: in an empty tuple or tuple ending with a comma, or in an empty argument list or argument list ending with a comma, such as in \texttt{max(1,2,)} or in \texttt{rand()}.
\__fp_parse_word_inf:N \__fp_parse_word_nan:N \__fp_parse_word_pi:N \__fp_parse_word_true:N \__fp_parse_word_false:N

Some words correspond to constant floating points. The floating point constant is left as a result of \__fp_parse_one:Nw after expanding \__fp_parse_infix:NN.

\cs_set_protected:Npn \__fp_tmp:w #1 #2 {
  \cs_new:cpn { __fp_parse_word_#1:N } { \exp_after:wN #2 \exp:w \exp_end_continue_f:w \__fp_parse_infix:NN }
\__fp_tmp:w { inf } \c_inf_fp \__fp_tmp:w { nan } \c_nan_fp \__fp_tmp:w { pi } \c_pi_fp \__fp_tmp:w { deg } \c_one_degree_fp \__fp_tmp:w { true } \c_one_fp \__fp_tmp:w { false } \c_zero_fp

(End definition for \__fp_parse_word_inf:N and others.)

\__fp_parse_caseless_inf:N \__fp_parse_caseless_infinity:N \__fp_parse_caseless_nan:N

Copies of \__fp_parse_word_...:N commands, to allow arbitrary case as mandated by the standard.

\cs_new_eq:NN \__fp_parse_caseless_inf:N \__fp_parse_word_inf:N \cs_new_eq:NN \__fp_parse_caseless_infinity:N \__fp_parse_word_inf:N \cs_new_eq:NN \__fp_parse_caseless_nan:N \__fp_parse_word_nan:N

(End definition for \__fp_parse_caseless_inf:N, \__fp_parse_caseless_infinity:N, and \__fp_parse_caseless_nan:N.)

\__fp_parse_word_pt:N \__fp_parse_word_in:N \__fp_parse_word_cm:N \__fp_parse_word_mm:N \__fp_parse_word_dd:N \__fp_parse_word_cc:N \__fp_parse_word_nd:N \__fp_parse_word_nc:N \__fp_parse_word_bp:N \__fp_parse_word_sp:N

Dimension units are also floating point constants but their value is not stored as a floating point constant. We give the values explicitly here.

\cs_set_protected:Npn \__fp_tmp:w #1 #2 {
  \cs_new:cpn { __fp_parse_word_#1:N } {
    \__fp_exp_after_f:nw { \__fp_parse_infix:NN }
    \s__fp \__fp_chk:w 10 #2 ;
  }
\__fp_tmp:w { pt } \__fp_pt
\__fp_tmp:w { in } \__fp_in
\__fp_tmp:w { cm } \__fp_cm
\__fp_tmp:w { mm } \__fp_mm
\__fp_tmp:w { dd } \__fp_dd
\__fp_tmp:w { cc } \__fp_cc
\__fp_tmp:w { nd } \__fp_nd
\__fp_tmp:w { nc } \__fp_nc
\__fp_tmp:w { bp } \__fp_bp
\__fp_tmp:w { sp } \__fp_sp

(End definition for \__fp_parse_prefix:):Nw)
The font-dependent units \texttt{em} and \texttt{ex} must be evaluated on the fly. We reuse an auxiliary of \texttt{dim_to_fp:n}.

\begin{verbatim}
\tl_map_inline:nn { {em} {ex} }
{ \cs_new:cpn { __fp_parse_word_#1:N } { \exp_after:wN \__fp_from_dim_test:ww \exp_after:wN 0 \exp_after:wN , \int_value:w \dim_to_decimal_in_sp:n { 1 #1 } \exp_after:wN ; \exp:w \exp_end_continue_f:w \__fp_parse_infix:NN }
\}
\end{verbatim}

(End definition for \texttt{\__fp_parse_word_em:N} and \texttt{\__fp_parse_word_ex:N}.)

\section*{69.5.3 Functions}

\begin{verbatim}
\cs_new:Npn \__fp_parse_unary_function:NNN #1#2#3
{ \exp_after:wN \__fp_parse_apply_unary:NNNwN \exp_after:wN #3 \exp_after:wN #2 \exp_after:wN #1 \exp:w \__fp_parse_operand:Nw \c__fp_prec_func_int \__fp_parse_expand:w }
\cs_new:Npn \__fp_parse_function:NNN #1#2#3
{ \exp_after:wN \__fp_parse_apply_function:NNNwN \exp_after:wN #3 \exp_after:wN #2 \exp_after:wN #1 \exp:w \__fp_parse_operand:Nw \c__fp_prec_func_int \__fp_parse_expand:w }
\end{verbatim}

(End definition for \texttt{\__fp_parse_unary_function:NNN} and \texttt{\__fp_parse_function:NNN}.)
69.6 Main functions

Start an \exp:w expansion so that \_\_fp_parse:n expands in two steps. The \_\_fp_parse_operand:Nw function performs computations until reaching an operation with precedence \_\_fp_prec_end_int or less, namely, the end of the expression. The marker \_\_fp_expr_mark indicates that the next token is an already parsed version of an infix operator, and \_\_fp_parse_infix_end:N has infinitely negative precedence. Finally, clean up a (well-defined) set of extra tokens and stop the initial expansion with \exp_end::

\begin{verbatim}
\cs_new:Npn \__fp_parse:n #1
{\exp:w \exp_after:wN \__fp_parse_after:ww \exp:w \__fp_parse_operand:Nw \c__fp_prec_end_int \__fp_parse_expand:w #1 \_\_fp_expr_mark \_\_fp_parse_infix_end:N \_\_fp_expr_stop \exp_end:}
\cs_new:Npn \__fp_parse_after:ww #1@ \__fp_parse_infix_end:N \_\_fp_expr_stop #2 { #2 #1 }
\cs_new:Npn \__fp_parse_o:n #1
{\exp:w \exp_after:wN \__fp_parse_operand:Nw \c__fp_prec_end_int \__fp_parse_expand:w #1 \_\_fp_expr_mark \_\_fp_parse_infix_end:N \exp_end_continue_f:w \_\_fp_parse_after_any_f:nw { \exp_after:wN \exp_stop_f: }
}
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_parse_operand:Nw #1
{\exp_end_continue_f:w \exp_after:wN \__fp_parse_continue:NwN \exp_after:wN #1 \exp:w \exp_end_continue_f:w \exp_after:wN \__fp_parse_one:Nw \exp_after:wN #1 \exp:w }
\cs_new:Npn \__fp_parse_continue:NwN #1 #2 @ #3 { #3 #1 #2 @ }
\end{verbatim}

This is just a shorthand which sets up both \_\_fp_parse_continue:Nw and \_\_fp_parse_one:N with the same precedence. Note the trailing \exp:w.

\begin{verbatim}
\cs_new:Npn \__fp_parse_operand:Nw #1
{\exp_end_continue_f:w \exp_after:wN \__fp_parse_continue:NwN \exp_after:wN #1 \exp:w \exp_end_continue_f:w \exp_after:wN \__fp_parse_one:Nw \exp_after:wN #1 \exp:w }
\end{verbatim}

(End definition for \_\_fp_parse:n, \_\_fp_parse_o:n, and \_\_fp_parse_after:ww.)
Receives \( \langle \text{precedence} \rangle \langle \text{operand}_1 \rangle \circ \langle \text{operation} \rangle \langle \text{operand}_2 \rangle \circ \langle \text{infix command} \rangle \). Builds the appropriate call to the \( \langle \text{operation} \rangle \#3 \), dispatching on both types. If the resulting control sequence does not exist, the operation is not allowed.

This is redefined in \texttt{13fp-extras}.

\begin{verbatim}
\cs_new:Npn \__fp_parse_apply_binary:NwNwN #1 #2#3@ #4 #5#6@ #7
\exp_after:wN \__fp_parse_continue:NwN
\exp_after:wN \__fp_parse_apply_binary_chk:NN
\exp:w \exp_end_continue_f:w
\exp_after:wN \__fp_parse_apply_binary_error:NNN
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_parse_apply_binary_chk:NN #1#2
\if_meaning:w \scan_stop: #1
\__fp_parse_apply_binary_error:NNN #2
\fi:
#1
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_parse_apply_binary_error:NNN #1#2#3
#2
\__fp_invalid_operation_o:Nww #1
\end{verbatim}

(End definition for \texttt{\__fp_parse_apply_binary:NwNwN}, \texttt{\__fp_parse_apply_binary_chk:NN}, and \texttt{\__fp_parse_apply_binary_error:NNN}.)

\begin{verbatim}
\__fp_binary_type_o:Nww \__fp_binary_rev_type_o:Nww
\end{verbatim}

Applies the operator \#1 to its two arguments, dispatching according to their types, and expands once after the result. The \texttt{rev} version swaps its arguments before doing this.

\begin{verbatim}
\cs_new:Npn \__fp_binary_type_o:Nww \__fp_binary_rev_type_o:Nww #1 #2#3 ; #4
\exp_after:wN \__fp_binary_type_o:Nww
\__fp_type_from_scan:N #1
\__fp_type_from_scan:N #4
\exp:w \exp_end_continue_f:w #4 #2
\__fp_binary_rev_type_o:Nww #1 #2#3 ; #4#5 ;
\end{verbatim}

1025
69.7 Infix operators

(End definition for \__fp_binary_type_o:Nww and \__fp_binary_rev_type_o:Nww.)
\_fp_parse_infix_after_paren:NN

Variant of \_fp_parse_infix:NN for use after a closing parenthesis. The only difference is that \_fp_parse_infix_juxt:N is replaced by \_fp_parse_infix_mul:N.

\cs_new:Npn \_fp_parse_infix_after_paren:NN #1 #2
\begin{verbatim}
  \if_catcode:w \scan_stop: \exp_not:N #2
  \if:w 0 \_fp_str_if_eq:nn { \s__fp_expr_mark } { \exp_not:N #2 }
  \else:
    \exp_after:wN \exp_after:wN \exp_after:wN \_fp_parse_infix_mark:NNN
  \fi:
\else:
  \exp_after:wN \exp_after:wN \exp_after:wN \_fp_parse_infix_mul:N
\fi:
\end{verbatim}

#1 #2
\end{verbatim}

{ End definition for \_fp_parse_infix_after_paren:NN.}

\_fp_parse_infix_mark:NNN

As an infix operator, \s__fp_expr_mark means that the next token (#3) has already
gone through \_fp_parse_infix:NN and should be provided the precedence #1. The
scan mark #2 is discarded.

\cs_new:Npn \_fp_parse_infix_mark:NNN #1#2#3 { #3 #1 #2 }

(End definition for \_fp_parse_infix_after_paren:NN.)

69.7.1 Closing parentheses and commas

As an infix operator, \s__fp_expr_mark means that the next token (#3) has already
gone through \_fp_parse_infix:NN and should be provided the precedence #1. The
scan mark #2 is discarded.
This one is a little bit odd: force every previous operator to end, regardless of the precedence.

\[
\text{\texttt{\__fp_parse_infix_end:N}}
\]

This is very similar to \texttt{\__fp_parse_infix_end:N}, complaining about an extra closing parenthesis if the previous operator was the beginning of the expression, with precedence \texttt{\c__fp_prec_end_int}.

As for other infix operations, if the previous operations has higher precedence the comma waits. Otherwise we call \texttt{\__fp_parse_apply_comma:NwNwN} whose job is to convert the first item of the tuple and an array of the remaining items into a tuple. In contrast to \texttt{\__fp_parse_apply_binary:NwNwN} this function's operands are not single-object arrays.
As described in the “work plan”, each infix operator has an associated \_\_fp_parse_infix\_\_ function, a computing function, and precedence, given as arguments to \_\_fp_tmp\_w. Using the general mechanism for arithmetic operations. The power operation must be associative in the opposite order from all others. For this, we use two distinct precedences.

69.7.2 Usual infix operators

\_\_fp_parse_infix_\+_N
\_\_fp_parse_infix_-\_N
\_\_fp_parse_infix_juxt\_N
\_\_fp_parse_infix_/:N
\_\_fp_parse_infix_mul\_N
\_\_fp_parse_infix_and\_N
\_\_fp_parse_infix_or\_N
\_\_fp_parse_infix_^\_N

\_\_fp_parse_infix_\+_N
\_\_fp_parse_infix_-\_N
\_\_fp_parse_infix_juxt\_N
\_\_fp_parse_infix_/:N
\_\_fp_parse_infix_mul\_N
\_\_fp_parse_infix_and\_N
\_\_fp_parse_infix_or\_N
\_\_fp_parse_infix_^\_N

(End definition for \_\_fp_parse_infix_-\_N, \_\_fp_parse_infix_comma:w, and \_\_fp_parse_apply_comma:NwNwN.)
69.7.3 Juxtaposition

\_fp\_parse\_infix\_(:N

When an opening parenthesis appears where we expect an infix operator, we compute the product of the previous operand and the contents of the parentheses using \_fp\_parse\_infix\_mul:N.

69.7.4 Multi-character cases

\_fp\_parse\_infix\_*:N
\_fp\_parse\_infix\_&:N
\_fp\_parse\_infix\_|:N

(End definition for \_fp\_parse\_infix\_*:N and others.)
69.7.5 Ternary operator
\_fp\_parse\_infix\_?:N
\_fp\_parse\_infix\_::N
\cs_set_protected:Npn \_fp\_tmp:w #1#2#3#4
  { \if_int_compare:w #1 < \c__fp\_prec\_quest\_int 
    \exp_after:wN @ \exp_after:wN #2 \exp:w \_fp\_parse\_operand:Nw #3 
    \exp_after:wN \_fp\_parse\_expand:w
  \else:
    \exp_after:wN @ \exp_after:wN \use\_none:n \exp_after:wN #1 \fi: }
\exp_args:Nc \_fp\_tmp:w { \_fp\_parse\_infix\_?:N } \_fp\_parse\_infix\_or:N
\exp_args:Nc \_fp\_tmp:w { \_fp\_parse\_infix\_&:N } \_fp\_parse\_infix\_and:N
(End definition for \_fp\_parse\_infix\_?:Nw and \_fp\_parse\_infix\_&:Nw.)

69.7.6 Comparisons
\_fp\_parse\_infix\_<:N
\_fp\_parse\_infix\_=:N
\_fp\_parse\_infix\>:N
\_fp\_parse\_infix\!:N
\_fp\_parse\_excl\_error:
\_fp\_parse\_compare:NNNNNNNN
\_fp\_parse\_compare\_auxi:NNNNNNNN
\_fp\_parse\_compare\_end:NNNN\_fp\_compare:wNNNN
\cs_new:cpn { \_fp\_parse\_infix\_<:N } \_fp\_parse\_infix\_<:N 
  \_fp\_parse\_compare\_compare:NNNNNN \_fp\_parse\_infix\_<:N \_fp\_parse\_infix\:_=:N 
  \_fp\_parse\_compare\_compare:NNNNNN \_fp\_parse\_infix\_:>:N \_fp\_parse\_infix\_<:N
(End definition for \_fp\_parse\_infix\_<:N and \_fp\_parse\_infix\_::N)
\exp_after:wN \__fp_parse_compare:NNNNNNN
\exp_after:wN #1
\exp_after:wN 0
\exp_after:wN 1
\exp_after:wN 1
\exp_after:wN 1
\exp_after:wN 1
\exp_after:wN { \msg_expandable_error:nnnn
{ fp } { missing } { = } { -after-!. }
}
\cs_new:Npn \__fp_parse_compare:NNNNNNN #1
{
\if_int_compare:w #1 < \c__fp_prec_comp_int
\exp_after:wN \__fp_parse_compare_auxi:NNNNNNN
\exp_after:wN \__fp_parse_excl_error:
\else:
\exp_after:wN \__fp_parse_compare:NNNNNNN
\exp_after:wN \use_none:n
\exp_after:wN \__fp_parse_compare:NNNNNNN
\fi:
}
\cs_new:Npn \__fp_parse_compare_auxi:NNNNNNN #1#2#3#4#5#6#7
{
\if_case:w
\__fp_int_eval:w \exp_after:wN ' \token_to_str:N #7 < ' \__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNNNN #2#4#5#6#7
\or: \__fp_parse_compare_auxii:NNNNNNN #2#3#5#6
\or: \__fp_parse_compare_auxii:NNNNNNN #2#3#4#6
\or: \__fp_parse_compare_auxii:NNNNNNN #2#3#4#5#2
\else: #1 \__fp_parse_compare_end:NNNNw #3#4#5#6#7
\fi:
}
\cs_new:Npn \__fp_parse_compare_end:NNNNw #1#2#3#4#5 \fi:
{
\if_case:w
\__fp_int_eval:w \exp_after:wN ' \token_to_str:N #7 = ' \__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNNNN #2#4#5#6
\or: \__fp_parse_compare_auxii:NNNNNNN #2#3#5#6
\or: \__fp_parse_compare_auxii:NNNNNNN #2#3#4#6
\or: \__fp_parse_compare_auxii:NNNNNNN #2#3#4#5#2
\else: #1 \__fp_parse_compare_end:NNNNw #3#4#5#6#7
\fi:
}
\cs_new:Npn \__fp_parse_compare:NNNNNNN #1
{
\exp_after:wN \__fp_parse_compare_auxi:NNNNNNN #1#2#3#4#5
\exp_after:wN \prg_do_nothing:
\exp_after:wN \__fp_parse_compare:NNNNNNN #1
\exp_after:wN \use_none:n
\exp:w \exp_after:wN \__fp_parse_compare:NNNNNNN #1#2#3#4#5
\fi:
\cs_new:Npn \__fp_parse_compare_end:NNNNw #1#2#3#4#5 \fi:
{
\if_case:w
\__fp_int_eval:w \exp_after:wN ' \token_to_str:N #7 - ' \__fp_int_eval_end:
\__fp_parse_compare_auxii:NNNNNNN #2#4#5#6#7
\or: \__fp_parse_compare_auxii:NNNNNNN #2#3#5#6
\or: \__fp_parse_compare_auxii:NNNNNNN #2#3#4#6
\or: \__fp_parse_compare_auxii:NNNNNNN #2#3#4#5#2
\else: #1 \__fp_parse_compare_end:NNNNw #3#4#5#6#7
\fi:
}
69.8 Tools for functions

\__fp_function_all_fp_o:fnw

Followed by \{\{function name\}\} \{\{code\}\} \{float array\} O this checks all floats are floating point numbers (no tuples).
This is followed by \{⟨function name⟩\} \{⟨code⟩\} ⟨float array⟩ @. It checks that the ⟨float array⟩ consists of one or two floating point numbers (not tuples), then leaves the ⟨code⟩ (if there is one float) or its tail (if there are two floats) followed by the ⟨float array⟩. The ⟨code⟩ should start with a single token such as \_\_fp_atan_default:w that deals with the single-float case.

The first \_\_fp_if_type_fp:NTwFw test catches the case of no argument and the case of a tuple argument. The next one distinguishes the case of a single argument (no error, just add \c_one_fp) from a tuple second argument. Finally check there is no further argument.
\if_meaning:w @ #5 \else:
\exp_after:wN \__fp_parse_function_one_two_error_o:w
\fi:
\use_iin {#1} { \use_none:n #2 } #3; #4; #5

(End definition for \_fp_parse_function_one_two:nnw and others.)

\__fp_tuple_map_o:nw \__fp_tuple_map_loop_o:nw

Apply \#1 to all items in the following tuple and expand once afterwards. The code \#1 should itself expand once after its result.
\cs_new:Npn \__fp_tuple_map_o:nw #1 \s__fp_tuple \__fp_tuple_chk:w #2 ;
{\exp:w \exp_end_continue_f:w \__fp_tuple_map_loop_o:nw {#1} #2 { \s__fp \prg_break: } ;}
\prg_break_point:
{ \exp_after:wN } \exp_after:wN ;

\cs_new:Npn \__fp_tuple_map_loop_o:nw #1#2#3 ; #4 @ #5#6 ;
{\use_none:n #2 \use_none:n #5 \use_none:n \__fp \prg_break_point: \exp_after:wN } \exp_after:wN ;

(End definition for \_fp_tuple_map_o:nw and \_fp_tuple_map_loop_o:nw.)

\__fp_tuple_mapthread_o:nww \__fp_tuple_mapthread_loop_o:nw

Apply \#1 to pairs of items in the two following tuples and expand once afterwards.
\cs_new:Npn \__fp_tuple_mapthread_o:nww #1 \s__fp_tuple \__fp_tuple_chk:w #2 ;
\s__fp_tuple \__fp_tuple_chk:w #3 ;
{\exp:w \exp_end_continue_f:w \__fp_tuple_mapthread_loop_o:nw {#1} #2 { \s__fp \prg_break: } ; @ #3 { \s__fp \prg_break: } ;}
\prg_break_point:
{ \exp_after:wN } \exp_after:wN ;

\cs_new:Npn \__fp_tuple_mapthread_loop_o:nw #1#2#3 ; #4 @ #5#6 ;
{\use_none:n \__fp \prg_break_point: \exp_after:wN } \exp_after:wN ;

(End definition for \_fp_tuple_mapthread_o:nww and \_fp_tuple_mapthread_loop_o:nw.)
### 69.9 Messages

\begin{verbatim}
\msg_new:nnn { fp } { deprecated } 
  { '#1'-deprecated;-use-'#2' }
\msg_new:nnn { fp } { unknown-fp-word }
  { Unknown-fp-word=#1. }
\msg_new:nnn { fp } { missing }
  { Missing=#1-inserted #2. }
\msg_new:nnn { fp } { extra }
  { Extra=#1-ignored. }
\msg_new:nnn { fp } { early-end }
  { Premature-end-in-fp-expression. }
\msg_new:nnn { fp } { after-e }
  { Cannot-use=#1 after-`e`. }
\msg_new:nnn { fp } { missing-number }
  { Missing-number-before-`#1`. }
\msg_new:nnn { fp } { unknown-symbol }
  { Unknown-symbol=#1-ignored. }
\msg_new:nnn { fp } { extra-comma }
  { Unexpected-comma-turned-to-nan-result. }
\msg_new:nnn { fp } { no-arg }
  { #1-got-no-argument;-used-nan. }
\msg_new:nnn { fp } { multi-arg }
  { #1-got-more-than-one-argument;-used-nan. }
\msg_new:nnn { fp } { num-args }
  { #1-expects-between-#2-and-#3-arguments. }
\msg_new:nnn { fp } { bad-args }
  { Arguments-in-`#1#2`-are-invalid. }
\msg_new:nnn { fp } { infty-pi }
  { Math-command=`#1` is-not-an-fp }
\cs_if_exist:cT { @unexpandable@protect }
  { \msg_new:nnn { fp } { robust-cmd }
    { Robust-command=`#1` invalid-in-fp-expression! } }
\end{verbatim}
Chapter 70

l3fp-assign implementation

70.1 Assigning values

\fp_new:N

Floating point variables are initialized to be $+0$.

\cs_new_protected:Npn \fp_new:N #1
\cs_generate_variant:Nn \fp_new:N { c }

(End definition for \fp_new:N. This function is documented on page 239.)

\fp_set:Nn
\fp_set:cn
\fp_gset:Nn
\fp_gset:cn
\fp_const:Nn
\fp_const:cn

Simply use \__fp_parse:n within various f-expanding assignments.

\cs_new_protected:Npn \fp_set:Nn #1#2
\cs_new_protected:Npn \fp_gset:Nn #1#2
\cs_new_protected:Npn \fp_const:Nn #1#2
\cs_generate_variant:Nn \fp_set:Nn { c }
\cs_generate_variant:Nn \fp_gset:Nn { c }
\cs_generate_variant:Nn \fp_const:Nn { c }

(End definition for \fp_set:Nn, \fp_gset:Nn, and \fp_const:Nn. These functions are documented on page 239.)

\fp_set_eq:NN
\fp_set_eq:cN
\fp_gset_eq:NN
\fp_gset_eq:cN
\fp_set_eq:cc
\fp_gset_eq:cc

Copying a floating point is the same as copying the underlying token list.

\cs_new_protected:Npn \fp_set_eq:NN #1#2
\cs_new_protected:Npn \fp_gset_eq:NN #1#2
\cs_generate_variant:Nn \fp_set_eq:NN { c , Nc , cc }
\cs_generate_variant:Nn \fp_gset_eq:NN { c , Nc , cc }

(End definition for \fp_set_eq:NN and \fp_gset_eq:NN. These functions are documented on page 239.)

\fp_zero:N
\fp_zero:c
\fp_gzero:N
\fp_gzero:c

Setting a floating point to zero: copy \c_zero_fp.

\cs_new_protected:Npn \fp_zero:N #1 { \fp_set_eq:NN #1 \c_zero_fp }
\cs_new_protected:Npn \fp_gzero:N #1 { \fp_gset_eq:NN #1 \c_zero_fp }
\cs_generate_variant:Nn \fp_zero:N { c }
\cs_generate_variant:Nn \fp_gzero:N { c }

1037
Set the floating point to zero, or define it if needed.

\begin{verbatim}
\fp_zero_new:N \fp_zero_new:c \fp_gzero_new:N \fp_gzero_new:c
\end{verbatim}

For the sake of error recovery we should not simply set #1 to #1±(#2): for instance, if #2
is 0+2, the parsing error would be raised at the last closing parenthesis rather than at
the closing parenthesis in the user argument. Thus we evaluate #2 instead of just putting
parentheses. As an optimization we use \texttt{\_\_fp_parse:n} rather than \texttt{\fp_eval:n}, which
would convert the result away from the internal representation and back.

\begin{verbatim}
\fp_add:Nn \fp_add:cn \fp_gadd:Nn \fp_gadd:cn \fp_sub:Nn \fp_sub:cn \fp_gsub:Nn \fp_gsub:cn
\end{verbatim}

This shows the result of computing its argument by passing the right data to \texttt{\tl_show:n}
or \texttt{\tl_log:n}.

\begin{verbatim}
\fp_show:N \fp_show:c \fp_log:N \fp_log:c \fp_gshow:NN \_\_fp_show:NN \_\_fp_show_validate:w
\end{verbatim}

\section{showing values}

This shows the result of computing its argument by passing the right data to \texttt{\tl_show:n}
or \texttt{\tl_log:n}.
\texttt{\exp_args:Nx \texttt{#1} \{ \token_to_str:N \texttt{#2} = \fp_to_tl:N \texttt{#2} \}}

\texttt{\cs_new:Npn \__fp\_show\_validate:w} 
\begin{Verbatim}
#1 \m_sfp \__fp\_chk:w \#2\#4\#5 ; \#6 \m_sfp\_stop
\end{Verbatim}

\texttt{\token_if_eq_meaning:NNTF \texttt{#2} 1} 
\begin{Verbatim}
\{ \m_sfp \__fp\_chk:w \#2 \#3 \{\#4\} \#5 ; \}
\{ \m_sfp \__fp\_chk:w \#2 \#3 \#4 \#5 ; \}
\end{Verbatim}

(End definition for \texttt{\fp\_show:N} and others. These functions are documented on page 247.)

\texttt{\fp\_show:n} 
\texttt{\fp\_log:n}

Use general tools.

\texttt{\cs_new_protected:Npn \texttt{\fp\_show:n}} 
\begin{Verbatim}
\texttt{\m_msg\_show\_eval:Nn \texttt{\fp\_to\_tl:n}}
\end{Verbatim}

\texttt{\cs_new_protected:Npn \texttt{\fp\_log:n}} 
\begin{Verbatim}
\texttt{\m_msg\_log\_eval:Nn \texttt{\fp\_to\_tl:n}}
\end{Verbatim}

(End definition for \texttt{\fp\_show:n} and \texttt{\fp\_log:n}. These functions are documented on page 247.)

\section*{70.4 Some useful constants and scratch variables}

\texttt{\c\_one\_fp} 
\texttt{\c\_e\_fp}

Some constants.

\begin{Verbatim}
\fp\_const:Nn \texttt{\c\_one\_fp} \{ 2.718 2818 2845 9045 \}
\fp\_const:Nn \texttt{\c\_e\_fp} \{ 1 \}
\end{Verbatim}

(End definition for \texttt{\c\_one\_fp} and \texttt{\c\_e\_fp}. These variables are documented on page 245.)

\texttt{\c\_pi\_fp} 
\texttt{\c\_one\_degree\_fp}

We simply round $\pi$ to and $\pi/180$ to 16 significant digits.

\begin{Verbatim}
\fp\_const:Nn \texttt{\c\_pi\_fp} \{ 3.141 5926 5358 9793 \}
\fp\_const:Nn \texttt{\c\_one\_degree\_fp} \{ 0.0 1745 3292 5199 4330 \}
\end{Verbatim}

(End definition for \texttt{\c\_pi\_fp} and \texttt{\c\_one\_degree\_fp}. These variables are documented on page 246.)

\texttt{\l\_tmpa\_fp} 
\texttt{\l\_tmpb\_fp} 
\texttt{\g\_tmpa\_fp} 
\texttt{\g\_tmpb\_fp}

Scratch variables are simply initialized there.

\begin{Verbatim}
\fp\_new:N \texttt{\l\_tmpa\_fp}
\fp\_new:N \texttt{\l\_tmpb\_fp}
\fp\_new:N \texttt{\g\_tmpa\_fp}
\fp\_new:N \texttt{\g\_tmpb\_fp}
\end{Verbatim}

(End definition for \texttt{\l\_tmpa\_fp} and others. These variables are documented on page 246.)
Chapter 71

l3fp-logic Implementation

Those functions may receive a variable number of arguments.

\cs_new:Npn \__fp_parse_word_max:N { \__fp_parse_function:NNN \__fp_minmax_o:Nw 2 }
\cs_new:Npn \__fp_parse_word_min:N { \__fp_parse_function:NNN \__fp_minmax_o:Nw 0 }

(End definition for \__fp_parse_word_max:N and \__fp_parse_word_min:N.)

71.1 Syntax of internal functions

- \__fp_compare_npos:nwnw \{\langle expo_1\rangle \langle body_1\rangle ; \langle expo_2\rangle \langle body_2\rangle ;
- \__fp_minmax_o:Nw \langle sign \rangle \langle floating point array \rangle
- \__fp_not_o:w ? \langle floating point array \rangle (with one floating point number only)
- \__fp_&_o:ww \langle floating point \rangle \langle floating point \rangle
- \__fp_|_o:ww \langle floating point \rangle \langle floating point \rangle
- \__fp_ternary:NwwN, \__fp_ternary_auxi:NwwN, \__fp_ternary_auxii:NwwN

have to be understood.

71.2 Tests

\fp_if_exist_p:N Copies of the cs functions defined in l3basics.
\fp_if_exist_p:c
\fp_if_exist:NTF
\fp_if_exist:cTF

(End definition for \fp_if_exist:NTF. This function is documented on page 241.)
Evaluate and check if the result is a floating point of the same kind as NaN.

\begin{verbatim}
\prg_new_conditional:Nnn \fp_if_nan:n \fp_if_nan:nTF
{ \fp:w 3 \exp_last_unbraced:Nf \__fp_kind:w \__fp_parse:n (#1) }
\prg_return_true:
\else:
\prg_return_false:
\fi:
\end{verbatim}

(End definition for \texttt{\fp_if_nan:nTF}. This function is documented on page 300.)

### 71.3 Comparison

Within floating point expressions, comparison operators are treated as operations, so we evaluate \#1, then compare with $\pm 0$. Tuples are \texttt{true}.

\begin{verbatim}
\prg_new_conditional:Nnn \fp_compare:n \fp_compare:nTF
{ \exp:w \exp_end_continue_f:w \__fp_parse:n (#1) }
\cs_new:Npn \__fp_compare_return:w #1#2#3;
{ \if_charcode:w 0 \__fp_if_type_fp:NTwFw #1 { \__fp_use_i_delimit_by_s_stop:nw #3 \s__fp_stop } \s__fp 1 \s__fp_stop \prg_return_false: \else: \prg_return_true: \fi: }
\end{verbatim}

(End definition for \texttt{\fp_compare:nTF} and \texttt{\__fp_compare_return:w}. This function is documented on page 243.)

Evaluate \#1 and \#3, using an auxiliary to expand both, and feed the two floating point numbers swapped to \texttt{\__fp_compare_back_any:ww}, defined below. Compare the result with `\#2`'=, which is $-1$ for $<$, $0$ for $=$, $1$ for $>$ and $2$ for ?.

\begin{verbatim}
\prg_new_conditional:Nnn \fp_compare:nNn \fp_compare:nNnTF \__fp_compare_aux:wn
{ \if_int_compare:w \__fp_int_eval:w '#2 - '=' \__fp_int_eval_end: \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End definition for \texttt{\fp_compare:nNn} and \texttt{\__fp_compare_aux:wn}. This function is documented on page 243.)
\_\_fp_compare_back_an\_\_any:ww
\_\_fp_compare_back:ww
\_\_fp_compare_nan:ww

\_\_fp_compare_back_an\_\_any:ww \{ \}
\_\_fp_compare_back:ww \{ \}
\_\_fp_compare_nan:ww

\cs_new:Npn \__fp_compare_back_an\_\_any:ww #1\#2 \#3
\{
\__fp_if_type_fp:NTwFw #1 { \__fp_if_type_fp:NTwFw #3 \use_i:nn \s__fp \use_ii:nn \s__fp_stop }
\s__fp \use_ii:nn \s__fp_stop
\__fp_compare_back:ww \{
\cs:w __fp \__fp_type_from_scan:N #1 \_compare_back \__fp_type_from_scan:N #3 :ww
\cs_end:
\}
\}
\cs_new:Npn \__fp_compare_back:ww
\s__fp \__fp_chk:w #1 #2 #3;
\s__fp \__fp_chk:w #4 #5 #6;
\{
\int_value:w \if_meaning:w 3 \#1 \exp_after:wN \__fp_compare_nan:ww \fi:
\if_meaning:w 3 \#4 \exp_after:wN \__fp_compare_nan:ww \fi:
\if_meaning:w 2 \#5 - \fi:
\if_meaning:w \#2 \#5
\if_meaning:w \#1 \#4
\if_meaning:w \#1 \#1
\if_meaning:w \#1 \#3;
\else:
0
\fi:
\else:
\if_int_compare:w \#4 < \#1 - \fi: 1
\fi:
\else:
\if_int_compare:w \#1\#4 = \c_zero_int 0
\else:
1
\fi:
Tuples and floating point numbers are not comparable so return 2 in mixed cases or when tuples have a different number of items. Otherwise compare pairs of items with \_\_fp\_compare\_back\_tuple and \_\_fp\_compare\_back\_tuple and others.

Within an \texttt{\int_value:w ... \exp_stop_f:} construction, this expands to 0 if the two numbers are equal, \(-1\) if the first is smaller, and 1 if the first is bigger. First compare the exponents: the larger one denotes the larger number. If they are equal, we must compare significands. If both the first 8 digits and the next 8 digits coincide, the numbers are equal. If only the first 8 digits coincide, the next 8 decide. Otherwise, the first 8 digits are compared.
\fp_do_until:nn
\fp_do_while:nn
\fp_until_do:nn
\fp_while_do:nn

These are quite easy given the above functions. The \texttt{do\_until} and \texttt{do\_while} versions execute the body, then test. The \texttt{until\_do} and \texttt{while\_do} do it the other way round.

\begin{verbatim}
\cs_new:Npn \fp_do_until:nn #1#2
{ #2 \fp_compare:nF {#1} \{ \fp_do_until:nn {#1} {#2} \} }
\cs_new:Npn \fp_do_while:nn #1#2
{ #2 \fp_compare:nT {#1} \{ \fp_do_while:nn {#1} {#2} \} }
\cs_new:Npn \fp_until_do:nn #1#2
{ \fp_compare:nF {#1} \{ \fp_until_do:nn {#1} {#2} \} }
\cs_new:Npn \fp_while_do:nn #1#2
{ \fp_compare:nT {#1} \{ \fp_while_do:nn {#1} {#2} \} }
\end{verbatim}

(End definition for \texttt{\_fp\_compare\_npos:nww} and \texttt{\_fp\_compare\_significand:nnnnnn}.)

\section{Floating point expression loops}

These are quite easy given the above functions. The \texttt{do\_until} and \texttt{do\_while} versions execute the body, then test. The \texttt{until\_do} and \texttt{while\_do} do it the other way round.

\begin{verbatim}
\cs_new:Npn \fp_do_until:nn #1#2
{ #2 \fp_compare:nF {#1} \{ \fp_do_until:nn {#1} {#2} \} }
\cs_new:Npn \fp_do_while:nn #1#2
{ #2 \fp_compare:nT {#1} \{ \fp_do_while:nn {#1} {#2} \} }
\cs_new:Npn \fp_until_do:nn #1#2
{ \fp_compare:nF {#1} \{ \fp_until_do:nn {#1} {#2} \} }
\cs_new:Npn \fp_while_do:nn #1#2
{ \fp_compare:nT {#1} \{ \fp_while_do:nn {#1} {#2} \} }
\end{verbatim}

(End definition for \texttt{\fp\_do\_until:nn} and others. These functions are documented on page 244.)
As above but not using the \texttt{nNn} syntax.

\begin{verbatim}
\cs_new:Npn \fp_do_until:nNnn \#1\#2\#3\#4
\{
\fp_compare:nNnF \#1 \#2 \#3
\{
\fp_do_until:nNnn \#1 \#2 \#3 \#4
\}
\cs_new:Npn \fp_do_while:nNnn \#1\#2\#3\#4
\{
\fp_compare:nNnT \#1 \#2 \#3
\{
\fp_do_while:nNnn \#1 \#2 \#3 \#4
\}
\cs_new:Npn \fp_until_do:nNnn \#1\#2\#3\#4
\{
\fp_compare:nNnF \#1 \#2 \#3
\{
\fp_until_do:nNnn \#1 \#2 \#3 \#4
\}
\}
\cs_new:Npn \fp_while_do:nNnn \#1\#2\#3\#4
\{
\fp_compare:nNnT \#1 \#2 \#3
\{
\fp_while_do:nNnn \#1 \#2 \#3 \#4
\}
\}
\end{verbatim}

(End definition for \texttt{\fp_do_until:nNnn} and others. These functions are documented on page 243.)

The approach here is somewhat similar to \texttt{\int_step_function:nnnN}. There are two subtleties: we use the internal parser \texttt{\_\_\_fp_parse:n} to avoid converting back and forth from the internal representation; and (due to rounding) even a non-zero step does not guarantee that the loop counter increases.

\begin{verbatim}
\cs_new:Npn \fp_step_function:nnnN \#1\#2\#3
\{
\exp_after:wN \__fp_step:wwwN
\exp:w \exp_end_continue_f:w \__fp_parse_o:n \#1
\exp:w \exp_end_continue_f:w \__fp_parse_o:n \#2
\exp:w \exp_end_continue_f:w \__fp_parse:n \#3
\}
\cs_generate_variant:Nn \fp_step_function:nnnN { nnnc }
\end{verbatim}

% Only floating point numbers (not tuples) are allowed arguments.
% Only \enquote{normal} floating points (not $\pm 0$, $\pm\texttt{inf}$, \texttt{nan}) can be used as step; if positive, call \texttt{\cs{\_\_\_fp_step:NNnnN\_\_\_fp_ParSe_o:n}} with argument $\#1$ otherwise $-\#1$. This function has one more argument than its integer counterpart, namely the previous value, to catch the case where the loop has made no progress. Conversion to decimal is done just before calling the user’s function.
\cs_new:Npn \__fp_step:wwwN #1#2; #3#4; #5#6; #7
{
\__fp_if_type_fp:NTwFw #1 { } \s__fp \prg_break: \s__fp_stop
\__fp_if_type_fp:NTwFw #3 { } \s__fp \prg_break: \s__fp_stop
\__fp_if_type_fp:NTwFw #5 { } \s__fp \prg_break: \s__fp_stop
\use_i:nnnn { \__fp_step_fp:wwwN #1#2; #3#4; #5#6; #7 }
\prg_break_point:
\use:n
\{ 
\__fp_error:nfff { step-tuple } { \fp_to_tl:n { #1#2 ; } }
{ \fp_to_tl:n { #3#4 ; } } { \fp_to_tl:n { #5#6 ; } }
\}
\}
\cs_new:Npn \__fp_step_fp:wwwN #1 ; \s__fp \__fp_chk:w #2#3#4 ; #5; #6
{ #1 ; } { \c_nan_fp } { \s__fp \__fp_chk:w #2#3#4 ; } {#5 ; } #6
\cs_generate_variant:Nn \__fp_step:NnnnnN { Nf }

(End definition for \fp_step_function:nnnN and others. This function is documented on page 245.)
As for \texttt{\texttt{\texttt{\texttt{\int_step_inline:nnnn}, create a global function and apply it, following up with a break point.}}},

\begin{verbatim}
cs_new_protected:Npn \fp_step_inline:nnnn  
{ \int_gincr:N \g__kernel_prg_map_int \exp_args:NNc \__fp_step:NNnnnn \cs_gset_protected:Npn \{ __fp_map_ \int_use:N \g__kernel_prg_map_int :w \} } 
cs_new_protected:Npn \fp_step_variable:nnnNn #1#2#3#4#5  
{ \int_gincr:N \g__kernel_prg_map_int \exp_args:NNc \__fp_step:NNnnnn \cs_gset_protected:Npx \{ __fp_map_ \int_use:N \g__kernel_prg_map_int :w \} \tl_set:Nn \exp_not:N #4 {##1} \exp_not:n {#5} } 
cs_new_protected:Npn \__fp_step:NNnnnn #1#2#3#4#5#6  
{ #1 #2 ##1 {#6} \fp_step_function:nnnN {#3} {#4} {#5} \prg_break_point:Nn \scan_stop: { \int_gdecr:N \g__kernel_prg_map_int } } 
\end{verbatim}

(End definition for \texttt{\texttt{\texttt{\int_step_inline:nnnn}, \fp_step_variable:nnnNn, and \__fp_step:NNnnnn. These functions are documented on page 245.)})

\begin{verbatim}
cs_new:Npn \__fp_minmax_o:Nw #1  
{ \__fp_parse_function_all_fp_o:fnw \token_if_eq_meaning:NNTF 0 #1 { min } { max } } 
cs_new:Npn \__fp_minmax_aux_o:Nw  
{ \__fp_minmax_loop:Nww } 
\end{verbatim}

71.5 Extrema

First check all operands are floating point numbers. The argument \#1 is 2 to find the maximum of an array \#2 of floating point numbers, and 0 to find the minimum. We read numbers sequentially, keeping track of the largest (smallest) number found so far. If numbers are equal (for instance \pm0), the first is kept. We append \(-\infty\) (\(\infty\)), for the case of an empty array. Since no number is smaller (larger) than that, this additional item only affects the maximum (minimum) in the case of \texttt{max()} and \texttt{min()} with no argument. The weird \texttt{fp}-like trailing marker breaks the loop correctly: see the precise definition of \texttt{\__fp_minmax_loop:Nww}.
\__fp_minmax_loop:Nww

The first argument is $-$ or $+$ to denote the case where the currently largest (smallest) number found (first floating point argument) should be replaced by the new number (second floating point argument). If the new number is \texttt{nan}, keep that as the extremum, unless that extremum is already a \texttt{nan}. Otherwise, compare the two numbers. If the new number is larger (in the case of \texttt{max}) or smaller (in the case of \texttt{min}), the test yields \texttt{true}, and we keep the second number as a new maximum; otherwise we keep the first number. Then loop.

\__fp_minmax_auxi:ww
\__fp_minmax_auxii:ww

Keep the first/second number, and remove the other.
This function is called from within an \texttt{if meaning} test. Skip to the end of the tests, close the current test with \\texttt{fi}, clean up, and return the appropriate number with one post-expansion.

\begin{verbatim}
\cs_new:Npn \__fp_minmax_break_o:w #1 \fi: \fi: #2 \s__fp #3; #4;
\end{verbatim}

\section{Boolean operations}

Return \texttt{true} or \texttt{false}, with two expansions, one to exit the conditional, and one to please \texttt{l3fp-parse}. The first argument is provided by \texttt{l3fp-parse} and is ignored.

\begin{verbatim}
\cs_new:Npn \__fp_not_o:w #1 \s__fp \__fp_chk:w #2#3; @
{ \if_meaning:w 0 #2
\exp_after:wN \exp_after:wN \exp_after:wN \c_one_fp
\else:
\exp_after:wN \exp_after:wN \exp_after:wN \c_zero_fp
\fi:
\__fp_exp_after_o:w}
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_tuple_not_o:w #1 @ { \exp_after:wN \c_zero_fp }
\end{verbatim}

For \texttt{and}, if the first number is zero, return it (with the same sign). Otherwise, return the second one. For \texttt{or}, the logic is reversed: if the first number is non-zero, return it, otherwise return the second number: we achieve that by hijacking \texttt{__fp\_\_o:ww}, inserting an extra argument, \texttt{\else:}, before \texttt{\s__fp}. In all cases, expand after the floating point number.

\begin{verbatim}
\group_begin:
\char_set_catcode_letter:N &
\char_set_catcode_letter:N |
\cs_new:Npn \__fp_&_o:ww #1 \s__fp \__fp_chk:w #2#3; 
{ \if_meaning:w 0 #2 #1 \__fp_and_return:wNw \s__fp \__fp_chk:w #2#3; 
\fi:
\__fp_exp_after_o:w }
\cs_new:Npn \__fp_&_tuple_o:ww #1 \s__fp \__fp_chk:w #2#3; 
{ \if_meaning:w 0 #2 #1 \__fp_and_return:ww \s__fp \__fp_chk:w #2#3; 
\fi:
\__fp_exp_after_tuple_o:w }
\cs_new:Npn \__fp\_\_o:ww \__fp\_\_tuple_o:ww #1; { \__fp_exp_after_tuple_o:w }
\cs_new:Npn \__fp\_\_tuple_o:ww #1; #2; { \__fp_exp_after_tuple_o:w #1; }
\end{verbatim}
\__fp_ternary:NwwN \__fp_ternary_auxi:NwwN \__fp_ternary_auxii:NwwN

The first function receives the test and the true branch of the ?:\ ternary operator. It calls \__fp_ternary_auxii:NwwN if the test branch is a floating point number \pm 0, and otherwise calls \__fp_ternary_auxi:NwwN. These functions select one of their two arguments.

\__fp_ternary:NwwN #1 #2#3@ #4@ #5

\__fp_ternary_auxi:NwwN #1#2@#3@#4

\__fp_ternary_auxii:NwwN #1#2@#3@#4

(End definition for \__fp_&_o:ww and others.)

\__fp_tuple_|_tuple_o:ww #1; #2;
{ \__fp_exp_after_tuple_o:w #1; }
\group_end:
\__fp_and_return:wNw #1; \fi: #2;
{ \fi: \__fp_exp_after_o:w #1; }

71.7 Ternary operator

The first function receives the test and the true branch of the :? ternary operator. It calls \__fp_ternary_auxii:NwwN if the test branch is a floating point number \pm 0, and otherwise calls \__fp_ternary_auxi:NwwN. These functions select one of their two arguments.
(End definition for \_fp_ternary:NwwN, \_fp_ternary_auxi:NwwN, and \_fp_ternary_auxii:NwwN.)
Chapter 72

l3fp-basics Implementation

The l3fp-basics module implements addition, subtraction, multiplication, and division of two floating points, and the absolute value and sign-changing operations on one floating point. All operations implemented in this module yield the outcome of rounding the infinitely precise result of the operation to the nearest floating point.

Some algorithms used below end up being quite similar to some described in “What Every Computer Scientist Should Know About Floating Point Arithmetic”, by David Goldberg, which can be found at http://cr.yp.to/2005-590/goldberg.pdf.

Unary functions.

\__fp_parse_word_abs:N
\__fp_parse_word_logb:N
\__fp_parse_word_sign:N
\__fp_parse_word_sqrt:N

(End definition for \__fp_parse_word_abs:N and others.)

72.1 Addition and subtraction

We define here two functions, \__fp_-_o:ww and \__fp_+_o:ww, which perform the subtraction and addition of their two floating point operands, and expand the tokens following the result once.

A more obscure function, \__fp_add_big_i_o:wNww, is used in l3fp-expo.

The logic goes as follows:

- \__fp_-_o:ww calls \__fp_+_o:ww to do the work, with the sign of the second operand flipped;
- \__fp_+_o:ww dispatches depending on the type of floating point, calling specialized auxiliaries;
• in all cases except summing two normal floating point numbers, we return one or the other operands depending on the signs, or detect an invalid operation in the case of $\infty - \infty$;

• for normal floating point numbers, compare the signs;

• to add two floating point numbers of the same sign or of opposite signs, shift the significand of the smaller one to match the bigger one, perform the addition or subtraction of significands, check for a carry, round, and pack using the `__fp_basics_pack_...` functions.

The trickiest part is to round correctly when adding or subtracting normal floating point numbers.

72.1.1 Sign, exponent, and special numbers

The `__fp_-o:ww` auxiliary has a hook: it takes one argument between the first `s_-fp` and `__fp_chk:w`, which is applied to the sign of the second operand. Positioning the hook there means that `__fp_-o:ww` can still perform the sanity check that it was followed by `s_-fp`.

\begin{verbatim}
\cs_new:cpx { __fp_-o:ww } \s__fp
{ \exp_not:c { __fp_+_o:ww } \s__fp_neg_sign:N }
\end{verbatim}

(End definition for `__fp_-o:ww`.)

This function is either called directly with an empty #1 to compute an addition, or it is called by `__fp_-o:ww` with `__fp_neg_sign:N` as #1 to compute a subtraction, in which case the second operand’s sign should be changed. If the ⟨types⟩ #2 and #4 are the same, dispatch to case #2 (0, 1, 2, or 3), where we call specialized functions: thanks to `\int_value:w`, those receive the tweaked ⟨sign⟩ (expansion of #1#5) as an argument. If the ⟨types⟩ are distinct, the result is simply the floating point number with the highest ⟨type⟩. Since case 3 (used for two nan) also picks the first operand, we can also use it when ⟨type1⟩ is greater than ⟨type2⟩. Also note that we don’t need to worry about ⟨sign⟩ in that case since the second operand is discarded.

\begin{verbatim}
\cs_new:cpn { __fp_+_o:ww }
\s__fp #1 \__fp_chk:w #2 #3 ; \s__fp \__fp_chk:w #4 #5
{ \if_case:w #2 \if_meaning:w #2 #4 \else:\if_int_compare:w #2 > #4 \exp_stop_f:
 3 \else:\4 \fi: \fi: \exp_stop_f:
\exp_after:wN \__fp_add_zeros_o:Nww \int_value:w
\or: \exp_after:wN \__fp_add_normal_o:Nww \int_value:w
\end{verbatim}

1053
\__fp_add_return_ii_o:Nww

Ignore the first operand, and return the second, but using the sign \#1 rather than \#4. As usual, expand after the floating point.
\cs_new:Npn \__fp_add_return_ii_o:Nww \#1 \#2 ; \s__fp \__fp_chk:w \#3 \#4
{ \__fp_exp_after_o:w \s__fp \__fp_chk:w \#3 \#1 }

(End definition for \__fp_add_return_ii_o:Nww.)

\__fp_add_zeros_o:Nww

Adding two zeros yields \c_zero_fp, except if both zeros were \texttt{-0}.
\cs_new:Npn \__fp_add_zeros_o:Nww \#1 \s__fp \__fp_chk:w \texttt{0} \#2
{ \if_int_compare:w \#2 \#1 = 20 \exp_stop_f: \else: \__fp_case_return_i_o:ww \fi: \#1 \#2 \s__fp \__fp_chk:w \texttt{0} \#2 }

(End definition for \__fp_add_zeros_o:Nww.)

\__fp_add_inf_o:Nww

If both infinities have the same sign, just return that infinity, otherwise, it is an invalid operation. We find out if that invalid operation is an addition or a subtraction by testing whether the tweaked \texttt{⟨sign2⟩} (\#1) and the \texttt{⟨sign2⟩} (\#4) are identical.
\cs_new:Npn \__fp_add_inf_o:Nww
\#1 \s__fp \__fp_chk:w \texttt{2} \#2 \#3; \s__fp \__fp_chk:w \texttt{2} \#4
{ \if_meaning:w \#1 \#2 \__fp_case_return_i_o:ww \else: \__fp_case_use:nw \exp_last_unbraced:Nf \__fp_invalid_operation_o:Nww \{ \token_if_eq_meaning:NNTF \#1 \#4 + - \} \fi: \s__fp \__fp_chk:w \texttt{2} \#2 \#3; \s__fp \__fp_chk:w \texttt{2} \#4 }

(End definition for \__fp_add_inf_o:Nww.)
We now have two normal numbers to add, and we have to check signs and exponents more carefully before performing the addition.

\[\text{24879} \cs_new:Npn \__fp_add_normal_o:Nww #1 \s__fp \__fp_chk:w 1 #2 \{
\text{if meaning:w #1#2}
\exp_after:wN \__fp_add_npos_o:NnwNnw
\else:
\exp_after:wN \__fp_sub_npos_o:NnwNnw
\fi:
\#2
\}
\]  
(End definition for \__fp_add_normal_o:Nww.)

### 72.1.2 Absolute addition

In this subsection, we perform the addition of two positive normal numbers.

\[\text{24888} \cs_new:Npn \__fp_add_npos_o:NnwNnw \{\text{sign1}\} \{\exp1\} \{\body1\} ; \s__fp \__fp_chk:w 1 \{\text{initial sign2}\} \{\exp2\} \{\body2\} ; \]

Since we are doing an addition, the final sign is \text{sign1}. Start an \__fp_int_eval:w, responsible for computing the exponent: the result, and the \text{final sign} are then given to \__fp_sanitize:Nw which checks for overflow. The exponent is computed as the largest exponent \#2 or \#5, incremented if there is a carry. To add the significands, we decimate the smaller number by the difference between the exponents. This is done by \__fp_add_big_i:oNWww or \__fp_add_big_ii:oNWww. We need to bring the final sign with us in the midst of the calculation to round properly at the end.

\[\text{24902} \cs_new:Npn \__fp_add_big_i:oNWww \{\text{shift}\} ; \{\text{final sign}\} \{\body1\} ; \{\body2\} ; \]

Used in l3fp-expo. Shift the significand of the small number, then add with \__fp_add_significand_o:Nnnwnnn.

\[\text{24903} \cs_new:Npn \__fp_add_big_ii:oNWww \{\text{shift}\} ; \{\text{final sign}\} \{\body1\} ; \{\body2\} ; \]

(End definition for \__fp_add_npos_o:NnwNnw.)
To round properly, we must know at which digit the rounding should occur. This requires to know whether the addition produces an overall carry or not. Thus, we do the computation now and check for a carry, then go back and do the rounding. The rounding may cause a carry in very rare cases such as 0.999999... → 1.000000..., but this situation always give an exact power of 10, for which it is easy to correct the result at the end.

If there’s no carry, grab all the digits again and round. The packing function \texttt{\_fp_basics\_pack\_high:NNNNNw} takes care of the case where rounding brings a carry.
The case where there is a carry is very similar. Rounding can even raise the first digit from 1 to 2, but we don’t care.

72.1.3 Absolute subtraction

Rounding properly in some modes requires to know what the sign of the result will be. Thus, we start by comparing the exponents and significands. If the numbers coincide, return zero. If the second number is larger, swap the numbers and call \_\_fp\_sub\_npo\_ii\_o with the opposite of \signi.

\cs_new:Npn \_\_fp\_sub\_npo\_o\_ii\_o #1; #2; #3; { \exp_after:wN \c_zero_fp };
\__fp_sub_npos_i_o:Nnwnw

\__fp_sub_npos_i_o:Nnwnw

After the computation is done, \__fp_sanitize:Nw checks for overflow/underflow. It expects the \textit{(final sign)} and the \textit{(exponent)} (delimited by ;). Start an integer expression for the exponent, which starts with the exponent of the largest number, and may be decreased if the two numbers are very close. If the two numbers have the same exponent, call the \textit{near} auxiliary. Otherwise, decimate \( y \), then call the \textit{far} auxiliary to evaluate the difference between the two significands. Note that we decimate by 1 less than one could expect.

\cs_new:Npn \__fp_sub_npos_i_o:Nnwnw #1 #2#3; #4#5;
{ \exp_after:wN \__fp_sanitize:Nw \exp_after:wN #1 \int_value:w \__fp_int_eval:w #2 \if_int_compare:w #2 = #4 \exp_stop_f: \exp_after:wN \__fp_sub_back_near_o:nnnnnnnnN \else: \exp_after:wN \__fp_decimate:nNnnnn \exp_after:wN \int_value:w \__fp_int_eval:w #2 - #4 - 1 \exp_after:wN \__fp_sub_back_far_o:NnnwnnnnN \fi: #5 #3 #1 }

(End definition for \__fp_sub_npos_i_o:Nnwnw.)

\__fp_sub_back_near_o:nnnnnnnnN \__fp_sub_back_near_pack:NNNNNNw \__fp_sub_back_near_after:wNNNNw

\__fp_sub_back_near_o:nnnnnnnnN \{{Y_1}\} \{{Y_2}\} \{{Y_3}\} \{{Y_4}\} \{{X_1}\} \{{X_2}\} \{{X_3}\} \{{X_4}\} \langle \text{final sign} \rangle

In this case, the subtraction is exact, so we discard the \textit{(final sign)} \#9. The very large shifts of \( 10^9 \) and \( 1.1 \cdot 10^9 \) are unnecessary here, but allow the auxiliaries to be reused later. Each integer expression produces a 10 digit result. If the resulting 16 digits start with a 0, then we need to shift the group, padding with trailing zeros.

\cs_new:Npn \__fp_sub_back_near_o:nnnnnnnnN #1#2#3#4 #5#6#7#8 #9
{ \exp_after:wN \__fp_sub_back_near_after:wNNNNw \int_value:w \__fp_int_eval:w 10#5#6 - #1#2 - 11 \exp_after:wN \__fp_sub_back_near_pack:NNNNNNw \int_value:w \__fp_int_eval:w 11#7#8 - #3#4 \exp_after:wN ; }

(End definition for \__fp_sub_back_near_o:nnnnnnnnN.)

\cs_new:Npn \__fp_sub_back_near_after:wNNNNw
{ + #1#2 ; \#3#4#5#6 \{\#7\} ; }

\cs_new:Npn \__fp_sub_back_near_after:wNNNNw #10 #1#2#3#4 #5
{ \if_meaning:w 0 #1 \exp_after:wN \__fp_sub_back_shift:wNNNNw }

\__fp_sub_back_shift:wNNNNw
This function is called with $\langle Z_1 \rangle \leq 999$. Act with \texttt{number} to trim leading zeros from $\langle Z_1 \rangle \langle Z_2 \rangle$ (we don’t do all four blocks at once, since non-zero blocks would then overflow \TeX’s integers). If the first two blocks are zero, the auxiliary receives an empty \#1 and trims \#2#30 from leading zeros, yielding a total shift between 7 and 16 to the exponent. Otherwise we get the shift from \#1 alone, yielding a result between 1 and 6. Once the exponent is taken care of, trim leading zeros from \#1#2#3#4 when \#1 is empty, the space before \#2#3 is ignored), get four blocks of 4 digits and finally clean up. Trailing zeros are added so that digits can be grabbed safely.

\begin{verbatim}
\cs_new:Npn \__fp_sub_back_shift:wnnnn \#1#2#3#4 \#5
\{
  \exp_after:wN \__fp_sub_back_shift_ii:ww
  \int_value:w \#1 \#2 0 ;
\}
\cs_new:Npn \__fp_sub_back_shift_ii:ww #1 0 ; #2#3 ;
\{
  \if_meaning:w @ #1 @ - 7
    \exp_after:wN \use_i:nnn
    \exp_after:wN \__fp_sub_back_shift_iii:NNNNNNNNw
    \int_value:w \#2#3 0 ~ 123456789;
  \else:
    - \__fp_sub_back_shift_iii:NNNNNNNNw #1 123456789;
  \fi:
  \exp_after:wN \__fp_pack_twice_four:wNNNNNNNN
  \exp_after:wN \__fp_pack_twice_four:wNNNNNNNN
  \exp_after:wN \__fp_sub_back_shift_iv:nnnnw
  \exp_after:wN ;\int_value:w
  \#1 ~ #2#3 0 ~ 0000 0000 0000 000 ;
\}
\cs_new:Npn \__fp_sub_back_shift_iii:NNNNNNNNw #1#2#3#4; #5#6#7#8
\{
  \exp_after:wN \__fp_pack_eight:wNNNNNNNN
  \exp_after:wN \__fp_sub_back_shift_iv:nnnnw
  \exp_after:wN \__fp_sub_back_shift_ii:ww \#1 \#2 0 ;
\}
\cs_new:Npn \__fp_sub_back_shift_ii:ww #1 0 ; #2#3 ;
\{
  \if_meaning:w @ #1 @ - 7
    \exp_after:wN \use_i:nnn
    \exp_after:wN \__fp_sub_back_shift_iii:NNNNNNNNw
    \int_value:w \#2#3 0 ~ 123456789;
  \else:
    - \__fp_sub_back_shift_iii:NNNNNNNNw #1 123456789;
  \fi:
  \exp_after:wN \__fp_pack_twice_four:wNNNNNNNN
  \exp_after:wN \__fp_pack_twice_four:wNNNNNNNN
  \exp_after:wN \__fp_sub_back_shift_iv:nnnnw
  \exp_after:wN \__fp_sub_back_shift_ii:ww \#1 \#2 0 ;
\}
\end{verbatim}
The easiest case is when \( x - y \) is extremely close to a power of 10, namely the first digit of \( x \) is 1, and all others vanish when subtracting \( y \). Then the \( \langle \text{r} \rangle \) \#3 and the \( \langle \text{f} \rangle \) \#4 control whether we get 1 or 0.9999999999999999.

In the usual round-to-nearest mode, we get 1 whenever the \( \langle \text{r} \rangle \) digit is less than or equal to 5 (remember that the \( \langle \text{r} \rangle \) digit is only equal to 5 if there was no further non-zero digit).

In the present case, \( x \) and \( y \) have different exponents, but \( y \) is large enough that \( x - y \) has a smaller exponent than \( x \). Decrement the exponent (with \(-1\)). Then proceed in a way similar to the near auxiliaries seen earlier, but multiplying \( x \) by 10 (\#30 and \#40 below), and with the added quirk that the \( \langle \text{r} \rangle \) digit has to be taken into account. Namely, we may have to decrease the result by one unit if \( \_\text{fp_round_neg}:\text{NNN} \) returns 1. This function expects the \( \langle \text{f} \rangle \) \#6, the last digit of 1100000000+\#40-\#2, and the \( \langle \text{r} \rangle \) digit. Instead of redoing the computation for the second argument, we note that \( \_\text{fp_round_neg}:\text{NNN} \) only cares about its parity, which is identical to that of the last digit of \#2.
\begin{verbatim}
\cs_new:Npn \__fp_sub_back_not_far_o:wwwwNN #1 - #2; #3 - #4; #5#6 {
  - 1
  \exp_after:wN \__fp_sub_back_near_after:wNNNNw
  \int_value:w \__fp_int_eval:w 1#30 - #1 - 11
  \exp_after:wN \__fp_sub_back_near_pack:NNNNNNw
  \int_value:w \__fp_int_eval:w 11 0000 0000 + #40 - #2
  - \exp_after:wN \__fp_round_neg:NNN
  \exp_after:wN #6
  \use_none:nnnnnn #2 #5
  \exp_after:wN ;
}
\cs_new:Npn \__fp_sub_back_very_far_o:wwwwNN #1#2#3#4#5#6#7 {
  \__fp_pack_eight:wNNNNNNNN
  \__fp_sub_back_very_far_ii_o:nnNwwNN { 0 #1#2#3 #4#5#6#7 };
}
\cs_new:Npn \__fp_sub_back_very_far_ii_o:nnNwwNN #1#2 ; #3 ; #4 ~ #5; #6#7 {
  \exp_after:wN \__fp_basics_pack_high:NNNNNw
  \int_value:w \__fp_int_eval:w 1#4 - #1 - 1
  \exp_after:wN \__fp_basics_pack_low:NNNNNw
  \int_value:w \__fp_int_eval:w 2#5 - #2
  - \exp_after:wN \__fp_round_neg:NNN
  \exp_after:wN #7
  \int_value:w
  \if_int_odd:w \__fp_int_eval:w #5 - #2 \__fp_int_eval_end:
    1 \else: 2 \fi:
  \int_value:w \__fp_int_eval:w \\
  \exp_after:wN ;
}
\end{verbatim}

The case where \( x - y \) and \( x \) have the same exponent is a bit more tricky, mostly because it cannot reuse the same auxiliaries. Shift the \( y \) significand by adding a leading \( 0 \). Then the logic is similar to the not_far functions above. Rounding is a bit more complicated: we have two \( \langle \text{rounding} \rangle \) digits \( #3 \) and \( #6 \) (from the decimation, and from the new shift) to take into account, and getting the parity of the main result requires a computation. The first \( \text{int_value:w} \) triggers the second one because the number is unfinished; we can thus not use 0 in place of 2 there.

\begin{verbatim}
\cs_new:Npn \__fp_sub_back_not_far_o:wwwwNN #1 - #2; #3 - #4; #5#6 {
  \__fp_sub_back_not_far_very_x:wwwwNN #1#2#3#4#5#6#7
}
\cs_new:Npn \__fp_sub_back_not_far_very_x:wwwwNN #1#2#3#4#5#6#7 {
  \__fp_pack_eight:wNNNNNNNN
  \__fp_sub_back_not_far_very_x_ii_o:nnNwwNN #1#2#3#4
  \__fp_sub_back_not_far_very_x_ii_o:nnNwwNN #1#2#3#4#5
  { 0 #1#2#3 #4#5#6#7 };
}
\cs_new:Npn \__fp_sub_back_not_far_very_x_ii_o:nnNwwNN #1#2 ; #3 ; #4 ~ #5; #6#7 {
  \exp_after:wN \__fp_basics_pack_high:NNNNNw
  \int_value:w \__fp_int_eval:w 1#4 - #1 - 1
  \exp_after:wN \__fp_basics_pack_low:NNNNNw
  \int_value:w \__fp_int_eval:w 2#5 - #2
  - \exp_after:wN \__fp_round_neg:NNN
  \exp_after:wN #7
  \int_value:w
  \if_int_odd:w \__fp_int_eval:w #5 - #2 \__fp_int_eval_end:
    1 \else: 2 \fi:
  \int_value:w \__fp_int_eval:w \\
  \exp_after:wN ;
}
\end{verbatim}

(End definition for \( \__fp_sub_back_not_far_o:wwwwNN \) and \( \__fp_sub_back_not_far_very_x:wwwwNN \).

\section{Multiplication}
\subsection{Signs, and special numbers}
We go through an auxiliary, which is common with \( \__fp_/_o:ww \). The first argument is the operation, used for the invalid operation exception. The second is inserted in a formula to dispatch cases slightly differently between multiplication and division. The
third is the operation for normal floating points. The fourth is there for extra cases needed in \_\_fp_/\_o:ww.

\cs_new:cpn { \_\_fp_*_o:ww }
{ \_\_fp_mul_cases_o:NnNnww *
 \_\_fp_mul_npos_o:Nww 
{ - 2 + }
}

(End definition for \_\_fp_*_o:ww.)

\_\_fp_mul_cases_o:nNnnww Split into 10 cases (12 for division). If both numbers are normal, go to case 0 (same sign) or case 1 (opposite signs): in both cases, call \_\_fp_mul_npos_o:Nww to do the work. If the first operand is nan, go to case 2, in which the second operand is discarded; if the second operand is nan, go to case 3, in which the first operand is discarded (note the weird interaction with the final test on signs). Then we separate the case where the first number is normal and the second is zero: this goes to cases 4 and 5 for multiplication, 10 and 11 for division. Otherwise, we do a computation which dispatches the products 0×0 = 0×1 = 1×0 = 0 to case 4 or 5 depending on the combined sign, the products 0×∞ and ∞×0 to case 6 or 7 (invalid operation), and the products 1×∞ = ∞×1 = ∞×∞ = ∞ to cases 8 and 9. Note that the code for these two cases (which return ±∞) is inserted as argument #4, because it differs in the case of divisions.

\cs_new:Npn \_\_fp_mul_cases_o:NnNnww
#1#2#3#4 \s__fp \__fp_chk:w #5#6#7; \s__fp \__fp_chk:w #8#9
{ \if_case:w \__fp_int_eval:w
  \if_int_compare:w #5 #8 = 11 ~
  1
  \else:
  \if_meaning:w 3 #8
  3
  \else:
  \if_meaning:w 3 #5
  2
  \else:
  \if_int_compare:w #5 #8 = 10 ~
  9 #2 - 2
  \else:
  (#5 #2 #8) / 2 * 2 + 7
  \fi:
  \fi:
  \fi:
  \fi:
  \if_meaning:w #6 #9 - 1 \fi:
  \_\_fp_int_eval_end:
  \_\_fp_case_use:nw { #3 0 }
  \or: \_\_fp_case_use:nw { #3 2 }
  #10 \or: \_\_fp_case_return_i_o:ww
  \or: \_\_fp_case_return_ii_o:ww
  \or: \_\_fp_case_return_o:Nww \c_zero_fp
  \or: \_\_fp_case_return_o:Nww \c_minus_zero_fp
1062
72.2.2 Absolute multiplication

In this subsection, we perform the multiplication of two positive normal numbers.

\begin{verbatim}
\cs_new:Npn \__fp_mul_npos_o:Nww #1 \s__fp \__fp_chk:w #2 #3 #4 #5 ; \s__fp \__fp_chk:w #6 #7 #8 #9 ;
\exp_after:wN \__fp_mul_significand_o:nnnnNnnnn #1 #2 #3 #4 #5 #6 #7 #8 #9 ;
\end{verbatim}

Note the three semicolons at the end of the definition. One is for the last \_\_fp_mul_significand_drop:NNNNNw; one is for \_\_fp_round_digit:Nw later on; and one, preceded by \exp_after:wN, which is correctly expanded (within an \_\_fp_int_eval:w), is used by \_\_fp_basics_pack_low:NNNNNw.

The product of two 16 digit integers has 31 or 32 digits, but it is impossible to know which one before computing. The place where we round depends on that number of digits, and may depend on all digits until the last in some rare cases. The approach is thus to compute the 5 first blocks of 4 digits (the first one is between 100 and 9999 inclusive), and a compact version of the remaining 3 blocks. Afterwards, the number of digits is known, and we can do the rounding within yet another set of \_\_fp_int_eval:w.

\begin{verbatim}
\exp_after:wN \__fp_mul_significand_test_f:NNN
\exp_after:wN #5
\int_value:w \__fp_int_eval:w 99990000 + #1*#6 +
\end{verbatim}
\exp_after:wN \__fp_mul_significand_keep:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #1*#7 + #2*#6 + #3*#7 + #4*#6 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #1*#9 + #2*#8 + #3*#7 + #4*#6 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #2*#9 + #3*#8 + #4*#7 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #3*#9 + #4*#8 + #4*#9 ;
\exp_after:wN ;
\exp_after:wN \__fp_mul_significand_test_f:NNN
\__fp_mul_significand_test_f:NNN
\__fp_mul_significand_large_f:NwwNNNN
\__fp_mul_significand_large_f:NwwNNNN
\__fp_mul_significand_test_f:NNN (sign) 1 \langle digits 1-8 \rangle; \langle digits 9-12 \rangle; \langle digits 13-16 \rangle; + \langle digits 17-20 \rangle + \langle digits 21-24 \rangle + \langle digits 25-28 \rangle + \langle digits 29-32 \rangle; \exp_after:wN ;
\exp_after:wN \__fp_basics_pack_high:NNNNNw
\int_value:w \__fp_int_eval:w 1#2
\exp_after:wN \__fp_basics_pack_low:NNNNNw
\int_value:w \__fp_int_eval:w 1#3#4#5#6#7
+ \exp_after:wN \__fp_round:NNN
\exp_after:wN #1
1064
\cs_new:Npn \__fp_mul_significand_test_f:NNN #1 #2 #3
{\if_meaning:w 0 #3
\exp_after:wN \__fp_mul_significand_small_f:NwwwN
\else:
\exp_after:wN \__fp_mul_significand_large_f:NwNNNN
\fi:
#1 #3}
\cs_new:Npn \__fp_mul_significand_large_f:NwwNNNN
In this branch, \langle digit \rangle is non-zero. The result is thus \langle digits 1-16 \rangle, plus some rounding which depends on the digits 16, 17, and whether all subsequent digits are zero or not. Here, \__fp_round_digit:Nw takes digits 17 and further (as an integer expression), and replaces it by a \langle rounding digit \rangle, suitable for \__fp_round:Nw.
\cs_new:Npn \__fp_mul_significand_large_f:NwNNNN #1 #2; #3; #4#5#6#7; +
{\exp_after:wN \__fp_basics_pack_high:NNNNNw
\int_value:w \__fp_int_eval:w 1#2
\exp_after:wN \__fp_basics_pack_low:NNNNNw
\int_value:w \__fp_int_eval:w 1#3#4#5#6#7
+ \exp_after:wN \__fp_round:NNN
\exp_after:wN #1
In this branch, \( \langle \text{digit 1} \rangle \) is zero. Our result is thus \( \langle \text{digits 2–17} \rangle \), plus some rounding which depends on the digits 17, 18, and whether all subsequent digits are zero or not. The 8 digits 1#3 are followed, after expansion of the small_pack auxiliary, by the next digit, to form a 9 digit number.

\[
\begin{align*}
&\exp_after:wN \_\text{fp}_\text{mul}_\text{significand}\_\text{large}_f:NNwwwN \exp_after:wN #7 \exp_after:wN \_\text{fp}_\text{round}_\text{digit}:Nw \\
&\text{(End definition for }\_\text{fp}_\text{mul}_\text{significand}\_\text{large}_f:NNwwwN).}
\end{align*}
\]

\section*{72.3 Division}

\subsection*{72.3.1 Signs, and special numbers}

Time is now ripe to tackle the hardest of the four elementary operations: division.

\[
\begin{align*}
&\text{Filtering special floating point is very similar to what we did for multiplications, with a few variations. Invalid operation exceptions display }/\text{ rather than }*. \text{ In the formula for dispatch, we replace } - 2 + \text{ by } -. \text{ The case of normal numbers is treated using }\_\text{fp}_-\div\_\text{npos}_o:Nww \text{ rather than }\_\text{fp}_\_\text{mul}_\_\text{npos}_o:Nww. \text{ There are two additional cases: if the first operand is normal and the second is a zero, then the division by zero exception is raised: cases 10 and 11 of the }\text{if}_\text{case}:w \text{ construction in }\_\text{fp}_\_\text{mul}_\_\text{cases}_o:NNwww \text{ are provided as the fourth argument here.}
\end{align*}
\]
We want to compute \( A/Z \). As for multiplication, \texttt{\_fp\_sanitize:Nw} checks for overflow or underflow; we provide it with the \langle final sign \rangle, and an integer expression in which we compute the exponent. We set up the arguments of \texttt{\_fp\_div\_significand\_i_o:wnnw}, namely an integer \langle y \rangle obtained by adding 1 to the first 5 digits of \( Z \) (explanation given soon below), then the four \{\langle A_i \rangle\}, then the four \{\langle Z_i \rangle\}, a semi-colon, and the \langle final sign \rangle, used for rounding at the end.

\begin{verbatim}
\cs_new:Npn \_fp_div_npos_o:Nww
\s__fp \__fp_chk:w 1 #1 \s__fp \__fp_chk:w 1 #2 #3 #4 ; \s__fp \__fp_chk:w 1 #5 #6 #7#8#9;
{ \exp_after:wN \__fp\_sanitize:Nw
\exp_after:wN \__fp\_chk:w 1 #1 \int_value:w \__fp\_int_eval:w
#3 - \#6
\exp_after:wN \__fp\_div\_significand\_i_o:wnnw
\int_value:w \__fp\_int_eval:w \#7 \use_i:nnnn \#8 + 1 ;
\#4}
\#7\#8\#9 ;
\#1
}
\end{verbatim}

(End definition for \texttt{\_fp\_div\_npos\_o:Nww}).

### 72.3.2 Work plan

In this subsection, we explain how to avoid overflowing \TeX{}’s integers when performing the division of two positive normal numbers.

We are given two numbers, \( A = 0.A_1A_2A_3A_4 \) and \( Z = 0.Z_1Z_2Z_3Z_4 \), in blocks of 4 digits, and we know that the first digits of \( A_1 \) and of \( Z_1 \) are non-zero. To compute \( A/Z \), we proceed as follows.

- Find an integer \( Q_A \simeq 10^4A/Z \).
- Replace \( A \) by \( B = 10^4A - Q_AZ \).
- Find an integer \( Q_B \simeq 10^4B/Z \).
- Replace \( B \) by \( C = 10^4B - Q_BZ \).
- Find an integer \( Q_C \simeq 10^4C/Z \).
- Replace \( C \) by \( D = 10^4C - Q_CZ \).
- Find an integer \( Q_D \simeq 10^4D/Z \).
- Consider \( E = 10^4D - Q_DZ \), and ensure correct rounding.
The result is then \[ Q = 10^{-4}Q_A + 10^{-8}Q_B + 10^{-12}Q_C + 10^{-16}Q_D + \text{rounding}. \]

Since the \( Q_i \) are integers, \( B, C, D, \) and \( E \) are all exact multiples of \( 10^{-16} \), in other words, computing with 16 digits after the decimal separator yields exact results. The problem is the risk of overflow: in general \( B, C, D, \) and \( E \) may be greater than 1.

Unfortunately, things are not as easy as they seem. In particular, we want all intermediate steps to be positive, since negative results would require extra calculations at the end. This requires that \( Q_A \leq 10^4A/Z \) etc. A reasonable attempt would be to define \( Q_A \) as

\[
\texttt{\int_eval:n} \left\{ \frac{A_1A_2}{Z_1+1} - 1 \right\} \leq 10^4A/Z
\]

Subtracting 1 at the end takes care of the fact that \( \varepsilon\text{-\TeX}'s \texttt{\_fp_int_eval:w} \) rounds divisions instead of truncating (really, \( 1/2 \) would be sufficient, but we work with integers). We add 1 to \( Z_1 \) because \( Z_1 \leq 10^4Z < Z_1 + 1 \) and we need \( Q_A \) to be an underestimate. However, we are now underestimating \( Q_A \) too much: it can be wrong by up to 100, for instance when \( Z = 0.1 \) and \( A \approx 1 \). Then \( B \) could take values up to 10 (maybe more), and a few steps down the line, we would run into arithmetic overflow, since \( \text{\TeX} \) can only handle integers less than roughly \( 2 \cdot 10^9 \).

A better formula is to take

\[
Q_A = \texttt{\int_eval:n} \left\{ \frac{10 \cdot A_1A_2}{10^{-3} \cdot Z_1Z_2} + 1 \right\}.
\]

This is always less than \( 10^9A/(10^5Z) \), as we wanted. In words, we take the 5 first digits of \( Z \) into account, and the 8 first digits of \( A \), using 0 as a 9-th digit rather than the true digit for efficiency reasons. We shall prove that using this formula to define all the \( Q_i \) avoids any overflow. For convenience, let us denote

\[
y = \left\lfloor 10^{-3} \cdot Z_1Z_2 \right\rfloor + 1,
\]

so that, taking into account the fact that \( \varepsilon\text{-\TeX} \) rounds ties away from zero,

\[
Q_A = \left\lfloor \frac{A_1A_20}{y} - \frac{1}{2} \right\rfloor > \frac{A_1A_20}{y} - \frac{3}{2}.
\]

Note that \( 10^4 < y \leq 10^5 \), and \( 999 \leq Q_A \leq 999899 \). Also note that this formula does not cause an overflow as long as \( A < (2^{31} - 1)/10^9 \approx 2.147\cdots, \) since the numerator involves an integer slightly smaller than \( 10^9A \).

Let us bound \( B \):

\[
10^5B = A_1A_20 + 10 \cdot 0 \cdot A_3A_4 - 10 \cdot Z_1Z_2Z_3Z_4 \cdot Q_A
\]
\[
< A_1A_20 \cdot \left( 1 - 10 \cdot \frac{Z_1Z_2Z_3Z_4}{y} \right) + \frac{3}{2} 10 \cdot Z_1Z_2Z_3Z_4 + 10
\]
\[
\leq A_1A_20 \cdot \frac{y - 10 \cdot Z_1Z_2Z_3Z_4}{y} + \frac{3}{2} y + 10
\]
\[
\leq \frac{A_1A_20 \cdot 1}{y} + \frac{3}{2} y + 10 \leq \frac{10^9A}{y} + 1.6 \cdot y.
\]
At the last step, we hide 10 into the second term for later convenience. The same reasoning yields

\[ 10^5 B < 10^9 A/y + 1.6y, \]
\[ 10^5 C < 10^9 B/y + 1.6y, \]
\[ 10^5 D < 10^9 C/y + 1.6y, \]
\[ 10^5 E < 10^9 D/y + 1.6y. \]

The goal is now to prove that none of $B$, $C$, $D$, and $E$ can go beyond \((2^{31} - 1)/10^9 = 2.147\ldots\).

Combining the various inequalities together with $A < 1$, we get

\[ 10^5 B < 10^9/y + 1.6y, \]
\[ 10^5 C < 10^{13}/y^2 + 1.6(y + 10^4), \]
\[ 10^5 D < 10^{17}/y^3 + 1.6(y + 10^4 + 10^8/y), \]
\[ 10^5 E < 10^{21}/y^4 + 1.6(y + 10^4 + 10^8/y + 10^{12}/y^2). \]

All of those bounds are convex functions of $y$ (since every power of $y$ involved is convex, and the coefficients are positive), and thus maximal at one of the end-points of the allowed range $10^4 < y \leq 10^5$. Thus,

\[ 10^5 B < \max(1.16 \cdot 10^5, 1.7 \cdot 10^5), \]
\[ 10^5 C < \max(1.32 \cdot 10^5, 1.77 \cdot 10^5), \]
\[ 10^5 D < \max(1.48 \cdot 10^5, 1.777 \cdot 10^5), \]
\[ 10^5 E < \max(1.64 \cdot 10^5, 1.7777 \cdot 10^5). \]

All of those bounds are less than $2.147 \cdot 10^5$, and we are thus within \TeX’s bounds in all cases!

We later need to have a bound on the $Q_i$. Their definitions imply that $Q_A < 10^9 A/y - 1/2 < 10^9 A$ and similarly for the other $Q_i$. Thus, all of them are less than 177770.

The last step is to ensure correct rounding. We have

\[ A/Z = \sum_{i=1}^{4} (10^{-4i} Q_i) + 10^{-16} E/Z \]

exactly. Furthermore, we know that the result is in $[0.1, 10)$, hence will be rounded to a multiple of $10^{-16}$ or of $10^{-15}$, so we only need to know the integer part of $E/Z$, and a “rounding” digit encoding the rest. Equivalently, we need to find the integer part of $2E/Z$, and determine whether it was an exact integer or not (this serves to detect ties). Since

\[ \frac{2E}{Z} = \frac{2 \cdot 10^5 E}{10^5 Z} \leq \frac{2 \cdot 10^5 E}{10^5} < 36, \]

1068
this integer part is between 0 and 35 inclusive. We let $\varepsilon$-\TeX{} round

$$P = \langle \text{\texttt{int\_eval:n}} \left\{ 2 \cdot E_1 E_2 \over Z_1 Z_2 \right\} \rangle,$$

which differs from $2E/Z$ by at most

$$\frac{1}{2} + 2 \left| {E \over Z} - {E \over 10^{-8} Z_1 Z_2} \right| + 2 \left| {10^8 E - E_1 E_2 \over Z_1 Z_2} \right| < 1,$$

(1/2 comes from $\varepsilon$-\TeX{}’s rounding) because each absolute value is less than $10^{-7}$. Thus $P$ is either the correct integer part, or is off by 1; furthermore, if $2E/Z$ is an integer, $P = 2E/Z$. We will check the sign of $2E - PZ$. If it is negative, then $E/Z \in (P - 1)/2, P/2)$. If it is zero, then $E/Z \in (P/2, (P - 1)/2)$. In each case, we know how to round to an integer, depending on the parity of $P$, and the rounding mode.

### 72.3.3 Implementing the significand division

```
\_\_fp\_div\_significand\_i\_o:wwnnn \{ \langle A_1 \rangle \} \{ \langle A_2 \rangle \} \{ \langle A_3 \rangle \} \{ \langle A_4 \rangle \} \{ \langle Z_1 \rangle \} \{ \langle Z_2 \rangle \} \{ \langle Z_3 \rangle \} \{ \langle Z_4 \rangle \} ; \langle \text{sign} \rangle \}
```

Compute $10^6 + Q_A$ (a 7 digit number thanks to the shift), unbrace $\langle A_1 \rangle$ and $\langle A_2 \rangle$, and prepare the $\langle \text{continuation} \rangle$ arguments for 4 consecutive calls to $\langle \_\_fp\_div\_significand\_calc:wwnnnn \rangle$. Each of these calls needs $\langle y \rangle$ $(\#1)$, and it turns out that we need post-expansion there, hence $\langle \text{\texttt{int\_value:w}} \rangle$. Here, #4 is six brace groups, which give the six first n-type arguments of the calc function.

```
\cs_new:Np \_\_fp\_div\_significand\_i\_o:wwnnn \#1 ; \#2\#3 \#4 ;
{
\exp_after:wN \_\_fp\_div\_significand\_test_o:w
\exp_after:wN \_\_fp\_int\_eval:w
\int_value:w \_\_fp\_int\_eval:w 999999 + \#2 \#3 0 / \#1 ;
\#2 \#3 ;
\#4
{ \exp_after:wN \_\_fp\_div\_significand\_ii:wwwn \int_value:w \#1 }
{ \exp_after:wN \_\_fp\_div\_significand\_ii:wwwn \int_value:w \#1 }
{ \exp_after:wN \_\_fp\_div\_significand\_ii:wwwn \int_value:w \#1 }
{ \exp_after:wN \_\_fp\_div\_significand\_ii:wwwn \int_value:w \#1 }
}
```

(End definition for $\langle \_\_fp\_div\_significand\_i\_o:wwnnn \rangle$)

```
\_\_fp\_div\_significand\_calc:wwnnnnnnn \{ 10^6 + Q_A \} \langle A_1 \rangle \langle A_2 \rangle ; \langle A_3 \rangle \{ \langle Z_1 \rangle \} \{ \langle Z_2 \rangle \} \{ \langle Z_3 \rangle \} \{ \langle Z_4 \rangle \} \langle \text{continuation} \rangle
```

expands to

$$\langle 10^6 + Q_A \rangle \langle \text{continuation} \rangle ; \langle B_1 \rangle \langle B_2 \rangle ; \langle B_3 \rangle \{ \langle B_4 \rangle \} \{ \langle Z_1 \rangle \} \{ \langle Z_2 \rangle \} \{ \langle Z_3 \rangle \} \{ \langle Z_4 \rangle \}$$

where $B = 10^4 A - Q_A \cdot Z$. This function is also used to compute $C, D, E$ (with the input shifted accordingly), and is used in \texttt{i3fp-expo}.

We know that $0 < Q_A < 1.8 \cdot 10^3$, so the product of $Q_A$ with each $Z_i$ is within \TeX’s bounds. However, it is a little bit too large for our purposes: we would not be able to
use the usual trick of adding a large power of 10 to ensure that the number of digits is fixed.

The bound on $Q_A$, implies that $10^6 + Q_A$ starts with the digit 1, followed by 0 or 1. We test, and call different auxiliaries for the two cases. An earlier implementation did the tests within the computation, but since we added a $\langle$continuation$\rangle$, this is not possible because the macro has 9 parameters.

The result we want is then (the overall power of 10 is arbitrary):

$$10^{-4}(#2 - #1 \cdot #5 - 10 \cdot \langle i \rangle \cdot #6) + 10^{-8}(#3 - #1 \cdot #6 - 10 \cdot \langle i \rangle \cdot #7)$$

$$+ 10^{-12}(#4 - #1 \cdot #7 - 10 \cdot \langle i \rangle \cdot #8) + 10^{-16}(-#1 \cdot #8),$$

where $\langle i \rangle$ stands for the $10^5$ digit of $Q_A$, which is 0 or 1, and #1, #2, etc. are the parameters of either auxiliary. The factors of 10 come from the fact that $Q_A = 10 \cdot 10^4 \cdot \langle i \rangle + #1$. As usual, to combine all the terms, we need to choose some shifts which must ensure that the number of digits of the second, third, and fourth terms are each fixed. Here, the positive contributions are at most $10^8$ and the negative contributions can go up to $10^9$. Indeed, for the auxiliary with $\langle i \rangle = 1$, #1 is at most 80000, leading to contributions of at worse $-8 \cdot 10^4$, while the other negative term is very small < $10^6$ (except in the first expression, where we don’t care about the number of digits); for the auxiliary with $\langle i \rangle = 0$, #1 can go up to 99999, but there is no other negative term. Hence, a good choice is $2 \cdot 10^9$, which produces totals in the range $[10^9, 2.1 \cdot 10^9]$. We are flirting with T\TeX{}’s limits once more.

\begin{verbatim}
\cs_new:Npn \__fp_div_significand_calc:wwnnnnnnn {#1}{#2}{#3}{#4}{#5}{#6}{#7}{#8}
\int_value:w \__fp_int_eval:w \c__fp_Bigg_leading_shift_int + #2 - #1 * #5 - #60
\exp_after:wN \__fp_pack_Bigg:NNNNNNw \int_value:w \__fp_int_eval:w \c__fp_Bigg_middle_shift_int + #3 - #1 * #6 - #70
\exp_after:wN \__fp_pack_Bigg:NNNNNNw \int_value:w \__fp_int_eval:w \c__fp_Bigg_middle_shift_int + #4 - #1 * #7 - #80
\exp_after:wN \__fp_pack_Bigg:NNNNNNw \int_value:w \__fp_int_eval:w \c__fp_Bigg_trailing_shift_int - #1 * #8 ;
\end{verbatim}
\__fp_div_significand_ii:wwn \__fp_div_significand_ii:wwn
\langle y \rangle ; \langle E_1 \rangle ; \{ \langle E_2 \rangle \} \{ \langle E_3 \rangle \} \{ \langle E_4 \rangle \}
\{ \langle Z_1 \rangle \} \{ \langle Z_2 \rangle \} \{ \langle Z_3 \rangle \} \{ \langle Z_4 \rangle \} \langle \text{sign} \rangle

We compute $P \approx 2E/Z$ by rounding $2E_1E_2/Z_1Z_2$. Note the first 0, which multiplies $Q_D$ by 10: we later add (roughly) $5 \cdot P$, which amounts to adding $P/2 \approx E/Z$ to $Q_D$, the appropriate correction from a hypothetical $Q_E$.  

```latex
\textbf{1071}
```
This adds to the current expression \((10^7 + 10 \cdot Q_D)\) a contribution of \(5 \cdot P + \text{sign}(T)\) with \(T = 2E - PZ\). This amounts to adding \(P/2\) to \(Q_D\), with an extra \(\text{(rounding)}\) digit. This \(\text{(rounding)}\) digit is 0 or 5 if \(T\) does not contribute, \text{i.e., if} \(0 = T = 2E - PZ\), in other words if \(10^{16} A/Z\) is an integer or half-integer. Otherwise it is in the appropriate range, \([1, 4]\) or \([6, 9]\). This is precise enough for rounding purposes (in any mode). It seems an overkill to compute \(T\) exactly as I do here, but I see no faster way right now.

Once more, we need to be careful and show that the calculation \#1 \#6 \#7 below does not cause an overflow: naively, \(P\) can be up to 35, and \#6 \#7 up to \(10^8\), but both cannot happen simultaneously. To show that things are fine, we split in two (non-disjoint) cases.

- For \(P < 10\), the product obeys \(P \cdot \#6 \#7 < 10^8 \cdot P < 10^9\).
- For large \(P \geq 3\), the rounding error on \(P\), which is at most 1, is less than a factor of 2, hence \(P \leq 4E/Z\). Also, \#6 \#7 \(\leq 10^8 \cdot Z\), hence \(P \cdot \#6 \#7 \leq 4E \cdot 10^8 < 10^9\).

Both inequalities could be made tighter if needed.

Note however that \(P \#8 \#9\) may overflow, since the two factors are now independent, and the result may reach \(3.5 \cdot 10^9\). Thus we compute the two lower levels separately. The rest is standard, except that we use \(+\) as a separator (ending integer expressions explicitly). \(T\) is negative if the first character is \(-\), it is positive if the first character is neither \(0\) nor \(-\). It is also positive if the first character is \(0\) and second argument of \text{	extbackslash__fp\_div\_significand\_vi:Nw}, a sum of several terms, is also zero. Otherwise, there was an exact agreement: \(T = 0\).

\begin{verbatim}
\cs_new:Npn \__fp_div_significand_iv:wwnnnnnnn #1; #2;#3#4#5 #6#7#8#9
25311 \exp_after:wN \__fp_div_significand_vi:Nw
25312 \int_value:w \__fp_int_eval:w -20 + 2*#2#3 - #1*#6#7 +
25313 \exp_after:wN \__fp_div_significand_v:NN
25314 \int_value:w \__fp_int_eval:w 199980 + 2*#4 - #1*#8 +
25315 \exp_after:wN \__fp_div_significand_v:NN
25316 \int_value:w \__fp_int_eval:w 200000 + 2*#5 - #1*#9 ;
25317 \__fp_div_significand_pack:NNN
\end{verbatim}

At this stage, we are in the following situation: \TeX is in the process of expanding several integer expressions, thus functions at the bottom expand before those above.
Here, $\varepsilon = \text{sign}(T)$ is 0 in case $2E = PZ$, 1 in case $2E > PZ$, which means that $P$ was the correct value, but not with an exact quotient, and $-1$ if $2E < PZ$, i.e., $P$ was an overestimate. The packing function we define now does nothing special: it removes the $10^6$ and carries two digits (for the $10^5$’s and the $10^4$’s).

\[\text{Here, } \varepsilon = \text{sign}(T) \text{ is 0 in case } 2E = PZ, \text{ 1 in case } 2E > PZ, \text{ which means that } P \text{ was the correct value, but not with an exact quotient, and } -1 \text{ if } 2E < PZ, \text{ i.e., } P \text{ was an overestimate. The packing function we define now does nothing special: it removes the } 10^6 \text{ and carries two digits (for the } 10^5 \text{'s and the } 10^4 \text{'s).}\]

\[\text{25339 } \text{\texttt{cs\_new:Hp}} \text{ \texttt{\_fp\_div\_significand\_pack:NNN}} \text{ 1 #1 #2 } \{ \text{ + #1 #2 } ; \}
\]

(End definition for \texttt{\_fp\_div\_significand\_pack:NNN}.)

\[\text{25340 } \text{\texttt{\_fp\_div\_significand\_test\_o:w}} \text{ 10 } \langle 5d \rangle ; \langle 4d \rangle ; \langle 4d \rangle ; \langle 5d \rangle ; \langle \text{sign} \rangle
\]

(End definition for \texttt{\_fp\_div\_significand\_test\_o:w}.)

\[\text{25341 } \text{\texttt{\_fp\_div\_significand\_small\_o:wwwNNNNw}} \text{ 0 } \langle 4d \rangle ; \langle 4d \rangle ; \langle 5d \rangle ; \langle \text{final sign} \rangle
\]

(End definition for \texttt{\_fp\_div\_significand\_small\_o:wwwNNNNw}.)

Standard use of the functions \texttt{\_fp\_basics\_pack\_low:NNNNw} and \texttt{\_fp\_basics\_pack\_high:NNNNw}. We finally get to use the (final sign) which has been sitting there for a while.

\[\text{25345 } \text{\texttt{\_fp\_round:NNN}} \text{ #9 #7 #8 } \text{\texttt{;}} \]

(End definition for \texttt{\_fp\_round:NNN}.)

We know that the final result cannot reach 10, hence 1#1#2, together with contributions from the level below, cannot reach $2 \cdot 10^9$. For rounding, we build the (rounding digit) from the last two of our 18 digits.
72.4 Square root

\_\_fp_sqrt_o:w

Zeros are unchanged: \( \sqrt{-0} = -0 \) and \( \sqrt{+0} = +0 \). Negative numbers (other than \(-0\)) have no real square root. Positive infinity, and \text{nan}, are unchanged. Finally, for normal positive numbers, there is some work to do.

\_\_fp_sqrt_npos_o:w

Prepare \_\_fp_sanitize:Nw to receive the final sign \( 0 \) (the result is always positive) and the exponent, equal to half of the exponent \#1 of the argument. If the exponent \#1 is even, find a first approximation of the square root of the significand \( 10^8a_1+a_2=10^8\#2\#3\#4\#5 \) through Newton's method, starting at \( x = 57234133 \approx 10^{7.75} \). Otherwise, first shift the significand of the argument by one digit, getting \( a_1' \in [10^6, 10^7) \) instead of \( [10^7, 10^8) \), then use Newton's method starting at \( 17782794 \approx 10^{7.25} \).
Newton’s method maps $x \mapsto ([x + [10^8 a_1/x]]/2)$ in each iteration, where $[b/c]$ denotes \TeX{}’s division. This division rounds the real number $b/c$ to the closest integer, rounding ties away from zero, hence when $c$ is even, $b/c - 1/2 + 1/c \leq [b/c] \leq b/c + 1/2$ and when $c$ is odd, $b/c - 1/2 + 1/(2c) \leq [b/c] \leq b/c + 1/2 - 1/(2c)$. For all $c$, $b/c - 1/2 + 1/(2c) \leq [b/c] \leq b/c + 1/2$.

Let us prove that the method converges when implemented with \TeX{}’s division, for any $10^6 \leq a_1 < 10^9$ and starting value $10^6 \leq x < 10^9$. Using the inequalities above and the arithmetic–geometric inequality $(x/t)/2 \geq \sqrt{xt}$ for $t = 10^8 a_1/x$, we find

$$x' = \left[\frac{[x + [10^8 a_1/x]]}{2}\right] \geq \frac{1}{2} \left(x + \frac{10^8 a_1}{2} - \frac{1}{2} + \frac{2}{x}\right) = \sqrt{10^8 a_1} - \frac{1}{4} + \frac{1}{4x}.$$  

After any step of iteration, we thus have $\delta = x - \sqrt{10^8 a_1} \geq -0.25 + 0.25 \cdot 10^{-8}$. The new difference $\delta' = x' - \sqrt{10^8 a_1}$ after one step is bounded above as

$$x' - \sqrt{10^8 a_1} \leq \frac{\delta}{2} - \frac{\delta}{2} \sqrt{10^8 a_1} + \frac{3}{4}.$$  

For $\delta > 3/2$, this last expression is $\leq \delta/2 + 3/4 < \delta$, hence $\delta$ decreases at each step: since all $x$ are integers, $\delta$ must reach a value $1/4 < \delta \leq 3/2$. In this range of values, we get $\delta' \leq \frac{3}{4} \sqrt{10^8 a_1} + \frac{3}{4} \leq 0.75 + 1.125 \cdot 10^{-7}$. We deduce that the difference $\delta = x - \sqrt{10^8 a_1}$ eventually reaches a value in the interval $[-0.25 + 0.25 \cdot 10^{-8}, 0.75 + 11.25 \cdot 10^{-8}]$, whose width is $1 + 11 \cdot 10^{-8}$. The corresponding interval for $x$ may contain two integers, hence $x$ might oscillate between those two values.

However, the fact that $x \mapsto x - 1$ and $x - 1 \mapsto x$ puts stronger constraints, which are not compatible: the first implies $x + [10^8 a_1/x] \leq 2x - 2$ hence $10^8 a_1/x \leq x - 3/2$, while the second implies $x - 1 + [10^8 a_1/(x - 1)] \geq 2x - 1$ hence $10^8 a_1/(x - 1) \geq x - 1/2$. Combining the two inequalities yields $x^2 - 3x/2 \geq 10^8 a_1 \geq x - 3x/2 + 1/2$, which cannot hold. Therefore, the iteration always converges to a single integer $x$. To stop the iteration when two consecutive results are equal, the function \texttt{\_\_fp_sqrt\_Newton\_o:w} receives the newly computed result as \#1, the previous result as \#2, and $a_1$ as \#3. Note that \TeX{} combines the computation of a multiplication and a following division, thus avoiding overflow in \#3 * 100000000 / \#1. In any case, the result is within $[10^7, 10^8]$. 

1075
The choice of \( a \) because this holds for the initial \( y \) underestimates \( \sqrt{a} \) and \( \sqrt{a} \). From the inequalities shown earlier, we know that \( a = 10^{-8}a_1 + 10^{-16}a_2 + 10^{-17}a_3 \). We set up an initial underestimate\[ y = (x - 1)10^{-8} + 0.2499998875 \cdot 10^{-8} \lesssim \sqrt{a}. \]

From the inequalities shown earlier, we know that \( y \leq \sqrt{10^{-8}a_1} \leq \sqrt{a} \) and that \( \sqrt{10^{-8}a_1} \leq y + 10^{-8} + 11 \cdot 10^{-16} \) hence (using \( 0.1 \leq y \leq \sqrt{a} \leq 1 \))

\[
\begin{align*}
\sqrt{a} - y^2 & \leq 10^{-8}a_1 + 10^{-8} - y^2 \leq (y + 10^{-8} + 11 \cdot 10^{-16})^2 - y^2 + 10^{-8} < 3.2 \cdot 10^{-8},
\end{align*}
\]

and \( \sqrt{a} - y = (a - y^2)/(\sqrt{a} + y) \leq 16 \cdot 10^{-8} \). Next, \( \text{\_\_fp_sqrt_auxii_o:NnnnnnnN} \) is called several times to get closer and closer underestimates of \( \sqrt{a} \). By construction, the underestimates \( y \) are always increasing, \( a - y^2 < 3.2 \cdot 10^{-8} \) for all. Also, \( y < 1 \).

This receives a continuation function \#1, then five blocks of 4 digits for \( y \), then two 8-digit blocks and a single digit for \( a \). A common estimate of \( \sqrt{a} - y = (a - y^2)/(\sqrt{a} + y) \) is \( (a - y^2)/(2y) \), which leads to alternating overestimates and underestimates. We tweak this, to only work with underestimates (no need then to worry about signs in the computation). Each step finds the largest integer \( j \leq 6 \) such that \( 10^{10j}(a - y^2) < 2 \cdot 10^8 \), then computes the integer (with \( \varepsilon \)-FPX's rounding division)

\[
10^{10j}z = \left[ \left( 10^{10j}(a - y^2) \right) - 257 \right] \cdot 0.5 \cdot 10^8 \left/ \left[ 10^{8}y + 1 \right] \right. \right].
\]

The choice of \( j \) ensures that \( 10^{10j}z < 2 \cdot 10^8 \cdot 0.5 \cdot 10^8/10^7 = 10^9 \), thus \( 10^9 + 10^{10j}z \) has exactly 10 digits, does not overflow FPX's integer range, and starts with 1. Incidentally, since all \( a - y^2 \leq 3.2 \cdot 10^{-8} \), we know that \( j \geq 3 \).

Let us show that \( z \) is an underestimate of \( \sqrt{a} - y \). On the one hand, \( \sqrt{a} - y \leq 16 \cdot 10^{-8} \) because this holds for the initial \( y \) and values of \( y \) can only increase. On the other hand, the choice of \( j \) implies that \( \sqrt{a} - y \leq 5(\sqrt{a} + y)(\sqrt{a} - y) = 5(a - y^2) < 10^{9-10j} \). For \( j = 3 \),
the first bound is better, while for larger \( j \), the second bound is better. For all \( j \in [3, 6] \), we find \( \sqrt{a} - y < 16 \cdot 10^{-2j} \). From this, we deduce that

\[
10^{k_j}(\sqrt{a} - y) = \frac{10^{k_j}(a - y^2 - (\sqrt{a} - y)^2)}{2y} \geq \frac{[10^{k_j}(a - y^2)] - 257}{2 \cdot 10^{-8}[10^y + 1]} + \frac{1}{2}
\]

where we have replaced the bound \( 10^{k_j}(16 \cdot 10^{-2j}) = 256 \) by \( 257 \) and extracted the corresponding term \( 1/(2 \cdot 10^{-8}[10^y + 1]) \geq 1/2 \). Given that \( c \cdot T_{2X} \)’s integer division obeys \( [b/c] \leq b/c + 1/2 \), we deduce that \( 10^{k_j}z \leq 10^{k_j}(\sqrt{a} - y) \), hence \( y + z \leq \sqrt{a} \) is an underestimate of \( \sqrt{a} \), as claimed. One implementation detail: because the computation involves \(-\#4\#4 - 2\#3\#5 - 2\#2\#6\) which may be as low as \(-5 \cdot 10^8\), we need to use the pack_big functions, and the big shifts.

We receive here the difference \( a - y^2 = d = \sum_i d_i \cdot 10^{-4i} \), as \( \langle d_2 \rangle \); \( \langle d_3 \rangle \) \ldots \( \langle d_{10} \rangle \), where each block has 4 digits, except \( \langle d_2 \rangle \). This function finds the largest \( j \leq 6 \) such that \( 10^{k_j}(a - y^2) < 2 \cdot 10^8 \), then leaves an open parenthesis and the integer \( [10^{k_j}(a - y^2)] \)
in an integer expression. The closing parenthesis is provided by the caller \___fp_sqrt__auxii_o:NnnnnnnN, which completes the expression

$$10^{4j}z = \left(\left\lfloor 10^{4j}(a - y^2) \right\rfloor - 257\right) \cdot (0.5 \cdot 10^8) / \left\lfloor 10^8y + 1 \right\rfloor$$

for an estimate of $10^{4j}(\sqrt{a} - y)$. If $d_2 \geq 2$, $j = 3$ and the auxiv auxiliary receives $10^{12}z$. If $d_2 \leq 1$ but $10^{4}d_2 + 10^4d_3 \geq 2$, $j = 4$ and the auxv auxiliary is called, and receives $10^{16}z$, and so on. In all those cases, the auxviii auxiliary is set up to add $z$ to $y$, then go back to the auxii step with continuation auxiii (the function we are currently describing). The maximum value of $j$ is 6, regardless of whether $10^{12}d_2 + 10^8d_3 + 10^4d_4 + d_5 \geq 1$. In this last case, we detect when $10^{24}z < 10^7$, which essentially means $\sqrt{a} - y \lesssim 10^{-17}$: once this threshold is reached, there is enough information to find the correctly rounded $\sqrt{a}$ with only one more call to \___fp_sqrt__auxii_o:NnnnnnnN. Note that the iteration cannot be stuck before reaching $j = 6$, because for $j < 6$, one has $2 \cdot 10^8 \leq 10^{4(j+1)}(a - y^2)$, hence

$$10^{4j}z \geq \frac{(20000 - 257)(0.5 \cdot 10^8)}{\left\lfloor 10^8y + 1 \right\rfloor} \geq (20000 - 257) \cdot 0.5 > 0.$$
Simply add the two 8-digit blocks of \( z \), aligned to the last four of the five 4-digit blocks of \( y \), then call the auxii auxiliary to evaluate \( y' = (y + z)^2 \).

At this stage, \( j = 6 \) and \( 10^{24} z < 10^7 \), hence

\[
10^7 + 1/2 > 10^{24} z + 1/2 \geq (10^{24}(a - y^2) - 258) \cdot (0.5 \cdot 10^8) / (10^8 y + 1),
\]

then \( 10^{24}(a - y^2) - 258 < 2(10^7 + 1/2) (y + 10^{-8}) \), and

\[
10^{24}(a - y^2) < (10^7 + 1290.5) (y + 0.5 \cdot 10^{-8}) (2y) < (10^7 + 1290.5) (1 + 10^{-7}) (y + \sqrt{a}),
\]

which finally implies \( 0 \leq \sqrt{a} - y < 0.2 \cdot 10^{-16} \). In particular, \( y \) is an underestimate of \( \sqrt{a} \) and \( y + 0.5 \cdot 10^{-16} \) is a (strict) overestimate. There is at exactly one multiple of \( 0.5 \cdot 10^{-16} \) in the interval \( [y, y + 0.5 \cdot 10^{-16}] \), so we search for the closest multiple of \( 0.5 \cdot 10^{-16} \) in the interval \( [y, y + 0.5 \cdot 10^{-16}] \); rather, only the last 4 digits \( #8 \) of \( y \) are considered, and we do not perform any carry yet. The auxxi auxiliary sets up auxii with a continuation function auxxii instead of auxiii as before. To prevent auxii from giving a negative results \( a - m^2 \), we compute \( a + 10^{-16} - m^2 \) instead, always positive since \( m < \sqrt{a} + 0.5 \cdot 10^{-16} \) and \( a \leq 1 - 10^{-16} \).
The difference $0 \leq a + 10^{-16} - m^2 \leq 10^{-16} + (\sqrt{a} - m)(\sqrt{a} + m) \leq 2 \cdot 10^{-16}$ was just computed: its first 8 digits vanish, as do the next four, and most of the following four, $\#2$. The guess $m$ is an overestimate if $a + 10^{-16} - m^2 < 10^{-16}$, that is, $\#1\#2$ vanishes. Otherwise it is an underestimate, unless $a + 10^{-16} - m^2 = 10^{-16}$ exactly. For an underestimate, call the auxxiv function with argument 9998. For an exact result call it with 9999, and for an overestimate call it with 10000.

This receives 9998, 9999 or 10000 as $\#1$ when $m$ is an underestimate, exact, or an overestimate, respectively. Then comes $m$ as five blocks of 4 digits, but where the last block $\#6$ may be 0, 5000, or 10000. In the latter case, we need to add a carry, unless $m$ is an overestimate ($\#1$ is then 10000). Then comes $a$ as three arguments. Rounding is done by \_\_fp_round:NNN, whose first argument is the final sign (square roots are positive). We fake its second argument. It should be the last digit kept, but this is only used when ties are “rounded to even”, and only when the result is exactly half-way between two representable numbers rational square roots of numbers with 16 significant digits have: this situation never arises for the square root, as any exact square root of a 16 digit number has at most 8 significant digits. Finally, the last argument is the next digit, possibly shifted by 1 when there are further nonzero digits. This is achieved by \_\_fp_\_\_round_digit:Nw, which receives (after removal of the 10000’s digit) one of 0000, 0001, 4999, 5000, 5001, or 9999, which it converts to 0, 1, 4, 5, 6, and 9, respectively.
72.5 About the sign and exponent

\_fp_logb\_o:w
\_fp_logb\_aux\_o:w

The exponent of a normal number is its \langle exponent \rangle minus one.

\cs_new:Npn \_fp_logb\_o:w \s__fp \_fp_chk:w #1#2; @
{ \if_case:w #1 \exp_stop_f: \_fp_case_use:nw
  \{ \_fp_division_by_zero\_o:Nnw \c_minus_inf_fp \{ logb \} \}
  \or: \exp_after:wN \_fp_logb\_aux\_o:w
  \or: \exp_after:wN \_fp_case_return_o:Nw \c_inf_fp
  \else: \_fp_case_return_same\_o:w
  \fi:
  \s__fp \_fp_chk:w #1 #2; }
\cs_new:Npn \_fp_logb\_aux\_o:w \s__fp \_fp_chk:w #1 #2 #3 #4 ;
{ \exp_after:wN \_fp_parse:n \exp_after:wN \int_value:w \int_eval:w #3 - 1 \exp_after:wN \}
\exp_after:wN \_exp_after:wN \int_value:w \__fp_int_eval:w #6 + 19999 - #1 ;
\exp_after:wN \_exp_after:wN \__fp_round:NNN
\exp_after:wN 0
\exp_after:wN 0
\int_value:w \use_i:nn
\exp_after:wN \__fp_round_digit:Nw
\int_value:w \__fp_int_eval:w #6 + 19999 - #1 ;
\exp_after:wN ;\exp_after:wN} (End definition for \_fp\_sqrt\_auxxiv\_o:wnnnnnnnN.)

\_fp_sign\_o:w
\_fp_sign\_aux\_o:w

Find the sign of the floating point: nan, +0, -0, +1 or -1.

\cs_new:Npn \_fp_sign\_o:w \s__fp \_fp_chk:w #1#2; @
{ \if_case:w #1 \exp_stop_f: \_fp_case_return_same\_o:w
  \or: \exp_after:wN \_fp_case_return_same\_o:w
  \else: \_fp_case_return_same\_o:w
  \fi:
  \s__fp \_fp_chk:w #1 #2; }
\cs_new:Npn \_fp_sign\_aux\_o:w \s__fp \_fp_chk:w #1 #2 #3 #4 ;
{ \exp_after:wN \_fp_parse:n \exp_after:wN \int_value:w \int_eval:w #3 - 1 \exp_after:wN }\exp_after:wN \_exp_after:wN \int_value:w \__fp_int_eval:w #6 + 19999 - #1 ;\exp_after:wN} (End definition for \_fp\_logb\_o:w and \_fp\_logb\_aux\_o:w.)

\_fp_sign\_o:w
\_fp_sign\_aux\_o:w

(End definition for \_fp\_sign\_o:w and \_fp\_sign\_aux\_o:w.)
This function is used for the unary minus and for \texttt{abs}. It leaves the sign of \texttt{nan} invariant, turns negative numbers (sign 2) to positive numbers (sign 0) and positive numbers (sign 0) to positive or negative numbers depending on \#1. It also expands after itself in the input stream, just like \texttt{__fp_set_sign_o:w}.

\begin{verbatim}
\cs_new:Npn \__fp_set_sign_o:w #1 \s__fp \__fp_chk:w #2#3#4; @
\exp_after:wN \__fp_exp_after_o:w
\exp_after:wN \s__fp
\exp_after:wN \__fp_chk:w
\exp_after:wN #2
\int_value:w
\if_case:w #3 \exp_stop_f: #1 \or: 1 \or: 0 \fi: \exp_stop_f:
#4;
\}
\end{verbatim}

\textit{(End definition for \texttt{\_fp_set_sign_o:w}.)}

\subsection*{72.6 Operations on tuples}

Two cases: \texttt{abs((tuple))} for which \#1 is 0 (invalid for tuples) and -(tuple) for which \#1 is 2. In that case, map over all items in the tuple an auxiliary that dispatches to the type-appropriate sign-flipping function.

\begin{verbatim}
\cs_new:Npn \__fp_tuple_set_sign_o:w #1
{ \if_meaning:w 2 #1 \exp_after:wN \__fp_tuple_set_sign_aux_o:Nnw \fi:
\__fp_invalid_operation_o:nw { abs }
}
\cs_new:Npn \__fp_tuple_set_sign_aux_o:Nnw #1#2#3 @
{ \__fp_tuple_map_o:nw \__fp_tuple_set_sign_aux_o:w #3 }
\cs_new:Npn \__fp_tuple_set_sign_aux_o:w #1#2 ;
{ \__fp_change_func_type:NNN #1 \__fp_set_sign_o:w
\__fp_parse_apply_unary_error:NNw
2 #1 #2 ; @
}
\end{verbatim}

\textit{(End definition for \texttt{\_fp_tuple_set_sign_o:w}, \texttt{\_fp_tuple_set_sign_aux_o:Nnw}, and \texttt{\_fp_tuple_set_sign_aux_o:w}.)}

\subsection*{72.6.1 \texttt{\_fp*} \_fp_tuple_o:ww, \texttt{\_fp_tuple_*} \_fp_o:ww, \texttt{\_fp_tuple_/} \_fp_o:ww}

For \langle number\rangle*\langle tuple\rangle and \langle tuple\rangle*\langle number\rangle and \langle tuple\rangle/\langle number\rangle, loop through the \langle tuple\rangle some code that multiplies or divides by the appropriate \langle number\rangle. Importantly we need to dispatch according to the type, and we make sure to apply the operator in the correct order.

\begin{verbatim}
\cs_new:cpn { \_fp_*_tuple_o:ww }
{ \_fp_tuple_map_o:nw \_fp_binary_type_o:Nnw * \#1 ; }
\cs_new:cpn { \_fp_tuple_*_o:ww }
{ \_fp_tuple_map_o:nw \_fp_binary_rev_type_o:Nnw * \#2 ; }
\cs_new:cpn { \_fp_tuple_/o:ww }
{ \_fp_tuple_map_o:nw \_fp_binary_rev_type_o:Nnw / \#2 ; }
\end{verbatim}

\textit{(End definition for \texttt{\_fp_*_tuple_o:ww}, \texttt{\_fp_tuple_*_o:ww}, and \texttt{\_fp_tuple_/o:ww}.)}
Check the two tuples have the same number of items and map through these a helper that dispatches appropriately depending on the types. This means \((1,2)+((1,1),2)\) gives \((\text{nan},4)\).

\[
\text{\texttt{\textbackslash cs\_set\_protected:Npn \textbackslash \_fp\_tmp:w \#1}}
\]
\[
\text{\texttt{\{}}
\]
\[
\text{\texttt{\cs\_new:cpn { \_fp\_tuple\_#1\_tuple\_o:ww }}}
\]
\[
\text{\texttt{\s__fp\_tuple \_fp\_tuple\_chk:w \#1 \;}}
\]
\[
\text{\texttt{\s__fp\_tuple \_fp\_tuple\_chk:w \#2 \;}}
\]
\[
\text{\texttt{\{}\texttt{}}
\]
\[
\text{\texttt{\int\_compare:nNnTF}}
\]
\[
\text{\texttt{\{ \_fp\_array\_count:n \{\#1\} \} = \{ \_fp\_array\_count:n \{\#2\} \}}
\]
\[
\text{\texttt{\{ \_fp\_tuple\_map\_thread\_o:nw \{ \_fp\_binary\_type\_o:Nw \#1 \}}
\]
\[
\text{\texttt{\{ \_fp\_invalid\_operation\_o:nw \#1 \}}
\]
\[
\text{\texttt{\s__fp\_tuple \_fp\_tuple\_chk:w \{\#1\} \;}}
\]
\[
\text{\texttt{\s__fp\_tuple \_fp\_tuple\_chk:w \{\#2\} \;}}
\]
\[
\text{\texttt{\}}\texttt{\}}
\]
\[
\text{\texttt{\texttt{\textbackslash \_fp\_tmp:w +}}}
\]
\[
\text{\texttt{\textbackslash \_fp\_tmp:w -}}
\]

\(\text{(End definition for } \_\_fp\_tuple\_+\_tuple\_o:ww \text{ and } \_\_fp\_tuple\_\_\_tuple\_o:ww)\)

\(\langle /\text{package} \rangle\)
Chapter 73

l3fp-extended implementation

73.1 Description of fixed point numbers

This module provides a few functions to manipulate positive floating point numbers with extended precision (24 digits), but mostly provides functions for fixed-point numbers with this precision (24 digits). Those are used in the computation of Taylor series for the logarithm, exponential, and trigonometric functions. Since we eventually only care about the 16 first digits of the final result, some of the calculations are not performed with the full 24-digit precision. In other words, the last two blocks of each fixed point number may be wrong as long as the error is small enough to be rounded away when converting back to a floating point number. The fixed point numbers are expressed as

\[ \{a_1\} \{a_2\} \{a_3\} \{a_4\} \{a_5\} \{a_6\} ; \]

where each \(a_i\) is exactly 4 digits (ranging from 0000 to 9999), except \(a_1\), which may be any “not-too-large” non-negative integer, with or without leading zeros. Here, “not-too-large” depends on the specific function (see the corresponding comments for details).

Checking for overflow is the responsibility of the code calling those functions. The fixed point number \(a\) corresponding to the representation above is \(a = \sum_{i=1}^{6} a_i \cdot 10^{-4i}\).

Most functions we define here have the form

\[ \_fp_fixed\_\langle\text{calculation}\rangle::\langle\text{operand}\_1\rangle;\langle\text{operand}\_2\rangle;\langle\text{continuation}\rangle \]

They perform the \(\langle\text{calculation}\rangle\) on the two \(\langle\text{operands}\rangle\), then feed the result (6 brace groups followed by a semicolon) to the \(\langle\text{continuation}\rangle\), responsible for the next step of the calculation. Some functions only accept an N-type \(\langle\text{continuation}\rangle\). This allows constructions such as

\[ \_fp_fixed_add::\langle\text{operand}\_1\rangle;\langle\text{operand}\_2\rangle;\langle\text{continuation}\rangle \]
\[ \_fp_fixed_mul::\langle\text{operand}\_1\rangle;\langle\text{operand}\_2\rangle;\langle\text{continuation}\rangle \]
\[ \_fp_fixed_add::\langle\text{operand}\_1\rangle;\langle\text{operand}\_2\rangle;\langle\text{continuation}\rangle \]
to compute \((X_1 + X_2) \cdot X_3 + X_4\). This turns out to be very appropriate for computing continued fractions and Taylor series.

At the end of the calculation, the result is turned back to a floating point number using \_fp-fixed-to-floating\_o:wN. This function has to change the exponent of the floating point number: it must be used after starting an integer expression for the overall exponent of the result.

### 73.2 Helpers for numbers with extended precision

#### \_c\_fp\_one\_fixed\_tl

The fixed-point number 1, used in \_fp-expo.

```latex
\texttt{tl_const:Nn \_c\_fp\_one\_fixed\_tl}
\texttt{\{0000\} \{0000\} \{0000\} \{0000\} ;}
```

(End definition for \_c\_fp\_one\_fixed\_tl.)

#### \_fp\_fixed\_continue:wn

This function simply calls the next function.

```latex
\texttt{cs_new:Npn \_fp\_fixed\_continue:wn \#1; \#2 \{ \#2 \#1;}}
```

(End definition for \_fp\_fixed\_continue:wn.)

#### \_fp\_fixed\_add\_one:wn

\_fp\_fixed\_add\_one:wn \(a\); continuation

This function adds 1 to the fixed point \(a\), by changing \(a_1\) to 10000 + \(a_1\), then calls the continuation. This requires \(a_1 + 10000 < 2^{31}\).

```latex
\texttt{cs_new:Npn \_fp\_fixed\_add\_one:wn \#1\#2\#3}{\#2\#3}
\texttt{\#1;}
```

(End definition for \_fp\_fixed\_add\_one:wn.)

#### \_fp\_fixed\_div\_myriad:wn

Divide a fixed point number by 10000. This is a little bit more subtle than just removing the last group and adding a leading group of zeros: the first group \#1 may have any number of digits, and we must split \#1 into the new first group and a second group of exactly 4 digits. The choice of shifts allows \#1 to be in the range \(0, 5 \cdot 10^8 - 1\).

```latex
\texttt{cs_new:Npn \_fp\_fixed\_div\_myriad:wn \#1\#2\#3\#4\#5\#6}{\#1\#2\#3\#4\#5\#6}
```

(End definition for \_fp\_fixed\_div\_myriad:wn.)

#### \_fp\_fixed\_mul\_after:wn

The fixed point operations which involve multiplication end by calling this auxiliary. It braces the last block of digits, and places the continuation \#3 in front.

```latex
\texttt{cs_new:Npn \_fp\_fixed\_mul\_after:wn \#1\#2\#3}{\#1\#2\#3\#1\#2;}
```

(End definition for \_fp\_fixed\_mul\_after:wn.)
73.3 Multiplying a fixed point number by a short one

\__fp_fixed_mul_short:wwn
\{\langle a_1 \rangle \} \{\langle a_2 \rangle \} \{\langle a_3 \rangle \} \{\langle a_4 \rangle \}\}
\{\langle b_0 \rangle \} \{\langle b_1 \rangle \} \{\langle b_2 \rangle \} \{\langle \text{continuation} \rangle \}

Computes the product \( c = ab \) of \( a = \sum_i \langle a_i \rangle 10^{-4i} \) and \( b = \sum_i \langle b_i \rangle 10^{-4i} \), rounds it to the closest multiple of \( 10^{-24} \), and leaves \( \{\langle \text{continuation} \rangle \} ... \{\langle c_6 \rangle \} \) in the input stream, where each of the \( \langle c_i \rangle \) are blocks of 4 digits, except \( \langle c_1 \rangle \), which is any \TeX integer. Note that indices for \( \langle b \rangle \) start at 0: for instance a second operand of \{0001\}{0000}{0000} leaves the first operand unchanged (rather than dividing it by \( 10^4 \), as \_fp_fixed_mul:wwn would).

\begin{verbatim}
\cs_new:Npn \__fp_fixed_mul_short:wwn #1#2#3#4#5#6; #7#8#9;
  \exp_after:wN \__fp_fixed_mul_after:wwn
  \int_value:w \__fp_int_eval:w \c__fp_leading_shift_int + #1*#7
  \exp_after:wN \__fp_pack:NNNNNw
  \int_value:w \__fp_int_eval:w \c__fp_middle_shift_int + #1*#8 + #2*#7
  \exp_after:wN \__fp_pack:NNNNNw
  \int_value:w \__fp_int_eval:w \c__fp_middle_shift_int + #1*#9 + #2*#8 + #3*#7
  \exp_after:wN \__fp_pack:NNNNNw
  \int_value:w \__fp_int_eval:w \c__fp_middle_shift_int + #1*#9 + #2*#8 + #3*#7
  \exp_after:wN \__fp_pack:NNNNNw
  \int_value:w \__fp_int_eval:w \c__fp_middle_shift_int + #2*#9 + #3*#8 + #4*#7
  \exp_after:wN \__fp_pack:NNNNNw
  \int_value:w \__fp_int_eval:w \c__fp_middle_shift_int + #3*#9 + #4*#8 + #5*#7
  \exp_after:wN \__fp_pack:NNNNNw
  \int_value:w \__fp_int_eval:w \c__fp_trailing_shift_int + #4*#9 + #5*#8 + #6*#7
  + ( #5*#9 + #6*#8 + #6*#9 / \c__fp_myriad_int )
  / \c__fp_myriad_int ; ;
\end{verbatim}

(End definition for \__fp_fixed_mul_short:wwn.)

73.4 Dividing a fixed point number by a small integer

\__fp_fixed_div_int:wwN \__fp_fixed_div_int:wnN \__fp_fixed_div_intAux:wnn \__fp_fixed_div_int Auxii:wnn \__fp_fixed_div_int_pack:Nw \__fp_fixed_div_int_after:Nw

\_fp_fixed_div_int:wwN (a) ; (n) ; (continuation)

Divides the fixed point number \( a \) by the (small) integer \( 0 < \langle n \rangle < 10^4 \) and feeds the result to the \( \{\langle \text{continuation} \rangle \} \). There is no bound on \( a_1 \).

The arguments of the i auxiliary are 1: one of the \( a_i \), 2: \( n \), 3: the ii or the iii auxiliary. It computes a (somewhat tight) lower bound \( Q_i \) for the ratio \( a_i/n \).

The ii auxiliary receives \( Q_i \), \( n \), and \( a_i \) as arguments. It adds \( Q_i \) to a surrounding integer expression, and starts a new one with the initial value 9999, which ensures that the result of this expression has 5 digits. The auxiliary also computes \( a_i - n \cdot Q_i \), placing the result in front of the 4 digits of \( a_{i+1} \). The resulting \( a_{i+1} = 10^4(a_i - n \cdot Q_i) + a_{i+1} \) serves as the first argument for a new call to the i auxiliary.

When the iii auxiliary is called, the situation looks like this:
\__fp_fixed_div_int_after:Nw (continuation)
\__fp_fixed_div_int_pack:Nw 9999 + Q_1
\__fp_fixed_div_int_pack:Nw 9999 + Q_2
\__fp_fixed_div_int_pack:Nw 9999 + Q_3
\__fp_fixed_div_int_pack:Nw 9999 + Q_4
\__fp_fixed_div_int_pack:Nw 9999 + Q_5
\__fp_fixed_div_int_pack:Nw 9999
\__fp_fixed_div_int_auxii:wnn \{ (n) \} \{ (a_6) \}

where expansion is happening from the last line up. The \textit{iii} auxiliary adds $Q_6 + 2 \simeq a_6/n + 1$ to the last 9999, giving the integer closest to 10000 + $a_6/n$.

Each \textit{pack} auxiliary receives 5 digits followed by a semicolon. The first digit is added as a carry to the integer expression above, and the 4 other digits are braced. Each call to the \textit{pack} auxiliary thus produces one brace group. The last brace group is produced by the \textit{after} auxiliary, which places the (continuation) as appropriate.

2567 \cs_new:Npn \__fp_fixed_div_int:wwN #1#2#3#4#5#6 ; #7 ; #8
2568 { \exp_after:wN \__fp_fixed_div_int_after:Nw \exp_after:wN #8
2569 \int_value:w \__fp_int_eval:w - 1 \__fp_fixed_div_int:wnN #1; {#7} \__fp_fixed_div_int_auxi:wnn
2570 #2; {#7} \__fp_fixed_div_int_auxi:wnn
2571 #3; {#7} \__fp_fixed_div_int_auxi:wnn
2572 #4; {#7} \__fp_fixed_div_int_auxi:wnn
2573 #5; {#7} \__fp_fixed_div_int_auxi:wnn
2574 #6; {#7} \__fp_fixed_div_int_auxii:wnn ;
2575 }
2576 \cs_new:Npn \__fp_fixed_div_int:wnN #1; #2 #3
2577 { \exp_after:wN #3 \int_value:w \__fp_int_eval:w #1 / #2 - 1 ; {#2}
2578 \{#1\}
2579 }
2580 \cs_new:Npn \__fp_fixed_div_int_auxi:wnn #1; #2 #3
2581 { + #1 \exp_after:wN \__fp_fixed_div_int_pack:Nw \int_value:w \__fp_int_eval:w 9999
2582 \exp_after:wN \__fp_fixed_div_int:wnN \int_value:w \__fp_int_eval:w #3 - #1*#2 \__fp_int_eval_end:
2583 }
2584 \cs_new:Npn \__fp_fixed_div_int_auxii:wnn \{ (n) \} \{ (a_6) \}
2585 \cs_new:Npn \__fp_fixed_div_int_after:Nw #1 #2; {#1} \{ #2 \}

(End definition for \__fp_fixed_div_int:wwN and others.)

73.5 Adding and subtracting fixed points
\__fp_fixed_add:wnn ⟨a⟩; ⟨b⟩; { (continuation) }
Computes \(a + b\) (resp. \(a - b\)) and feeds the result to the \((\text{continuation})\). This function requires \(0 \leq a_1, b_1 \leq 114748\), its result must be positive (this happens automatically for addition) and its first group must have at most 5 digits: \((a \pm b)_1 < 100000\). The two functions only differ by a sign, hence use a common auxiliary. It would be nice to grab the 12 brace groups in one go; only 9 parameters are allowed. Start by grabbing the sign, \(a_1, \ldots, a_4\), the rest of \(a\), and \(b_1\) and \(b_2\). The second auxiliary receives the rest of \(a\), the sign multiplying \(b\), the rest of \(b\), and the \((\text{continuation})\) as arguments. After going down through the various level, we go back up, packing digits and bringing the \((\text{continuation})\) \((\#8, \text{then } \#7)\) from the end of the argument list to its start.

\[
\begin{align*}
\text{\texttt{\_fp\_fixed\_add:wwn \{ \_fp\_fixed\_add:Nnnnnnnn \ + \}}}
\text{\texttt{\_fp\_fixed\_sub:wwn \{ \_fp\_fixed\_add:Nnnnnnnn \ - \}}}
\text{\texttt{\_fp\_fixed\_add:Nnnnnnnn \#1 \#2\#3\#4\#5 \#6; \#7\#8}}
\end{align*}
\]

\[
\text{\texttt{\exp\_after:wN \_fp\_fixed\_add\_after:NNNNNNw}}
\text{\texttt{\_fp\_int\_eval:w \_fp\_fixed\_add\_pack:NNNNNNw}}
\text{\texttt{\_fp\_int\_eval:w \_fp\_fixed\_add:nnNNNNNw \#1 \#2 \#3 \#4\#5 \#6 \#7 \#8}}
\]

\[
\text{\texttt{\_fp\_fixed\_add:nnNNNNNw \#1 \#2 \#3 \#4\#5 \#6 \#7 \#8}}
\]

\[
\text{\texttt{\_fp\_fixed\_add\_pack:NNNNNNw \#1 \#2\#3\#4\#5 \#6}}
\]

\[
\text{\texttt{\_fp\_fixed\_add\_after:NNNNNNw}}
\]

\[
\text{\texttt{\_fp\_fixed\_mul:wwn \_fp\_fixed\_mul:nnNNNNNw}}
\]

\[
\text{\texttt{\_fp\_fixed\_mul:nnNNNNNw \{ \_fp\_fixed\_add:wwn \ ; \{ (\text{continuation}) \}}} \]

73.6 Multiplying fixed points

\[
\text{\texttt{\_fp\_fixed\_mul:wwn \(a\); \(b\); \{ (\text{continuation}) \}}} \]

Computes \(a \times b\) and feeds the result to \((\text{continuation})\). This function requires \(0 \leq a_1, b_1 < 10000\). Once more, we need to play around the limit of 9 arguments for \TeX\ macros. Note that we don’t need to obtain an exact rounding, contrarily to the \(*\) operator, so things could be harder. We wish to perform carries in

\[
\begin{align*}
a \times b &= a_1 \cdot b_1 \cdot 10^{-8}
+ \left( a_1 \cdot b_2 + a_2 \cdot b_1 \right) \cdot 10^{-12}
+ \left( a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1 \right) \cdot 10^{-16}
+ \left( a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1 \right) \cdot 10^{-20}
+ \left( a_2 \cdot b_4 + a_3 \cdot b_3 + a_4 \cdot b_2
+ a_3 \cdot b_4 + a_4 \cdot b_3 + a_1 \cdot b_6 + a_2 \cdot b_5 + a_5 \cdot b_2 + a_6 \cdot b_1 \right)
\frac{10^4}{10^{24}} + O(10^{-24}),
\end{align*}
\]

1088
where the $O(10^{-24})$ stands for terms which are at most $5 \cdot 10^{-24}$, ignoring those leads to an error of at most 5 ulp. Note how the first 15 terms only depend on $a_1, \ldots, a_4$ and $b_1, \ldots, b_4$, while the last 6 terms only depend on $a_1, a_2, a_5, a_6$ and the corresponding parts of $b$. Hence, the first function grabs $a_1, \ldots, a_4$, the rest of $a$, and $b_1, \ldots, b_4$, and writes the 15 first terms of the expression, including a left parenthesis for the fraction. The $i$ auxiliary receives $a_5, a_6, b_1, b_2, a_1, a_2, b_5, b_6$ and finally the (continuation) as arguments. It writes the end of the expression, including the right parenthesis and the denominator of the fraction. The (continuation) is finally placed in front of the 6 brace groups by \_\_fp_fixed_mul_after:wwn.

\begin{verbatim}
\cs_new:Npn \_\_fp_fixed_mul_add:wwn #1#2#3#4 #5; #6#7#8#9
\exp_after:wN \_\_fp_fixed_mul_after:wwn
\int_value:w \_\_fp_int_eval:w \c__fp_leading_shift_int + #1*#6
\exp_after:wN \_\_fp_pack:NNNNw
\int_value:w \_\_fp_int_eval:w \c__fp_middle_shift_int + #1*#7 + #2*#6
\exp_after:wN \_\_fp_pack:NNNNw
\int_value:w \_\_fp_int_eval:w \c__fp_middle_shift_int + #1*#8 + #2*#7 + #3*#6
\exp_after:wN \_\_fp_pack:NNNNw
\int_value:w \_\_fp_int_eval:w \c__fp_middle_shift_int + #1*#9 + #2*#8 + #3*#7
\int_value:w \_\_fp_int_eval:w \c__fp_middle_shift_int + #3*#9 + #4*#8 + ( #3*#9 + #4*#8
+ \_\_fp_fixed_mul:nnnnnnnw #5 {#6} {#7} {#1}{#2} ) / \_\_fp_fixed_mul:nnnnnnw #5 {#6} {#7} #1#2 #3#4 #5#6 #7#8 ;
\end{verbatim}

(End definition for \_\_fp_fixed_mul:wwn and \_\_fp_fixed_mul:nnnnnnw.)

### 73.7 Combining product and sum of fixed points

\_\_fp_fixed_mul_add:wwn (a); (b); (c); {\{continuation\}}
\_\_fp_fixed_mul_sub_back:wwn (a); (b); (c); {\{continuation\}}
\_\_fp_fixed_one_minus_mul:wwn (a); (b); {\{continuation\}}

Sometimes called FMA (fused multiply-add), these functions compute $a \times b + c$, $c - a \times b$, and $1 - a \times b$ and feed the result to the (continuation). Those functions require $0 \leq a_1, b_1, c_1 \leq 10000$. Since those functions are at the heart of the computation of Taylor expansions, we over-optimize them a bit, and in particular we do not factor out the common parts of the three functions.
For definiteness, consider the task of computing \( a \times b + c \). We perform carries in
\[
\begin{align*}
\text{\( a \times b + c = (a_1 \cdot b_1 + c_1c_2) \cdot 10^{-8} \)}
+ (a_1 \cdot b_2 + a_2 \cdot b_1) \cdot 10^{-12} \\
+ (a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1 + c_3c_4) \cdot 10^{-16} \\
+ (a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1) \cdot 10^{-20} \\
+ \left( a_2 \cdot b_4 + a_3 \cdot b_3 + a_4 \cdot b_2 \\
+ a_1 \cdot b_4 + a_4 \cdot b_3 + a_1 \cdot b_6 + a_2 \cdot b_5 + a_5 \cdot b_2 + a_6 \cdot b_1 \right) \cdot 10^{4} \\
+ a_1 \cdot b_5 + a_5 \cdot b_1 + c_5c_6 \right) \cdot 10^{-24} + O(10^{-24}),
\end{align*}
\]
where \( c_1c_2, c_3c_4, c_5c_6 \) denote the 8-digit number obtained by juxtaposing the two blocks of digits of \( c \), and \( \cdot \) denotes multiplication. The task is obviously tough because we have 18 brace groups in front of us.

Each of the three function starts the first two levels (the first, corresponding to \( 10^{-4} \), is empty), with \( c_1c_2 \) in the first level, calls the \( i \) auxiliary with arguments described later, and adds a trailing + \( c_5c_6 \); \{(continuation)\} ;. The + \( c_5c_6 \) piece, which is omitted for \(--\_fp_fixed_one_minus_mul:wwn\) is taken in the integer expression for the \( 10^{-24} \) level.
\__fp_fixed_mul_add:Nwnnnwnnn
\__fp_fixed_mul_add:nnnnwnnnn (op) + \langle c_3 \rangle \langle c_4 \rangle ;
\langle b \rangle ; \langle a \rangle ; \langle b \rangle ; (op)
+ \langle c_5 \rangle \langle c_6 \rangle ;

Here, \langle op \rangle is either + or -. Arguments \#3, \#4, \#5 are \langle b_1 \rangle, \langle b_2 \rangle, \langle b_3 \rangle; arguments \#7, \#8, \#9 are \langle a_1 \rangle, \langle a_2 \rangle, \langle a_3 \rangle. We can build three levels: \( a_1 \cdot b_1 \) for \( 10^{-8} \), \( a_1 \cdot b_2 + a_2 \cdot b_1 \) for \( 10^{-12} \), and \( a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1 + c_3 c_4 \) for \( 10^{-16} \). The \( a-b \) products use the sign \#1. Note that \#2 is empty for \__fp_fixed_one_minus_mul:w. We call the \#1 auxiliary for levels \( 10^{-20} \) and \( 10^{-24} \), keeping the pieces of \langle a \rangle we’ve read, but not \langle b \rangle, since there is another copy later in the input stream.

\cs_new:Npn \__fp_fixed_mul_add:nnnnwnnnn \#1 \#2; \#3\#4\#5; \#7\#8\#9
{\#1 \#7*\#3
\exp_after:wN \__fp_pack_big:NNNNNw}
\int_value:w \__fp_int_eval:w \c__fp_big_middle_shift_int
\#1 \#7*\#4 \#1 \#8*\#3
\exp_after:wN \__fp_pack_big:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_big_middle_shift_int
\#1 \#7*\#5 \#1 \#8*\#4 \#1 \#9*\#3 \#2
\exp_after:wN \__fp_pack_big:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_big_middle_shift_int
\#1 \__fp_fixed_mul_add:nnnnwnnnn \#7\#8\#9
}

\__fp_fixed_mul_add:nnnnwnnnn \langle a \rangle ; \langle b \rangle ; (op)
+ \langle c_5 \rangle \langle c_6 \rangle ;

Level \( 10^{-20} \) is \( (a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1) \), multiplied by the sign, which was inserted by the \#1 auxiliary. Then we prepare level \( 10^{-24} \). We don’t have access to all parts of \langle a \rangle and \langle b \rangle needed to make all products. Instead, we prepare the partial expressions

\[
\begin{align*}
&b_1 + a_4 \cdot b_2 + a_3 \cdot b_3 + a_2 \cdot b_4 + a_1 \\
&b_2 + a_4 \cdot b_3 + a_3 \cdot b_4 + a_2.
\end{align*}
\]

Obviously, those expressions make no mathematical sense: we complete them with \( a_5 \cdot b_5 \), and \( a_6 \cdot b_1 + a_5 \cdot b_5 \) and \( a_6 \cdot b_1 + a_5 \cdot b_6 \), and of course with the trailing \( +c_5 c_6 \).

To do all this, we keep \( a_1, a_5, a_6 \), and the corresponding pieces of \langle b \rangle.

\cs_new:Npn \__fp_fixed_mul_add:nnnnwnnnn \#1\#2\#3\#4\#5; \#6\#7\#8\#9
{\#1*\#9 + \#2*\#8 + \#3*\#7 + \#4*\#6
\exp_after:wN \__fp_pack_big:NNNNNw}
\int_value:w \__fp_int_eval:w \c__fp_big_trailing_shift_int
\__fp_fixed_mul_add:nnnnwnnnN
{\#6 + \#4*\#7 + \#3*\#8 + \#2*\#9 + \#1 }
{\#7 + \#4*\#8 + \#3*\#9 + \#2 }
{\#1} \#5;
{\#6}
The second one is divided by 10000; this is the carry from level $10^{-28}$. The trailing
+ $c_5c_6$ is taken into the expression for level $10^{-24}$. Note that the total of level $10^{-24}$ is
in the interval $[-5 \cdot 10^8, 6 \cdot 10^8]$ (give or take a couple of 10000), hence adding it to the
shift gives a 10-digit number, as expected by the packing auxiliaries. See \l3fp-aux for the
definition of the shifts and packing auxiliaries.

In other words, an extended-precision number is an exponent ending in a comma, then a
fixed point number. The corresponding value is $0.\langle \text{digits} \rangle \cdot 10^{\langle \text{exponent} \rangle}$. This convention
differs from floating points.

\begin{verbatim}
\cs_new:Npn \__fp_ep_to_fixed:wwn #1,#2
{\exp_after:wN \__fp_ep_to_fixed_auxi:www
\int_value:w \__fp_int_eval:w 1 0000 0000 + #2 \exp_after:wN ;}
\cs_new:Npn \__fp_ep_to_fixed_auxi:www 1#1; #2; #3#4#5#6#7;
{\__fp_pack_eight:wNNNNNNNN \__fp_pack_twice_four:wNNNNNNNN \__fp_pack_twice_four:wNNNNNNNN \__fp_pack_twice_four:wNNNNNNNN
\__fp_ep_to_fixed_auxii:nnnnnnnwn ; #2 #1#3#4#5#6#7 0000 !}
\cs_new:Npn \__fp_ep_to_fixed_auxii:nnnnnnnwn #1#2#3#4#5#6#7; #8! #9
{#9 {#1#{#2}{#3}{#4}{#5}{#6}{#7};}
(End definition for \__fp_fixed_mul_add:nnnnwnnn.)
\end{verbatim}

\section{Extended-precision floating point numbers}

In this section we manipulate floating point numbers with roughly 24 significant figures
("extended-precision" numbers, in short, “ep”), which take the form of an integer expo-
ponent, followed by a comma, then six groups of digits, ending with a semicolon. The first
group of digit may be any non-negative integer, while other groups of digits have 4 digits.
In other words, an extended-precision number is an exponent ending in a comma, then a
fixed point number. The corresponding value is $0.\langle \text{digits} \rangle \cdot 10^{\langle \text{exponent} \rangle}$. This convention
diffs from floating points.

Converts an extended-precision number with an exponent at most 4 and a first block less
than $10^8$ to a fixed point number whose first block has 12 digits, hopefully starting with
many zeros.

\begin{verbatim}
\cs_new:Npn \__fp_ep_to_fixed:wwn \__fp_ep_to_fixed_auxi:www \__fp_ep_to_fixed_auxii:nnnnnnnwn
{\exp_after:wN \__fp_ep_to_fixed_auxi:www \int_value:w \__fp_int_eval:w 1 0000 0000 + #2 \exp_after:wN ;}
\cs_new:Npn \__fp_ep_to_fixed_auxi:www 1#1; #2; #3#4#5#6#7;
{\__fp_pack_eight:wNNNNNNNN \__fp_pack_twice_four:wNNNNNNNN \__fp_pack_twice_four:wNNNNNNNN \__fp_pack_twice_four:wNNNNNNNN
\__fp_ep_to_fixed_auxii:nnnnnnnwn ; #2 #1#3#4#5#6#7 0000 !}
\cs_new:Npn \__fp_ep_to_fixed_auxii:nnnnnnnwn #1#2#3#4#5#6#7; #8! #9
{#9 {#1#{#2}{#3}{#4}{#5}{#6}{#7};}
(End definition for \__fp_ep_to_fixed:wwn, \__fp_ep_to_fixed_auxi:www, and \__fp_ep_to_fixed_auxii:nnnnnnnwn.)
\end{verbatim}
Normalize an extended-precision number. More precisely, leading zeros are removed from the mantissa of the argument, decreasing its exponent as appropriate. Then the digits are packed into 6 groups of 4 (discarding any remaining digit, not rounding). Finally, the continuation \#8 is placed before the resulting exponent–mantissa pair. The input exponent may in fact be given as an integer expression. The loop auxiliary grabs a digit: if it is 0, decrement the exponent and continue looping, and otherwise call the end auxiliary, which places all digits in the right order (the digit that was not 0, and any remaining digits), followed by some 0, then packs them up neatly in $3 \times 2 = 6$ blocks of four. At the end of the day, remove with \_fp_use_i:ww any digit that did not make it in the final mantissa (typically only zeros, unless the original first block has more than 4 digits).

```
\cs_new:Npn \_fp_ep_to_ep:wwN \#1,#2#3#4#5#6#7; \#8
\exp_after:wN \#8
\int_value:w \__fp_int_eval:w \#1 + 4
\exp_after:wN \use_i:nn
\exp_after:wN \__fp_ep_to_ep_loop:N
\int_value:w \__fp_int_eval:w \#3\#4\#5\#6\#7 ; ; !
\}
```

```
\cs_new:Npn \__fp_ep_to_ep_loop:N \#1
\if_meaning:w 0 \#1
- 1
\else:
\_fp_ep_to_ep_end:www \#1
\fi:
\_fp_ep_to_ep_loop:N

\}
```

```
\cs_new:Npn \__fp_ep_to_ep_end:www \#1 \fi: \_fp_ep_to_ep_loop:N \#2; \#3!
\{
\fi:
\if_meaning:w \#1
- 2 \_fp_max_exponent_int
\_fp_ep_to_ep_zero:ww
\fi:
\_fp_pack_twice_four:wNNNNNNNN
\_fp_pack_twice_four:wNNNNNNNN
\_fp_pack_twice_four:wNNNNNNNN
\_fp_use_i:ww , ;
\#1 \#2 0000 0000 0000 0000 0000 0000 ;
\}
```

```
\cs_new:Npn \_fp_ep_to_ep_zero:ww \fi: \_fp_ep_to_ep_loop:N \#1; \#2; \#3;
\{
\fi: , \{1000\}{0000}\{0000\}{0000}\{0000\}; \}
```

(End definition for \_fp_ep_to_ep:wwN and others.)

In l3fp-trig we need to compare two extended-precision numbers. This is based on the same function for positive floating point numbers, with an extra test if comparing only 16 decimals is not enough to distinguish the numbers. Note that this function only works if the numbers are normalized so that their first block is in $[1000, 9999]$.

```
\cs_new:Npn \_fp_ep_compare:wwww \#1,#2#3#4#5#6#7;
```

1093
Multiply two extended-precision numbers: first normalize them to avoid losing too much precision, then multiply the mantissas #2 and #4 as fixed point numbers, and sum the exponents #1 and #3. The result’s first block is in $[100, 9999]$. }

```
\cs_new:Npn \__fp_ep_mul:wwwwn #1,#2; #3,#4; #5
{ \__fp_ep_to_ep:wwN #3,#4; \__fp_fixed_continue:wn
  \__fp_ep_to_ep:wwN #1,#2; \__fp_ep_mul_raw:wwwwN
} \__fp_fixed_continue:wn

\cs_new:Npn \__fp_ep_mul_raw:wwwwN #1,#2; #3,#4; #5
{ \__fp_fixed_mul:wwn #2; #4; \exp_after:wN \__fp_int_eval:w \__fp_int_eval:w #1 + #3 , } \__fp_int_eval:w #5
```

(End definition for \__fp_ep_mul:wwwwn and \__fp_ep_mul_raw:wwwwN.)

73.9 Dividing extended-precision numbers

Divisions of extended-precision numbers are difficult to perform with exact rounding: the technique used in l3fp-basics for 16-digit floating point numbers does not generalize easily to 24-digit numbers. Thankfully, there is no need for exact rounding.

Let us call $\langle n \rangle$ the numerator and $\langle d \rangle$ the denominator. After a simple normalization step, we can assume that $\langle n \rangle \in [0.1, 1)$ and $\langle d \rangle \in [0.1, 1)$, and compute $\langle n \rangle/(10\langle d \rangle) \in (0.01, 1)$. In terms of the 6 blocks of digits $\langle n_1 \rangle\cdots\langle n_6 \rangle$ and the 6 blocks $\langle d_1 \rangle\cdots\langle d_6 \rangle$, the condition translates to $\langle n_1 \rangle, \langle d_1 \rangle \in [1000, 9999]$.  

1094
We first find an integer estimate \( a \approx 10^8/\langle d \rangle \) by computing
\[
\alpha = \left[ \frac{10^9}{\langle d \rangle + 1} \right],
\beta = \left[ \frac{10^9}{\langle d \rangle} \right],
\]
\[
a = 10^3 \alpha + (\beta - \alpha) \cdot \left( 10^3 - \left[ \frac{\langle d \rangle}{10} \right] \right) - 1250,
\]
where \( [\cdot] \) denotes \( \varepsilon \)-TEX’s rounding division, which rounds ties away from zero. The idea is to interpolate between \( 10^3 \alpha \) and \( 10^3 \beta \) with a parameter \( \langle d \rangle \)/10\(^4\), so that when \( \langle d \rangle = 0 \) one gets \( a = 10^3 \beta - 1250 \approx 10^{12}/\langle d \rangle \) \( \approx 10^8/\langle d \rangle \), while when \( \langle d \rangle = 9999 \) one gets \( a = 10^3 \alpha - 1250 \approx 10^{12}/(\langle d \rangle + 1) \approx 10^8/\langle d \rangle \). The shift by 1250 helps to ensure that \( a \) is an underestimate of the correct value. We shall prove that
\[
1 - 1.755 \cdot 10^{-5} < \frac{\langle d \rangle a}{10^8} < 1.
\]
We can then compute the inverse of \( \langle d \rangle a/10^8 = 1 - \epsilon \) using the relation \( 1/(1 - \epsilon) \approx (1 + \epsilon)(1 + \epsilon^2) + \epsilon^4 \), which is correct up to a relative error of \( \epsilon^5 < 1.6 \cdot 10^{-24} \). This allows us to find the desired ratio as
\[
\left( \frac{n}{d} \right) - \frac{\langle a \rangle a}{10^8} (1 + \epsilon)(1 + \epsilon^2) + \epsilon^4.
\]

Let us prove the upper bound first (multiplied by 10\(^{15}\)). Note that 10\(^7\langle d \rangle < 10^3 \langle d \rangle + 10^{-1}(\langle d \rangle + 1) \), and that \( \varepsilon \)-TEX’s division \( \left[ \frac{\langle d \rangle}{10} \right] \) underestimates \( 10^{-1}(\langle d \rangle + 1) \) by 0.5 at most, as can be checked for each possible last digit of \( \langle d \rangle \). Then,
\[
10^7 \langle d \rangle a < \left( 10^3 \langle d \rangle + \left[ \frac{\langle d \rangle}{10} \right] + \frac{1}{2} \right) \left( 10^3 - \left[ \frac{\langle d \rangle}{10} \right] \right) \beta + \left[ \frac{\langle d \rangle}{10} \right] \alpha - 1250 \tag{1}
\]
\[
< 10^3 \langle d \rangle + \left[ \frac{\langle d \rangle}{10} \right] + \frac{1}{2} \left( 10^3 - \left[ \frac{\langle d \rangle}{10} \right] \right) \left( \left[ \frac{\langle d \rangle}{10} \right] + \frac{1}{2} \right) + \left[ \frac{\langle d \rangle}{10} \right] \left( \frac{10^9}{\langle d \rangle + 1} + \frac{1}{2} \right) - 1250 \tag{2}
\]
\[
< \left( 10^3 \langle d \rangle + \left[ \frac{\langle d \rangle}{10} \right] + \frac{1}{2} \right) \left( \frac{10^{12}}{\langle d \rangle} - \left[ \frac{\langle d \rangle}{10} \right] \right) - 1250 \tag{3}
\]
\[
< \left( 10^3 \langle d \rangle + \left[ \frac{\langle d \rangle}{10} \right] + \frac{1}{2} \right) \left( \frac{10^{12}}{\langle d \rangle} - \left[ \frac{\langle d \rangle}{10} \right] \right) - 1250 \tag{4}
\]
We recognize a quadratic polynomial in \( \left[ \langle d \rangle/10 \right] \) with a negative leading coefficient: this polynomial is bounded above, according to \( (\langle d \rangle/10 + a)(b - c(\langle d \rangle/10)) \leq (b + ca)^2/(4c) \). Hence,
\[
10^7 \langle d \rangle a < \frac{10^{15}}{\langle d \rangle(\langle d \rangle + 1)} \left( \langle d \rangle + \frac{1}{2} + \frac{1}{4} \cdot 10^{-3} - \frac{3}{8} \cdot 10^{-9} \langle d \rangle(\langle d \rangle + 1) \right)^2
\]
Since \( \langle d \rangle \) takes integer values within \( [1000, 9999] \), it is a simple programming exercise to check that the squared expression is always less than \( \langle d \rangle(\langle d \rangle + 1) \), hence \( 10^7 \langle d \rangle a < 10^{15} \). The upper bound is proven. We also find that \( \frac{8}{3} \) can be replaced by slightly smaller numbers, but nothing less than 0.374563..., and going back through the derivation of
the upper bound, we find that 1250 is as small a shift as we can obtain without breaking the bound.

Now, the lower bound. The same computation as for the upper bound implies

$$10^7 \langle d \rangle a > \left( \frac{10^3 \langle d_1 \rangle}{10} + \left\lceil \frac{10^9}{\langle d_1 \rangle} \right\rceil - \frac{1}{2} \right) \left( \frac{10^{12}}{\langle d_1 \rangle} - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \frac{10^9}{\langle d_1 \rangle \langle d(1) + 1 \rangle} \right) - 1750$$

This time, we want to find the minimum of this quadratic polynomial. Since the leading coefficient is still negative, the minimum is reached for one of the extreme values $y/10 = 0$ or $y/10 = 100$, and we easily check the bound for those values.

We have proven that the algorithm gives us a precise enough answer. Incidentally, the upper bound that we derived tells us that

$$a < \frac{10^8}{\langle d \rangle} \leq 10^9,$$

hence we can compute $a$ safely as a TeX integer, and even add $10^9$ to it to ease grabbing of all the digits. The lower bound implies $10^8 - 1755 < a$, which we do not care about.

The \_fp\_ep\_div:wwwn function computes the ratio of two extended-precision numbers. The result is an extended-precision number whose first block lies in the range $[100, 9999]$, and is placed after the \(\langle\)continuation\rangle once we are done. First normalize the inputs so that both first block lie in $[1000, 9999]$, then call \_fp\_ep\_div\_esti:wwwwn (denominator \langle\)numerator\rangle, responsible for estimating the inverse of the denominator.

\begin{verbatim}
\cs_new:Npn \_fp\_ep\_div:wwwn #1,#2; #3,#4; }
  \__fp\_ep\_to\_ep:wwN #1,#2; \\_fp\_fixed\_continue:wn 
  {#2} #3; 
  \_fp\_ep\_div\_esti:wwwwn 
}
\end{verbatim}

The \_fp\_ep\_div\_esti:wwwn function evaluates $a = 10^9 / (\langle d_1 \rangle + 1)$, which is used twice in the expression for $a$, and combines the exponents #1 and #4 (with a shift by 1 because we later compute $\langle n \rangle/(10\langle d \rangle)$). Then the \_fp\_ep\_div\_esti:wwwn function evaluates $10^9 + a$, and puts the exponent #2 after the continuation #7: from there on we can forget exponents and focus on the mantissa. The \_fp\_ep\_div\_esti:wwwn function multiplies the denominator #7 by $10^{-8}a$ (obtained as $a$ split into the single digit #1 and two blocks of 4 digits, #2#3#4#5 and #6). The result $10^{-8}a(d) = (1 - \epsilon)$, and a partially packed $10^{-9}a$ (as a block of four digits, and five individual digits, not packed by lack of available macro parameters here) are passed to \_fp\_ep\_div\_epsi:wnNNNNn, which computes $10^{-9}a/(1 - \epsilon)$, that is, $1/(10\langle d \rangle)$ and we finally multiply this by the numerator #8.

\begin{verbatim}
\cs_new:Npn \_fp\_ep\_div\_esti:wwwn #1,#2; #3,#4; }
  \int_value:w \__fp_int_eval:w 10 0000 0000 / ( #2 + 1 ) 
  \exp_after:wN \_fp\_ep\_div\_epsi:wnNNNNn 
  \int_value:w \__fp_int_eval:w #4 - #1 + 1 , 
  \exp_after:wN \__fp\_int\_eval:w #2} #3; 
}
\end{verbatim}

The \_fp\_ep\_div\_epsi:wnNNNNn function computes $10^{-9}a/(1 - \epsilon)$, that is, $1/(10\langle d \rangle)$ and we finally multiply this by the numerator #8.

\begin{verbatim}
\cs_new:Npn \_fp\_ep\_div\_epsi:wnNNNNn #1,#2; #3,#4; #5; }
  \exp_after:wN \__fp\_int\_eval:w \_fp\_int\_eval:w #1 \_fp\_ep\_div\_epsi:wnNNNNn 
  \exp_after:wN \__fp\_int\_eval:w \_fp\_int\_eval:w 10 0000 0000 / ( #2 + 1 ) 
  \exp_after:wN \__fp\_int\_eval:w \int_value:w #3} #5; 
}
\end{verbatim}

The bounds shown above imply that the epsi function’s first operand is \((1 - \epsilon)\) with 
\(\epsilon \in [0, 1.755 \times 10^{-5}]\). The epsi function computes \(\epsilon\) as \(1 - (1 - \epsilon)\). Since \(\epsilon < 10^{-4}\), its first block vanishes and there is no need to explicitly use \#1 (which is 9999). Then epsii evaluates \(10^{-9}a/(1 - \epsilon)\) as \((1 + \epsilon^2)(1 + \epsilon)(10^{-9}a\epsilon) + 10^{-9}a\). Importantly, we compute \(10^{-9}a\epsilon\) before multiplying it with the rest, rather than multiplying by \(\epsilon\) and then \(10^{-9}a\), as this second option loses more precision. Also, the combination of short_mul and div_myriad is both faster and more precise than a simple mul.

(End definition for \_\_fp_ep_div_\_estiii:NNNNNw, \_\_fp_ep_div_\_estii:wwnnw, and \_\_fp_ep_div_\_estiii:NNNNNw.)
### 73.10 Inverse square root of extended precision numbers

The idea here is similar to division. Normalize the input, multiplying by powers of 100 until we have \( x \in [0.01, 1) \). Then find an integer approximation \( r \in [101, 1003] \) of \( 10^2/\sqrt{x} \), as the fixed point of iterations of the Newton method: essentially \( r \mapsto \left( r + \frac{10^8}{(x - r^2)} \right)/2 \), starting from a guess that optimizes the number of steps before convergence. In fact, just as there is a slight shift when computing divisions to ensure that some inequalities hold, we replace \( 10^8 \) by a slightly larger number which ensures that \( r^2 x \geq 10^4 \). This also causes \( r \in [101, 1003] \). Another correction to the above is that the input is actually normalized to \([0.1, 1)\), and we use either \( 10^8 \) or \( 10^9 \) in the Newton method, depending on the parity of the exponent. Skipping those technical hurdles, once we have the approximation \( r \), we set \( y = 10^{-4} r^2 x \) (or rather, the correct power of 10 to get \( y \approx 1 \)) and compute \( y - 1/2 \) through another application of Newton's method. This time, the starting value is \( z = 1 \), each step maps \( z \mapsto z(1.5 - 0.5 y z^2) \), and we perform a fixed number of steps. Our final result combines \( r \) with \( y - 1/2 \) as \( x - 1/2 = 10^{-2} ry - 1/2 \).

First normalize the input, then check the parity of the exponent \( \#1 \). If it is even, the result’s exponent will be \( -\#1/2 \), otherwise it will be \( (\#1 - 1)/2 \) (except in the case where the input was an exact power of 100). The auxii function receives as \#1 the result’s exponent just computed, as \#2 the starting value for the iteration giving \( r \) (the values 168 and 535 lead to the least number of iterations before convergence, on average), as \#3 and \#4 one empty argument and one 0, depending on the parity of the original exponent, as \#5 and \#6 the normalized mantissa (\( \#5 \in [1000, 9999] \)), and as \#7 the continuation. It sets up the iteration giving \( r \): the esti function thus receives the initial two guesses \#2 and 0, an approximation \#5 of \( 10^4 x \) (its first block of digits), and the empty/zero arguments \#3 and \#4, followed by the mantissa and an altered continuation where we have stored the result’s exponent.

If the last two approximations gave the same result, we are done: call the estii function to clean up. Otherwise, evaluate \((\langle\text{prev}\rangle + 1.005 \cdot 10^8)/(\langle\text{prev}\rangle \cdot x)\)/2, as the next approximation: omitting the 1.005 factor, this would be Newton’s method. We can
check by brute force that if \( #4 \) is empty (the original exponent was even), the process computes an integer slightly larger than \( \frac{100}{\sqrt{x}} \), while if \( #4 \) is 0 (the original exponent was odd), the result is an integer slightly larger than \( \frac{100}{\sqrt{x}/10} \). Once we are done, we evaluate \( 100r^2/2 \) or \( 10r^2/2 \) (when the exponent is even or odd, respectively) and feed that to \texttt{estiii}. This third auxiliary finds \( y_{\text{even}}/2 = 10^{-4}r^2x/2 \) or \( y_{\text{odd}}/2 = 10^{-5}r^2x/2 \) (again, depending on earlier parity). A simple program shows that \( y \in [1, 1.0201] \). The number \( y/2 \) is fed to \texttt{\_\_fp_ep_isqrt_estii:wN}, which computes \( 1/\sqrt{y} \), and we finally multiply the result by \( r. \)

\[
\texttt{\_\_fp_ep_isqrt_esti:wwwnn} \ #1, \ #2, \ #3, \ #4
\]

\[
\texttt{\_\_fp_ep_isqrt_estii:wwwnn} \ #1, \ #2, \ #3, \ #4\#5
\]

\[
\texttt{\_\_fp_ep_isqrt_estiii:NNNNNwwwn} \ #1\#2\#3\#4\#5\#6, \ #1\#7\#8; \ #9;
\]

\[
\texttt{\_\_fp_fixed_mul_short:wwn} \ #9; \ \{\#1 \ \#2\#3\#4\#5 \ \#600\} ;
\]

\[
\texttt{\_\_fp_ep_isqrt_epsii:wN}
\]

\[
\texttt{\_\_fp_fixed_mul_short:wwn} \ \{\#7 \ \#80 \ \{0000\} ;
\]

\[
\texttt{\_\_fp_ep_isqrt_epsii:wN}
\]

Here, we receive a fixed point number \( y/2 \) with \( y \in [1, 1.0201] \). Starting from \( z = 1 \) we iterate \( z \to z(3/2 - z^2y/2) \). In fact, we start from the first iteration \( z = 3/2 - y/2 \) to avoid useless multiplications. The \texttt{epsii} auxiliary receives \( z \) as \#1 and \( y \) as \#2.
Converting from fixed point to floating point

After computing Taylor series, we wish to convert the result from extended precision (with or without an exponent) to the public floating point format. The functions here should be called within an integer expression for the overall exponent of the floating point.

An extended-precision number is simply a comma-delimited exponent followed by a fixed point number. Leave the exponent in the current integer expression then convert the fixed point number.

```latex
\__fp_ep_to_float_o:wwN
\__fp_ep_inv_to_float_o:wwN
\__fp_fixed_inv_to_float_o:wN
\__fp_fixed_to_float_rad_o:wN
\__fp_fixed_to_float_o:wN
\__fp_fixed_to_float_o:Nw
```

(End definition for \_fp_ep_to_float_o:wwN and \_fp_ep_inv_to_float_o:wwN.)

Another function which reduces to converting an extended precision number to a float.

```
\cs_new:Npn \_fp_fixed_to_float_rad_o:wN #1; {
  \_fp_fixed_mul:wn #1; {5729}{5779}{5130}{8232}{0876}{7981};
  { \_fp_ep_to_float_o:wwN 2, }
}
```

(End definition for \_fp_fixed_to_float_rad_o:wN.)

Converts the fixed point number #1 from degrees to radians then to a floating point number. This could perhaps remain in l3fp-trig.

```
\cs_new:Npn \_fp_fixed_to_float_rad_o:wwN #1; {
  \_fp_fixed_to_loop:N 11
}
```

(End definition for \_fp_fixed_to_float_rad_o:wwN.)

And the to_fixed version gives six brace groups instead of 4, ensuring that 1000 ≤ \langle a_1 \rangle ≤ 9999. At this stage, we know that \langle a_1 \rangle is positive (otherwise, it is sign of an error before), and we assume that it is less than 10^8.\footnote{Bruno: I must double check this assumption.}

```
\cs_new:Npn \_fp_fixed_to_float_rad_o:Nw #1#2; {
  \_fp_fixed_to_float_o:wwN #2; #1 
}
```

\% for the 8-digit-at-the-start thing

```
\exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN \_fp_fixed_to_loop:N
```

1100
\cs_new:Npn \__fp_fixed_to_loop:N #1
\cs_new:Npn \__fp_fixed_to_loop_end:w #1 #2 ; 
\cs_new:Npn \__fp_fixed_to_float_zero:w ; 0000 0000 0000 0000 ; 
\cs_new:Npn \__fp_fixed_to_float_pack:ww #1 ; #2#3 ; ; 
\cs_new:Npn \__fp_fixed_to_float_round_up:wnnnw 
\cs_new:Npn \__fp_fixed_to_float_o:wN and \__fp_fixed_to_float_o:Nw. 

(End definition for \__fp_fixed_to_float_o:wN and \__fp_fixed_to_float_o:Nw.)
Chapter 74

l3fp-expo implementation

Unary functions.

\__fp_parse_word_exp:N
\__fp_parse_word_ln:N
\__fp_parse_word_fact:N

(End definition for \__fp_parse_word_exp:N, \__fp_parse_word_ln:N, and \__fp_parse_word_fact:N)

74.1 Logarithm

74.1.1 Work plan

As for many other functions, we filter out special cases in \__fp_ln_o:w. Then \__fp_-ln_npos_o:w receives a positive normal number, which we write in the form \(a \cdot 10^b\) with \(a \in [0,1)\).

The rest of this section is actually not in sync with the code. Or is the code not in sync with the section? In the current code, \(c \in [1,10]\) is such that \(0.7 \leq ac < 1.4\).

We are given a positive normal number, of the form \(a \cdot 10^b\) with \(a \in [0,1,1)\). To compute its logarithm, we find a small integer \(5 \leq c < 50\) such that \(0.91 \leq ac/5 < 1.04\), and use the relation

\[
\ln(a \cdot 10^b) = b \cdot \ln(10) - \ln(c/5) + \ln(ac/5).
\]

The logarithms \(\ln(10)\) and \(\ln(c/5)\) are looked up in a table. The last term is computed using the following Taylor series of \(\ln\) near 1:

\[
\ln \left( \frac{ac}{5} \right) = \ln \left( \frac{1 + t}{1 - t} \right) = 2t \left( 1 + t^2 \left( \frac{1}{3} + \frac{t^2}{5} + \frac{t^4}{9} + \cdots \right) \right)
\]

where \(t = 1 - 10/(ac + 5)\). We can now see one reason for the choice of \(ac \sim 5\): then \(ac + 5 = 10(1 - \epsilon)\) with \(-0.05 < \epsilon \leq 0.045\), hence

\[
t = \frac{\epsilon}{1 - \epsilon} = \epsilon(1 + \epsilon)(1 + \epsilon^2)(1 + \epsilon^4)\ldots,
\]

1102
is not too difficult to compute.

74.1.2 Some constants

A few values of the logarithm as extended fixed point numbers. Those are needed in the implementation. It turns out that we don’t need the value of ln(5).

\begin{verbatim}
\__fp_ln_i_fixed_tl \__fp_ln_ii_fixed_tl \__fp_ln_iii_fixed_tl \__fp_ln_iv_fixed_tl \__fp_ln_vi_fixed_tl \__fp_ln_vii_fixed_tl \__fp_ln_viii_fixed_tl \__fp_ln_ix_fixed_tl \__fp_ln_x_fixed_tl
\end{verbatim}

(End definition for \c__fp_ln_i_fixed_tl and others.)

74.1.3 Sign, exponent, and special numbers

The logarithm of negative numbers (including $-\infty$ and $-0$) raises the “invalid” exception. The logarithm of $+0$ is $-\infty$, raising a division by zero exception. The logarithm of $+\infty$ or a nan is itself. Positive normal numbers call \__fp_ln_npos_o:w.

\begin{verbatim}
\cs_new:Npn \__fp_ln_o:w #1 \s__fp \__fp.chk:w #2#3#4; @
{ \if_meaning:w 2 #3
  \__fp_case_use:nw { \__fp_invalid_operation_o:nw { ln } }
  \fi:
  \if_case:w #2 \exp_stop_f:
    \__fp_case_use:nw { \__fp_division_by_zero_o:Nnw \c_minus_inf_fp { ln } }
  \or:
    \__fp_case_return_same_o:w
  \fi:
  \__fp_ln_npos_o:w \s__fp \__fp.chk:w #2#3#4; }
\end{verbatim}

(End definition for \c__fp_ln_o:w.)

74.1.4 Absolute ln

We catch the case of a significand very close to 0.1 or to 1. In all other cases, the final result is at least $10^{-4}$, and then an error of $0.5 \cdot 10^{-20}$ is acceptable.

\begin{verbatim}
\cs_new:Npn \__fp_ln_npos_o:w \s__fp \__fp.chk:w \#2#3#4; \if_int_compare:w #1 < \c_one_int
  2
\else:
  0
\fi;
\end{verbatim}
This function expands to
\[
\text{continuation} \{ (Y_1) \} \{ (Y_2) \} \{ (Y_3) \} \{ (Y_4) \} \{ (Y_5) \} \{ (Y_6) \};
\]
where \( Y = -\ln(X) \) as an extended fixed point.

We have thus found \( c \in [1, 10] \) such that \( 0.7 \leq ac < 1.4 \) in all cases. Compute \( 1 + x = 1 + ac \in [1.7, 2.4) \).

\( \exp_{after}:wN \exp_{stop}:f; \)
\( \int_value:w _\_fp_int_eval:w \% \text{for the exponent} \)
\( _\_fp_ln_significand:NNNNnnnN \#2#3 \)
\( _\_fp_ln_exponent:wn \{#1\} \)
\( _\_fp_ln_npos_o:w \)
(End definition for \( _\_fp_ln_npos_o:w \)).

\( _\_fp_ln_significand:NNNNnnnN \)
\( _\_fp_ln_x_ii:wnnnn \)
(End definition for \( _\_fp_ln_significand:NNNNnnnN \)).

\( _\_fp_ln_x_ii:wnnnn \)
(End definition for \( _\_fp_ln_x_ii:wnnnn \)).
\begin{verbatim}
{1#2#3#4#5 + 1 ;}{1#2#3#4#5} {#6}
\end{verbatim}

The Taylor series to be used is expressed in terms of \( t = (x - 1)/(x + 1) = 1 - 2/(x + 1) \).
We now compute the quotient with extended precision, reusing some code from \texttt{fp_oo}. Note that \( 1 + x \) is known exactly.

To reuse notations from \texttt{l3fp-basics}, we want to compute \( A/Z \) with \( A = 2 \) and \( Z = x + 1 \). In \texttt{l3fp-basics}, we considered the case where both \( A \) and \( Z \) are arbitrary, in the range \([0.1, 1)\), and we had to monitor the growth of the sequence of remainders \( A, B, C, \) etc. to ensure that no overflow occurred during the computation of the next quotient.

The main source of risk was our choice to define the quotient as roughly \( 10^9 \cdot A/10^5 \cdot Z \): then \( A \) was bound to be below \( 2^{147} \cdots \), and this limit was never far.

In our case, we can simply work with \( 10^8 \cdot A \) and \( 10^4 \cdot Z \), because our reason to work with higher powers has gone: we needed the integer \( y \approx 10^5 \cdot Z \) to be at least \( 10^4 \), and now, the definition \( y \approx 10^4 \cdot Z \) suffices.

Let us thus define \( y = \lfloor 10^4 \cdot Z \rfloor + 1 \in (1.7 \cdot 10^4, 2.4 \cdot 10^4] \), and

\[
Q_1 = \left\lfloor \frac{10^8 \cdot A}{y} - \frac{1}{2} \right\rfloor.
\]

(The 1/2 comes from how \( \varepsilon \)-\TeX{} rounds.) As for division, it is easy to see that \( Q_1 \leq 10^4 A/Z \), i.e., \( Q_1 \) is an underestimate.

Exactly as we did for division, we set \( B = 10^4 A - Q_1 Z \). Then

\[
10^4 B \leq A_1 A_2 A_3 A_4 - \left( \frac{A_1 A_2}{y} - \frac{3}{2} \right) 10^4 Z
\]

\[
\leq A_1 A_2 \left( 1 - \frac{10^4 Z}{y} \right) + 1 + \frac{3}{2} y
\]

\[
\leq 10^8 \frac{A}{y} + 1 + \frac{3}{2} y
\]

In the same way, and using \( 1.7 \cdot 10^4 \leq y \leq 2.4 \cdot 10^4 \), and convexity, we get

\[
10^4 A = 2 \cdot 10^4
\]

\[
10^4 B \leq 10^8 \frac{A}{y} + 1.6 y \leq 4.7 \cdot 10^4
\]

\[
10^4 C \leq 10^8 \frac{B}{y} + 1.6 y \leq 5.8 \cdot 10^4
\]

\[
10^4 D \leq 10^8 \frac{C}{y} + 1.6 y \leq 6.3 \cdot 10^4
\]

\[
10^4 E \leq 10^8 \frac{D}{y} + 1.6 y \leq 6.5 \cdot 10^4
\]

\[
10^4 F \leq 10^8 \frac{E}{y} + 1.6 y \leq 6.6 \cdot 10^4
\]

Note that we compute more steps than for division: since \( t \) is not the end result, we need to know it with more accuracy (on the other hand, the ending is much simpler, as we don’t need an exact rounding for transcendental functions, but just a faithful rounding).
The number is \( x \). Compute \( y \) by adding 1 to the five first digits.

We now have essentially

\[
\__fp_div_significand_pack:NNN \ 10^6 + Q_1; \ 10^6 + Q_2; \ 10^6 + Q_3; \ 10^6 + Q_4; \ 10^6 + Q_5; \ 10^6 + Q_6; \langle \text{exponent} \rangle; \langle \text{continuation} \rangle
\]

where \( \langle \text{fixed tl} \rangle \) holds the logarithm of a number in \([1, 10]\), and \( \langle \text{exponent} \rangle \) is the exponent. Also, the expansion is done backwards. Then \( \__fp_div_significand_pack:NNN \) puts things in the correct order to add the \( Q_i \) together and put semicolons between each piece. Once those have been expanded, we get

\[
\__fp_ln_x_iv:wnnnnnnn \ 1 \ or \ 2 \ \langle 8d \rangle ; \ \langle 4d \rangle \ \langle 4d \rangle \ \langle 4d \rangle \ \langle 8d \rangle ; \ \langle 4d \rangle \ \langle 4d \rangle \ \langle 4d \rangle \ \langle 4d \rangle \ \langle 4d \rangle \ \langle \text{fixed-tl} \rangle
\]
Just as with division, we know that the first two digits are 1 and 0 because of bounds on the final result of the division \( \frac{2}{(x + 1)} \), which is between roughly 0.8 and 1.2. We then compute \( 1 - \frac{2}{(x + 1)} \), after testing whether \( \frac{2}{(x + 1)} \) is greater than or smaller than 1.

\[
\texttt{\_fp\_ln\_div\_after:Nw} \#1\#2;
\]

\[
\texttt{\_fp\_ln\_t\_small:Nw} \#1 \#2; \#3; \#4; \#5; \#6; \#7;
\]

\[
\texttt{\_fp\_ln\_t\_large:NNw} (\text{sign}) \langle \text{fixed tl} \rangle \langle t_1 \rangle; \langle t_2 \rangle; \langle t_3 \rangle; \langle t_4 \rangle; \langle t_5 \rangle; \langle t_6 \rangle; \langle \text{exponent} \rangle; \langle \text{continuation} \rangle
\]

Compute the square \( t^2 \), and keep \( t \) at the end with its sign. We know that \( t < 0.1765 \), so every piece has at most 4 digits. However, since we were not careful in \_fp\_ln\_t\_small:w, they can have less than 4 digits.
\int_value:w \__fp_int_eval:w 9999 + 2*#5
\exp_after:wN \__fp_int_twice_t_pack:Nw
\int_value:w \__fp_int_eval:w 9999 + 2*#6
\exp_after:wN \__fp_int_twice_t_pack:Nw
\int_value:w \__fp_int_eval:w 9999 + 2*#7
\exp_after:wN \__fp_int_twice_t_pack:Nw
\int_value:w \__fp_int_eval:w 10000 + 2*#8 ; ;
{ \__fp_ln_c:NwNw #1 }
#2
}
\cs_new:Npn \__fp_ln_twice_t_pack:Nw #1 #2; { + #1 ; {#2} }
\cs_new:Npn \__fp_ln_twice_t_after:w #1; { ;;; {#1} }
\cs_new:Npn \__fp_ln_square_t_pack:NNNNNw #1 #2#3#4#5 #6; { + #1#2#3#4#5 ; {#6} }
\cs_new:Npn \__fp_ln_square_t_after:w 1 0 #1#2#3 #4; { \__fp_ln_Taylor:wwNw {0#1#2#3} {#4} }

(End definition for \__fp_ln_x_ii:wnnnn.)
\__fp_ln_Taylor:wwNw

Denoting $T = t^2$, we get
\__fp_ln_Taylor:wwNw
{(T_1)} {(T_2)} {(T_3)} {(T_4)} ; ;
{(2t)} {(2t)} {(2t)} {(2t)} ; ;
\__fp_ln_Taylor:wwNw (sign) ; ;
\__fp_ln_Taylor:wwNw (fixed tl) (exponent) ; ;(continuation)

And we want to compute
$$\ln \left( \frac{1+t}{1-t} \right) = 2t \left( 1 + T \left( \frac{1}{3} + T \left( \frac{1}{5} + T \left( \frac{1}{7} + T \left( \frac{1}{9} + \cdots \right) \right) \right) \right)$$

The process looks as follows
\\loop 5; A;
\\div_int 5; 1.0; \add A; \mul T; {\\loop \eval 5-2;}
\add 0.2; A; \mul T; {\\loop \eval 5-2;}
\mul B; T; {\\loop 3;}
\loop 3; C;

This uses the routine for dividing a number by a small integer ($< 10^4$).
We are now reduced to finding $\ln(c)$ and $\langle$exponent$\rangle \ln(10)$ in a table, and adding it to the mixture. The first step is to get $\ln(c) - \ln(x) = -\ln(a)$, then we get $b \ln(10)$ and add or subtract.

For now, $\ln(x)$ is given as $\cdot 10^0$. Unless both the exponent is 1 and $c = 1$, we shift to working in units of $\cdot 10^4$, since the final result is at least $\ln(10^7) \simeq 3.5$.

Compute $\langle$exponent$\rangle \times \ln(10)$. Apart from the cases where $\langle$exponent$\rangle$ is 0 or 1, the result is necessarily at least $\ln(10) \approx 2.3$ in magnitude. We can thus drop the least significant 4 digits. In the case of a very large (positive or negative) exponent, we can (and we need to) drop 4 additional digits, since the result is of order $10^4$. Naively, one would think that in both cases we can drop 4 more digits than we do, but that would be slightly too tight for rounding to happen correctly. Besides, we already have addition and subtraction for 24 digits fixed point numbers.

We are now reduced to finding $\ln(c)$ and $\langle$exponent$\rangle \ln(10)$ in a table, and adding it to the mixture. The first step is to get $\ln(c) - \ln(x) = -\ln(a)$, then we get $b \ln(10)$ and add or subtract.

For now, $\ln(x)$ is given as $\cdot 10^0$. Unless both the exponent is 1 and $c = 1$, we shift to working in units of $\cdot 10^4$, since the final result is at least $\ln(10^7) \simeq 3.5$.
Now we painfully write all the cases.¹² No overflow nor underflow can happen, except when computing \( \ln(1) \).

For small exponents, we just drop one block of digits, and set the exponent of the log to 4 (minus any shift coming from leading zeros in the conversion from fixed point to floating point). Note that here the exponent has been made positive.

(End definition for \texttt{\_\_fp\_ln\_exponent:wm}.)

### 74.2 Exponential

#### 74.2.1 Sign, exponent, and special numbers

¹²Bruno: do rounding.
\or:
\__fp_case_return_same:o:w
\fi:
\s__fp \__fp_chk:w #2#3#4;
}

(End definition for \__fp_exp:o:w)

\cs_new:Npn \__fp_exp_normal:o:w \s__fp \__fp_chk:w 1#1
{\if_meaning:w 0 #1
  \__fp_exp_pos:o:NNwnw + \__fp_fixed_to_float:o:wN
\else:
  \__fp_exp_pos:o:NNwnw - \__fp_fixed_inv_to_float:o:wN
  \fi:}
\cs_new:Npn \__fp_exp_pos:o:NNwnw #1#2#3 \fi: #4#5;
{\fi:
  \if_int_compare:w #4 > \c__fp_max_exp_exponent_int
  \token_if_eq_charcode:NNTF + #1
  \{ \__fp_exp_overflow:o:NN \__fp_overflow:w \c_inf_fp \}
  \{ \__fp_exp_overflow:o:NN \__fp_underflow:w \c_zero_fp \}
  \exp:w
  \exp_after:wN \__fp_sanitize:Nw
  \exp_after:wN 0
  \int_value:w #1 \__fp_int_eval:w
  \if_int_compare:w #4 < \c_zero_int
    \exp_after:wN \use_i:nn
  \else:
    \exp_after:wN \use_ii:nn
  \fi:
{ \fi:
  0
  \__fp_decimate:nNnnnn { - #4 }
  \__fp_exp_Taylor:Nnnnn
}
{ \__fp_decimate:nNnnnn { \c__fp_prec_int - #4 }
  \__fp_exp_pos_large:Nnnnn
}#5
#{4}
\#1 #2 0
\exp:w
\fi:
\exp_after:wN \exp_end:
}
\cs_new:Npn \__fp_exp_overflow:o:NN #1#2\#3 \fi: #4#5;
{\fi:
  \if_int_compare:w #4 > \c__fp_max_exp_exponent_int
  \token_if_eq_charcode:NNTF + #1
  \{ \__fp_exp_overflow:o:NN \__fp_overflow:w \c_inf_fp \}
  \{ \__fp_exp_overflow:o:NN \__fp_underflow:w \c_zero_fp \}
  \exp:w
  \exp_after:wN \__fp_sanitize:Nw
  \exp_after:wN 0
  \int_value:w #1 \__fp_int_eval:w
  \if_int_compare:w #4 < \c_zero_int
    \exp_after:wN \use_i:nn
  \else:
    \exp_after:wN \use_ii:nn
  \fi:
\fi:
\exp_after:wN \exp_end:
}
\cs_new:Npn \__fp_exp_overflow:o:NN #1#2\#3 \fi: #4#5;
{\fi:
This function is called for numbers in the range \(10^{-9}, 10^{-1}\). We compute 10 terms of the Taylor series. The first argument is irrelevant (rounding digit used by some other functions). The next three arguments, at least 16 digits, delimited by a semicolon, form a fixed point number, so we pack it in blocks of 4 digits.

\begin{verbatim}
\cs_new:Npn \_fp_exp_Taylor:Nnnwn #1#2#3 #4; #5 #6
{ #6 \_fp_pack_twice_four:wNNNNNNNN \_fp_pack_twice_four:wNNNNNNNN \_fp_pack_twice_four:wNNNNNNNN \_fp_exp_Taylor_ii:ww
; #2#3#4 0000 0000 ;}
\cs_new:Npn \_fp_exp_Taylor_ii:ww #1; #2;
{ \_fp_exp_Taylor_loop:www 10 ; #1 ; #1 ; \s__fp_stop }
\cs_new:Npn \_fp_exp_Taylor_loop:www #1; #2; #3;
{ \if_int_compare:w #1 = \c_one_int \exp_after:wN \_fp_exp_Taylor_break:Nww \fi:
\_fp_fixed_div_int:wwN #3 ; #1 ; \_fp_fixed_add_one:wn #2 ;
{ \exp_after:wN \_fp_exp_Taylor_loop:www \int_value:w \_fp_int_eval:w #1 - 1 ; #2 ;}
}
\cs_new:Npn \_fp_exp_Taylor_break:Nww #1 #2; #3 \s__fp_stop
{ \_fp_fixed_add_one:wn #2 ; }
\end{verbatim}

The integer array has \(6 \times 9 \times 4 = 216\) items encoding the values of \(\exp(j \times 10^i)\) for \(j = 1, \ldots, 9\) and \(i = -1, \ldots, 4\). Each value is expressed as \(\approx 10^p \times 0.m_1m_2m_3\) with three 8-digit blocks \(m_1\), \(m_2\), \(m_3\) and an integer exponent \(p\) (one more than the scientific exponent), and these are stored in the integer array as four items: \(p, 10^8 + m_1, 10^8 + m_2, 10^8 + m_3\). The various exponentials are stored in increasing order of \(j \times 10^i\).

Storing this data in an integer array makes it slightly harder to access (slower, too), but uses 16 bytes of memory per exponential stored, while storing as tokens used around 40 tokens; tokens have an especially large footprint in Unicode-aware engines.

\begin{verbatim}
\intarray_const_from_clist:Nn \c__fp_exp_intarray
{ 1 , 1 1105 1709 , 1 1807 5647 , 1 6248 1171 , 1 , 1 1221 4027 , 1 5816 0169 , 1 8339 2107 , 1 , 1 1349 8588 , 1 0757 6003 , 1 1039 8374 , 1 , 1 1491 8246 , 1 9764 1270 , 1 3178 2485 ,
}
\end{verbatim}
The first two arguments are irrelevant (a rounding digit, and a brace group with 8 zeros).

1113
The third argument is the integer part of our number, then we have the decimal part delimited by a semicolon, and finally the exponent, in the range $[0, 5]$. Remove leading zeros from the integer part: putting #4 in there too ensures that an integer part of 0 is also removed. Then read digits one by one, looking up $\exp((\text{digit}) \cdot 10^{\text{(exponent)}})$ in a table, and multiplying that to the current total. The loop is done by \texttt{\_\_fp\_exp\_large:NwN}, whose #1 is the (exponent), #2 is the current mantissa, and #3 is the (digit). At the end, \texttt{\_\_fp\_exp\_large\_after:wwn} moves on to the Taylor series, eventually multiplied with the mantissa that we have just computed.

\begin{verbatim}
\cs_new:Npn \__fp_exp_pos_large:NnnNwn #1#2#3 #4#5; #6
\exp_after:wN \exp_after:wN \exp_after:wN \__fp_exp_large:NwN
\exp_after:wN \exp_after:wN \exp_after:wN \c__fp_one_fixed_tl
\int_value:w #3 #4 \exp_stop_f:
\__fp_exp_large_after:wwn
\__fp_exp_intarray:w #1 #2 #3
\exp_after:wN \__fp_fixed_continue:wn
\exp_after:wN \__fp_exp_intarray:w
\int_value:w \__fp_int_eval:w 36 * #1 + 4 * #3 \exp_after:wN \scan_stop:
\__fp_exp_intarray_aux:w 1 #1#2#3#4#5 ;
\cs_new:Npn \__fp_exp_intarray_aux:w #1 \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 - 2 } \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 - 1 } \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 - 0 } \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 + 1 } \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 + 2 } \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 + 3 } \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 + 4 } \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 + 5 } \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 + 6 }
\end{verbatim}

1114
74.3 Power

Raising a number \( a \) to a power \( b \) leads to many distinct situations.

<table>
<thead>
<tr>
<th>( a^b )</th>
<th>(-\infty )</th>
<th>((\infty, -0))</th>
<th>(-\text{integer} )</th>
<th>( \pm 0 )</th>
<th>(+\text{integer} )</th>
<th>((0, \infty))</th>
<th>(+\infty)</th>
<th>( \text{NaN} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( +\infty )</td>
<td>(+0)</td>
<td>(+0)</td>
<td>(+1)</td>
<td>(+\infty)</td>
<td>(+\infty)</td>
<td>( \text{NaN} )</td>
<td>( \text{NaN} )</td>
<td>( \text{NaN} )</td>
</tr>
<tr>
<td>(1, ( \infty ))</td>
<td>(+0)</td>
<td>(+</td>
<td>a</td>
<td>^b)</td>
<td>(+1)</td>
<td>(+</td>
<td>a</td>
<td>^b)</td>
</tr>
<tr>
<td>(+1)</td>
<td>(+1)</td>
<td>(+1)</td>
<td>(+1)</td>
<td>(+1)</td>
<td>(+1)</td>
<td>( \text{NaN} )</td>
<td>( \text{NaN} )</td>
<td>( \text{NaN} )</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>(+\infty)</td>
<td>(+</td>
<td>a</td>
<td>^b)</td>
<td>(+1)</td>
<td>(+</td>
<td>a</td>
<td>^b)</td>
</tr>
<tr>
<td>(+0)</td>
<td>(+\infty)</td>
<td>(+\infty)</td>
<td>(+1)</td>
<td>(+0)</td>
<td>(+0)</td>
<td>( \text{NaN} )</td>
<td>( \text{NaN} )</td>
<td>( \text{NaN} )</td>
</tr>
<tr>
<td>(-0)</td>
<td>(+\infty)</td>
<td>(\text{NaN})</td>
<td>((-1)^b\infty)</td>
<td>(+1)</td>
<td>((-1)^b0)</td>
<td>(+0)</td>
<td>(+0)</td>
<td>( \text{NaN} )</td>
</tr>
<tr>
<td>((-1, 0))</td>
<td>(+\infty)</td>
<td>(\text{NaN})</td>
<td>((-1)^b</td>
<td>a</td>
<td>^b)</td>
<td>(+1)</td>
<td>((-1)^b</td>
<td>a</td>
</tr>
<tr>
<td>(-1)</td>
<td>(+1)</td>
<td>(\text{NaN})</td>
<td>((-1)^b)</td>
<td>(+1)</td>
<td>((-1)^b)</td>
<td>(\text{NaN})</td>
<td>(+1)</td>
<td>( \text{NaN} )</td>
</tr>
<tr>
<td>((\infty, -1))</td>
<td>(+0)</td>
<td>(\text{NaN})</td>
<td>((-1)^b</td>
<td>a</td>
<td>^b)</td>
<td>(+1)</td>
<td>((-1)^b</td>
<td>a</td>
</tr>
<tr>
<td>(-\infty)</td>
<td>(+0)</td>
<td>(+0)</td>
<td>((-1)^b0)</td>
<td>(+1)</td>
<td>((-1)^b\infty)</td>
<td>(\text{NaN})</td>
<td>(+\infty)</td>
<td>( \text{NaN} )</td>
</tr>
<tr>
<td>(\text{NaN})</td>
<td>(\text{NaN})</td>
<td>(\text{NaN})</td>
<td>(\text{NaN})</td>
<td>(+1)</td>
<td>(\text{NaN})</td>
<td>(\text{NaN})</td>
<td>(\text{NaN})</td>
<td>( \text{NaN} )</td>
</tr>
</tbody>
</table>

We distinguished in this table the cases of finite (positive or negative) integer exponents, as \((-1)^b\) is defined in that case. One peculiarity of this operation is that \(\text{NaN}^0 = 1\) \(\text{NaN}\), because this relation is obeyed for any number, even \(\pm\infty\).

\(\text{\_fp\textunderscore ^\textunderscore o:ww}\) We cram most of the tests into a single function to save csnames. First treat the case \(b = 0\): \(a^0 = 1\) for any \(a\), even \(\text{NaN}\). Then test the sign of \(a\).

- If it is positive, and \(a\) is a normal number, call \(\text{\_fp\textunderscore pow\textunderscore normal\textunderscore o:ww}\) followed by the two \text{fp} \(a\) and \(b\). For \(a = +0\) or \(+\infty\), call \(\text{\_fp\textunderscore pow\textunderscore zero\textunderscore or\textunderscore inf:ww}\) instead, to return either \(+0\) or \(+\infty\) as appropriate.

- If \(a\) is a \(\text{NaN}\), then skip to the next semicolon (which happens to be conveniently the end of \(b\)) and return \(\text{NaN}\).

- Finally, if \(a\) is negative, compute \(a^b\) (\(\text{\_fp\textunderscore pow\textunderscore normal\textunderscore o:ww\) which ignores the sign of its first operand}), and keep an extra copy of \(a\) and \(b\) (the second brace group, containing \{ \(b\) \(a\) \}, is inserted between \(a\) and \(b\)). Then do some tests to find the final sign of the result if it exists.
\__fp_pow_zero_or_inf:ww
Raising \(-0\) or \(-\infty\) to \texttt{nan} yields \texttt{nan}. For other powers, the result is \(+0\) if 0 is raised to a positive power or \(+\infty\), and \(+\infty\) otherwise. Thus, if the type of \(a\) and the sign of \(b\) coincide, the result is 0, since those conveniently take the same possible values, 0 and 2. Otherwise, either \(a = \pm\infty\) and \(b > 0\) and the result is \(+\infty\), or \(a = \pm0\) with \(b < 0\) and we have a division by zero unless \(b = -\infty\).

\__fp_pow_normal_o:ww
We have in front of us \(a\), and \(b \neq 0\), we know that \(a\) is a normal number, and we wish to compute \(|a|^b\). If \(|a| = 1\), we return 1, unless \(a = -1\) and \(b\) is \texttt{nan}. Indeed, returning 1 at this point would wrongly raise “invalid” when the sign is considered. If \(|a| \neq 1\), test the type of \(b\):

0 Impossible, we already filtered \(b = \pm0\).
1 Call \_\_fp_pow\_npos\_o:Nww.

2 Return +∞ or +0 depending on the sign of b and whether the exponent of a is positive or not.

3 Return b.

\cs_new:Npn \_\_fp_pow\_normal\_o:ww \s__fp \_\_fp\_chk:w #1 #2 #3 #4 #5
\{ 
\if:w 0 \_\_fp\_str\_if:eq:n { #2 #3 } { 1 {1000} {0000} {0000} {0000} } \if_int_compare:w #4 #1 = 32 \exp_stop_f: \exp_after:wN \_\_fp\_case\_return\_ii\_o:ww \fi: \_\_fp\_case\_return\_o:Nww \c_one_fp \fi: \if_case:w #4 \exp_stop_f: \or: \exp_after:wN \_\_fp\_pow\_npos\_aux:NNww \exp_after:wN + \_\_fp\_fixed\_to\_float\_o:wN \fi: \fi: \_\_fp\_case\_return\_ii\_o:ww \fi: \_\_fp\_chk:w 1 #1 #2 #3 ; \_\_fp\_chk:w #4 #5 \}
\end{definition} for \_\_fp\_pow\_normal\_o:ww

\_\_fp_pow\_npos\_o:Nww We now know that \( a \neq \pm 1 \) is a normal number, and \( b \) is a normal number too. We want to compute \( |
\begin{align*}
\exp(\ln |x| + n \ln(10)) \cdot y \cdot 10^p) = \exp(z).
\end{align*}
\end{frac}

To compute the exponential accurately, we need to know the digits of \( z \) up to the 16-th position. Since the exponential of \( 10^5 \) is infinite, we only need at most 21 digits, hence the fixed point result of \( \_\_fp\_ln\_o:w \) is precise enough for our needs. Start an integer expression for the decimal exponent of \( e^z \). If \( z \) is negative, negate that decimal exponent, and prepare to take the inverse when converting from the fixed point to the floating point result.

\cs_new:Npn \_\_fp_pow\_npos\_o:Nww \s__fp \_\_fp\_chk:w #1 \_\_fp\_sanitize:Nw \exp_after:wN 0 \_\_fp\_value:w
\if:w #1 \if_int_compare:w #3 > \c_zero_int 0 \else: 2 \fi: \exp_after:wN \_\_fp\_pow\_npos\_aux:NNww \exp_after:wN + \exp_after:wN \_\_fp\_fixed\_to\_float\_o:wN
\_\_p \_o \_n \_p \_o \_a \_u \_x:NNnw

The first argument is the conversion function from fixed point to float. Then comes an exponent and the 4 brace groups of \textit{x}, followed by \textit{b}. Compute $-\ln(x)$.

\texttt{\cs_new:Npn \_\_fp \_pow \_n \_p \_o \_a \_x:NNnw \#1\#2\#3\#4\#5; \_s \_fp \_\_fp \_c \_h \_k:w 1\#6\#7\#8;}

\{
  \#1
  \_\_fp \_i \_n \_e \_v \_a \_l:w
  \_\_fp \_l \_n \_s \_i \_g \_i \_c \_a \_n \_d:NNNNnnn \#4#5
  \_\_fp \_p \_o \_w \_e \_x \_n \_o \_e \_n \_t:Nw \#3
  \_\_fp \_f \_i \_x \_e \_d \_i \_m \_u \_l:Nw \#5 \{0000\}(0000) ;
  \_\_fp \_f \_i \_x \_e \_d:B:wwN \#7;
  \#1 \#2 0 \% \_f \_i \_x \_t \_d \_i \_f \_l \_o \_t \_o \_f \_l \_o \_f:o:w
\}
\texttt{\cs_new:Npn \_\_fp \_p \_o \_w \_e \_x \_n \_o \_e \_n \_t:Nw \#1; \#2}
\{
  \_f \_i \_n \_c \_o \_m \_p \_a:w \#2 > \_c \_z \_e \_r \_o \_i \_n \t
  \_e \_x \_p \_a:w \_\_fp \_p \_o \_w \_e \_x \_n \_o \_e \_n \_t:Nw \#n\ln(10) - (-\ln(x))
  \_e \_x \_p \_a:w +
  \_e \_x \_p \_a:w \_\_fp \_p \_o \_w \_e \_x \_n \_o \_e \_n \_t:Nw \#n\ln(10) + (-\ln(x))
  \_e \_x \_p \_a:w -
  \_f \_i:
  \#2; \#1;
\}
\texttt{\cs_new:Npn \_\_fp \_p \_o \_w \_e \_x \_n \_o \_e \_n \_t:Nw \#1\#2; \#3\#4\#5\#6\#7\#8;}
\{
  \%"A todo: use that in ln.
  \_e \_x \_p \_a:w \_\_fp \_f \_i \_x \_e \_d \_m \_u \_l:w:w \_\_fp \_l \_n \_s \_i \_g \_i \_c \_a \_n \_d:NNNNnw \#1#2#3#4#5#6#7#8;
  \_f \_i \_n \_c \_o \m \_p \_a:w \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5
  \_e \_x \_p \_a:w \_\_fp \_l \_n \_s \_i \g \_i \c \_a \_n \_d:NNNNnw \#1 #2+2994 - \#1 \#5

1118
\cs_new:Npn \__fp_pow_B:wwN \#1\#2\#3\#4\#5\#6; \#7; 
{ 
  \if_int_compare:w \#7 < \c_zero_int 
    \exp_after:wN \__fp_pow_C_neg:w \int_value:w - 
  \else: 
    \if_int_compare:w \#7 < 22 \exp_stop_f: 
      \exp_after:wN \__fp_pow_C_pos:w \int_value:w 
    \else: 
      \exp_after:wN \__fp_pow_C_overflow:w \int_value:w 
    \fi: 
  \fi: 
  \#7 \exp_after:wN ; 
  \int_value:w \__fp_int_eval:w 10 0000 + \#1 \__fp_int_eval_end: 
} \cs_new:Npn \__fp_pow_C_overflow:w \#1; \#2; \#3 
{ 
  \c__fp_max_exponent_int + 2 * \c__fp_max_exponent_int 
  \exp_after:wN \__fp_fixed_continue:wn \c__fp_one_fixed_tl 
} \cs_new:Npn \__fp_pow_C_neg:w \#1; 1 
{ \__fp_pow_C_pack:w 
  \prg_replicate:nn \{\#1\} \{0\} } \cs_new:Npn \__fp_pow_C_pos:w \#1; 1 
{ \__fp_pow_C_pos_loop:wN \#1; } \cs_new:Npn \__fp_pow_C_pos_loop:wN \#1; \#2 
{ \if_meaning:w 0 \#1 
    \exp_after:wN \__fp_pow_C_pack:w 
  \exp_after:wN \#2 
  \else: 
    \if_meaning:w 0 \#2 
      \exp_after:wN \__fp_pow_C_pos_loop:wN \int_value:w 
    \else: 
      \exp_after:wN \__fp_pow_C_overflow:w \int_value:w 
    \fi: 
    \__fp_int_eval:w \#1 - 1 \exp_after:wN ; 
  \fi: 
} \cs_new:Npn \__fp_pow_C_pack:w 
{ \exp_after:wN \__fp_exp_large:NwN \exp_after:wN 5 \c__fp_one_fixed_tl } 
(End definition for \__fp_pow_npos_aux:NNww.)

This function \textit{is} followed by three floating point numbers: \(a^b\), \(a \in [-\infty, -0]\), and \(b\). If \(b\) is an even integer (case \(-1\)), \(a^b = a^b\). If \(b\) is an odd integer (case \(0\)), \(a^b = -a^b\), obtained by a call to \__fp_pow_neg_aux:NNww. Otherwise, the sign is undefined. This is invalid, unless \(a^b\) turns out to be +0 or \texttt{nan}, in which case we return that as \(a^b\). In particular, since the
underflow detection occurs before `\_fp_pow_neg:ww` is called, \((-0.1)**(12345.67)`
gives \(+0\) rather than complaining that the sign is not defined.

\begin{verbatim}
\cs_new:Npn \_fp_pow_neg:www \s__fp \_fp_chk:w #1#2; #3; #4;
{ \if_case:w \_fp_pow_neg_case:w #4 ; \exp_after:wN \_fp_pow_neg_aux:wNN \or:
  \if_int_compare:w \_fp_int_eval:w #1 / 2 = \c_one_int
  \_fp_invalid_operation:o:NNw ^ #3; #4;
  \exp:w \exp_end_continue_f:w
  \exp_after:wN \exp_after:wN \_fp_after:w \_fp_after:w \_fp_use_none_until:s:w
  \fi:
  \fi:
\__fp_exp_after_o:w \_fp_chk:w #1#2;
}\}
\cs_new:Npn \_fp_pow_neg_case:w \s__fp \_fp_chk:w #1#2#3; \_fp_pow_neg_case_aux:ww \_fp_pow_neg_case_aux:NNw
This function expects a floating point number, and determines its “parity”. It should
be used after `\if_case:w` or in an integer expression. It gives \(-1\) if the number is an
even integer, 0 if the number is an odd integer, and 1 otherwise. Zeros and ±\(\infty\) are even
(because very large finite floating points are even), while `nan` is a non-integer. The sign
of normal numbers is irrelevant to parity. After `\_fp_decimate:nNnnnn` the argument
#1 of `\_fp_pow_neg_case_aux:NNww` is a rounding digit, 0 if and only if the number
was an integer, and #3 is the 8 least significant digits of that integer.
\begin{verbatim}
\cs_new:Npn \_fp_pow_neg_case:w \s__fp \_fp_chk:w #1#2#3;
{ \if_case:w #1 \exp_stop_f:
  -1 \or: \_fp_pow_neg_case_aux:NNNNN \_fp_pow_neg_case_aux:nnnn #3 \or:
    -1 \else: 1 \fi:
  \exp_stop_f:
}\}
\cs_new:Npn \_fp_pow_neg_case_aux:NNNNN \_fp_pow_neg_case_aux:nnnn \_fp_pow_neg_case_aux:Nnnw
\_fp_decimate:nNnnnn \c__fp_prec_int \_fp_pow_neg_case_aux:nnnn \_fp_pow_neg_case_aux:NnNn
{ \c__fp_prec_int - #1 } \_fp_pow_neg_case_aux:Nnnw \_fp_pow_neg_case_aux:NnNn
{#2} {#3} {#4} {#5}
\end{verbatim}

1120
\fi:
}\fi:
\cs_new:Npn \__fp_pow_neg_case_aux:Nnnw #1#2#3#4 ;
\if_meaning:w 0 #1
\if_int_odd:w #3 \exp_stop_f:
  0
\else:
  -1
\fi:
\else:
  1
\fi:
}

(End definition for \__fp_pow_neg_case:w, \__fp_pow_neg_case_aux:nwnn, and \__fp_pow_neg_case_aux:Nnnw.)

74.4 Factorial

\c__fp_fact_max_arg_int
The maximum integer whose factorial fits in the exponent range is 3248, as 3249! ∼ \int_{10^{10000}}.

\int_const:Nn \c__fp_fact_max_arg_int { 3248 }
(End definition for \c__fp_fact_max_arg_int.)

\__fp_fact_o:w
First detect \pm 0 and +∞ and nan. Then note that factorial of anything with a negative sign (except −0) is undefined. Then call \__fp_small_int:TF to get an integer as the argument, and start a loop. This is not the most efficient way of computing the factorial, but it works all right. Of course we work with 24 digits instead of 16. It is easy to check that computing factorials with this precision is enough.

\cs_new:Npn \__fp_fact_o:w #1 \s__fp \__fp_chk:w #2#3#4; @
\if_case:w #2 \exp_stop_f:
  \__fp_case_return_o:Nw \c_one_fp
\or:
  \__fp_case_return_same_o:w
\or:
  \__fp_case_use:nw { \__fp_invalid_operation_o:fw { fact } }
\fi:
\__fp_fact_pos_o:w
\s__fp \__fp_chk:w #2 #3 #4 ;
}

(End definition for \__fp_fact_o:w.)
Then check the input is an integer, and call \_fp_factorial_int:o:n with that int as an argument. If it's too big the factorial overflows. Otherwise call \_fp-sanitize:Nw with a positive sign marker 0 and an integer expression that will mop up any exponent in the calculation.

\_fp_fact_pos:o:w\_fp_fact_int:o:w

The loop receives an integer \#1 whose factorial we want to compute, which we progressively decrement, and the result so far as an extended-precision number \#2 in the form ⟨exponent⟩,⟨mantissa⟩. The loop goes in steps of two because we compute \#1*\#1-1 as an integer expression (it must fit since \#1 is at most 3248), then multiply with the result so far. We don’t need to fill in most of the mantissa with zeros because \_fp-ep_mul:wwwwn first normalizes the extended precision number to avoid loss of precision. When reaching a small enough number simply use a table of factorials less than 10^8. This limit is chosen because the normalization step cannot deal with larger integers.

\_fp_fact_loop:o:w
\exp_after:wN \__fp_mul:wwww
\exp_after:wN 4 \exp_after:wN ,
\exp_after:wN
{
\int_value:w
\if_case:w #1 \exp_stop_f:
1 \or: 1 \or: 2 \or: 6 \or: 24 \or: 120 \or: 720 \or: 5040
\or: 40320 \or: 362880 \or: 3628800 \or: 39916800
\fi:
\}
\}
\}
\}
\}
#3 ;
\__fp_to_float_o:wwN 0
\}
\}
(End definition for \__fp_fact_loop_o:w.)
\}

{/package}
Chapter 75

l3fp-trig Implementation

Unary functions.

\tl_map_inline:nn
\{\acos \acsc \asec \asin \cos \cot \csc \sec \sin \tan\}
\}
\cs_new:cpx { __fp_parse_word_#1:N }
\exp_not:N \__fp_parse_unary_function:NNN
\exp_not:c { __fp_#1_o:w }
\exp_not:N \use_i:nn
\cs_new:cpx { __fp_parse_word_#1d:N }
\exp_not:N \__fp_parse_unary_function:NNN
\exp_not:c { __fp_#1_o:w }
\exp_not:N \use_ii:nn
\}

(End definition for \__fp_parse_word_acos:N and others.)

Those functions may receive a variable number of arguments.

\cs_new:Npn \__fp_parse_word_acot:N
\{ \__fp_parse_function:NNN \__fp_acot_o:Nw \use_i:nn \}
\cs_new:Npn \__fp_parse_word_acotd:N
\{ \__fp_parse_function:NNN \__fp_acot_o:Nw \use_ii:nn \}
\cs_new:Npn \__fp_parse_word_atan:N
\{ \__fp_parse_function:NNN \__fp_atan_o:Nw \use_i:nn \}
\cs_new:Npn \__fp_parse_word_atand:N
\{ \__fp_parse_function:NNN \__fp_atan_o:Nw \use_ii:nn \}

(End definition for \__fp_parse_word_acot:N and others.)
75.1 Direct trigonometric functions

The approach for all trigonometric functions (sine, cosine, tangent, cotangent, cosecant, and secant), with arguments given in radians or in degrees, is the same.

- Filter out special cases (±0, ±inf and NaN).
- Keep the sign for later, and work with the absolute value |x| of the argument.
- Small numbers (|x| < 1 in radians, |x| < 10 in degrees) are converted to fixed point numbers (and to radians if |x| is in degrees).
- For larger numbers, we need argument reduction. Subtract a multiple of π/2 (in degrees, 90) to bring the number to the range to [0, π/2) (in degrees, [0, 90)).
- Reduce further to [0, π/4] (in degrees, [0, 45]) using \( \sin x = \cos(\pi/2 - x) \), and when working in degrees, convert to radians.
- Use the appropriate power series depending on the octant \( \lfloor x \pi/4 \rfloor \mod 8 \) (in degrees, the same formula with \( \pi/4 \to 45 \)), the sign, and the function to compute.

75.1.1 Filtering special cases

\_fp_sin_o:w

This function, and its analogs for \( \cos, \csc, \sec, \tan \), and \( \cot \) instead of \( \sin \), are followed either by \texttt{use_i:nn} and a float in radians or by \texttt{use_ii:nn} and a float in degrees. The sine of ±0 or NaN is the same float. The sine of ±∞ raises an invalid operation exception with the appropriate function name. Otherwise, call the \texttt{trig} function to perform argument reduction and if necessary convert the reduced argument to radians. Then, \_fp_sin_series_o:NNwwww is called to compute the Taylor series: this function receives a sign \#3, an initial octant of 0, and the function \_fp_ep_to_float_o:wwN which converts the result of the series to a floating point directly rather than taking its inverse, since \( \sin(x) = \#3 \sin|x| \).

\_fp_cos_o:w

The cosine of ±0 is 1. The cosine of ±∞ raises an invalid operation exception. The cosine of NaN is itself. Otherwise, the \texttt{trig} function reduces the argument to at most half a right-angle and converts if necessary to radians. We then call the same series as
for sine, but using a positive sign 0 regardless of the sign of $x$, and with an initial octant of 2, because $\cos(x) = +\sin(\pi/2 + |x|)$.

\[
\text{\_\_fp_cos_o:} #1 \text{ \_\_fp_chk:w \#2\#3; @}
\]

(End definition for \_\_fp_cos_o:)

\_\_fp_csc_o: The cosecant of $\pm 0$ is $\pm \infty$ with the same sign, with a division by zero exception (see \_\_fp_cot_zero_o:Nfw defined below), which requires the function name. The cosecant of $\pm \infty$ raises an invalid operation exception. The cosecant of NaN is itself. Otherwise, the trig function performs the argument reduction, and converts if necessary to radians before calling the same series as for sine, using the sign \#3, a starting octant of 0, and inverting during the conversion from the fixed point sine to the floating point result, because $\csc(x) = \#3(\sin|\pi/2 + |x|)|^{-1}$.

\[
\text{\_\_fp_csc_o:} #1 \text{ \_\_fp_chk:w \#2\#3\#4; @}
\]

(End definition for \_\_fp_csc_o:)

\_\_fp_sec_o: The secant of $\pm 0$ is 1. The secant of $\pm \infty$ raises an invalid operation exception. The secant of NaN is itself. Otherwise, the trig function reduces the argument and turns it to radians before calling the same series as for sine, using a positive sign 0, a starting octant of 2, and inverting upon conversion, because $\sec(x) = +1/\sin(\pi/2 + |x|)$.

\[
\text{\_\_fp_sec_o:} #1 \text{ \_\_fp_chk:w \#2\#3; @}
\]

1126
\_fp\_tan\_o:w

The tangent of ±0 or NaN is the same floating point number. The tangent of ±\infty raises an invalid operation exception. Once more, the trig function does the argument reduction step and conversion to radians before calling \_fp\_tan\_series\_o:NNwwww, with a sign #3 and an initial octant of 1 (this shift is somewhat arbitrary). See \_fp\_\-cot\_o:w for an explanation of the 0 argument.

\cs\_new:Npn \_fp\_tan\_o:w #1 \_s\_fp \_s\_fp\_chk:w #2#3#4; @
\{ 
  \if_case:w #2 \exp_stop_f:
    \_fp\_case\_return\_same\_o:w
  \or: \_fp\_case\_use:nw
    \{ 
      \_fp\_trig:NNNNNwm #1 \_fp\_sin\_series\_o:NNwwww
      \_fp\_ep\_inv\_to\_float\_o:wwN 0 2
    \}
  \or: \_fp\_case\_use:nw
    { \_fp\_invalid\_operation\_o:fw \{ #1 \{ sec \} \{ secd \} \} }
  \else: \_fp\_case\_return\_same\_o:w
  \fi:
\s\_fp \_s\_fp\_chk:w #2 #3;
\}

(End definition for \_fp\_sec\_o:w.)

\_fp\_cot\_o:w
\_fp\_cot\_zero\_o:Nfw

The cotangent of ±0 is ±\infty with the same sign, with a division by zero exception (see \_fp\_cot\_zero\_o:Nfw. The cotangent of ±\infty raises an invalid operation exception. The cotangent of NaN is itself. We use cot x = -tan(\pi/2 + x), and the initial octant for the tangent was chosen to be 1, so the octant here starts at 3. The change in sign is obtained by feeding \_fp\_tan\_series\_o:NNwwww two signs rather than just the sign of the argument: the first of those indicates whether we compute tangent or cotangent. Those signs are eventually combined.

\cs\_new:Npn \_fp\_cot\_o:w #1 \_s\_fp \_s\_fp\_chk:w #2#3#4; @
\{ 
  \if_case:w #2 \exp_stop_f:
    \_fp\_cot\_zero\_o:Nfw #3 \{ #1 \{ cot \} \{ cotd \} \} 
  \or: \_fp\_case\_use:nw
    \{ 
      \_fp\_trig:NNNNNwm #1
      \_fp\_tan\_series\_o:NNwwww 2 \#3 \#3
    \}
  \or: \_fp\_case\_use:nw

1127
The first argument is \texttt{\use_i:nn} if the operand is in radians and \texttt{\use_ii:nn} if it is in degrees. Arguments \#2 to \#5 control what trigonometric function we compute, and \#6 to \#8 are pieces of a normal floating point number. Call the \texttt{\_series} function \#2, with arguments \#3, either a conversion function (\texttt{\_fp_to_float:o:Nw} or \texttt{\_fp_\-inv_to_float:o:Nw}) or a sign 0 or 2 when computing tangent or cotangent; \#4, a sign 0 or 2; the octant, computed in an integer expression starting with \#5 and stopped by a period; and a fixed point number obtained from the floating point number by argument reduction (if necessary) and conversion to radians (if necessary). Any argument reduction adjusts the octant accordingly by leaving a (positive) shift into its integer expression. Let us explain the integer comparison. Two of the four \texttt{\exp_after:wN} are expanded, the expansion hits the test, which is true if the float is at least 1 when working in radians, and at least 10 when working in degrees. Then one of the remaining \texttt{\exp_after:wN} hits \#1, which picks the \texttt{trig} or \texttt{trigd} function in whichever branch of the conditional was taken. The final \texttt{\exp_after:wN} closes the conditional. At the end of the day, a number is large if it is $\geq 1$ in radians or $\geq 10$ in degrees, and small otherwise. All four \texttt{trig/trigd} auxiliaries receive the operand as an extended-precision number.
75.1.3 Small arguments

\_fp\_trig\_small:ww

This receives a small extended-precision number in radians and converts it to a fixed point number. Some trailing digits may be lost in the conversion, so we keep the original floating point number around: when computing sine or tangent (or their inverses), the last step is to multiply by the floating point number (as an extended-precision number) rather than the fixed point number. The period serves to end the integer expression for the octant.

\_fp\_trigd\_small:ww

\cs_new:Npn \_fp\_trigd\_small:ww #1,#2;
{ \_fp\_ep\_mul\_raw:wwwwN -1,{1745}{3292}{5199}{4329}{5769}{2369}; #1,#2; \_fp\_trig\_small:ww }

(End definition for \_fp\_trigd\_small:ww.)

75.1.4 Argument reduction in degrees

\_fp\_trigd\_large:ww

\_fp\_trigd\_large\_auxi:nnnnwNNNN \_fp\_trigd\_large\_auxii:wNw \_fp\_trigd\_large\_auxiii:www

Note that 25 \times 360 = 9000, so 10^{k+1} \equiv 10^k \pmod{360} for k \geq 3. When the exponent \#1 is very large, we can thus safely replace it by 22 (or even 19). We turn the floating point number into a fixed point number with two blocks of 8 digits followed by five blocks of 4 digits. The original float is \times (block_1) \cdots (block_3), (block_4) \cdots (block_7), or is equal to it modulo 360 if the exponent \#1 is very large. The first auxiliary finds (block_1) + (block_2) (mod 9), a single digit, and prepends it to the 4 digits of (block_3). It also unpacks (block_4) and grabs the 4 digits of (block_7). The second auxiliary grabs the (block_3) plus any contribution from the first two blocks as \#1, the first digit of (block_4) (just after the decimal point in hundreds of degrees) as \#2, and the three other digits as \#3. It finds the quotient and remainder of \#1\#2 modulo 9, adds twice the quotient to the integer expression for the octant, and places the remainder (between 0 and 8) before \#3 to form a new \langle block_1 \rangle. The resulting fixed point number is x \in [0, 0.9]. If x \geq 0.45, we add 1 to the octant and feed 0.9 - x with an exponent of 2 to compensate the fact that we are working in units of hundreds of degrees rather than degrees to \_fp\_trigd\_small:ww. Otherwise, we feed it x with an exponent of 2. The third auxiliary also discards digits which were not packed into the various \langle blocks \rangle. Since the original exponent \#1 is at least 2, those are all 0 and no precision is lost (#6 and #7 are four 0 each).

\cs_new:Npn \_fp\_trigd\_large:ww #1, #2#3#4#5#6#7;
{ \exp_after:wN \_fp\_pack\_eight:wNNNNNNNN \exp_after:wN \_fp\_pack\_eight:wNNNNNNNN \exp_after:wN \_fp\_pack\_twice\_four:wNNNNNNNN \exp_after:wN \_fp\_pack\_twice\_four:wNNNNNNNN \exp_after:wN \_fp\_trig\_large\_auxi:mmmmwNNNN \exp_after:wN ; \exp:w \exp_end_continue_f:w
### 75.1.5 Argument reduction in radians

Arguments greater or equal to 1 need to be reduced to a range where we only need a few terms of the Taylor series. We reduce to the range \([0, 2\pi]\) by subtracting multiples of \(2\pi\), then to the smaller range \([0, \pi/2]\) by subtracting multiples of \(\pi/2\) (keeping track of how many times \(\pi/2\) is subtracted), then to \([0, \pi/4]\) by mapping \(x \rightarrow \pi/2 - x\) if appropriate. When the argument is very large, say, \(10^{100}\), an equally large multiple of \(2\pi\) must be subtracted, hence we must work with a very good approximation of \(2\pi\) in order to get a sensible remainder modulo \(2\pi\).

Specifically, we multiply the argument by an approximation of \(1/(2\pi)\) with 10048 digits, then discard the integer part of the result, keeping 52 digits of the fractional part. From the fractional part of \(x/(2\pi)\) we deduce the octant (quotient of the first three digits by 125). We then multiply by 8 or \(-8\) (the latter when the octant is odd), ignore any integer part (related to the octant), and convert the fractional part to an extended precision number, before multiplying by \(\pi/4\) to convert back to a value in radians in \([0, \pi/4]\).

It is possible to prove that given the precision of floating points and their range of exponents, the 52 digits may start at most with 24 zeros. The 5 last digits are affected by carries from computations which are not done, hence we are left with at least 52 - 24 - 5 = 23 significant digits, enough to round correctly up to \(0.6 \cdot \text{ulp}\) in all cases.
This integer array stores blocks of 8 decimals of $10^{-16}/(2\pi)$. Each entry is $10^8$ plus an 8 digit number storing 8 decimals. In total we store 10112 decimals of $10^{-16}/(2\pi)$. The number of decimals we really need is the maximum exponent plus the number of digits we later need, 52, plus 12 (4 − 1 groups of 4 digits). The memory footprint (1/2 byte per digit) is the same as an earlier method of storing the data as a control sequence name, but the major advantage is that we can unpack specific subsets of the digits without unpacking the 10112 decimals.
The exponent #1 is between 1 and 10000. We wish to look up decimals $10^{#1-16}/(2\pi)$ starting from the digit #1 + 1. Since they are stored in batches of 8, compute $[#1/8]$ and fetch blocks of 8 digits starting there. The numbering of items in \c__fp_trig_intarray starts at 1, so the block $[#1/8]+1$ contains the digit we want, at one of the eight positions. Each call to \int_value:w \__kernel_intarray_item:Nn expands the next, until being stopped by \__fp_trig_large_auxiii:w using \exp_stop_f:. Once all these blocks are unpacked, the \exp_stop_f: and 0 to 7 digits are removed by \use_none:n...n. Finally, \__fp_trig_large_auxii:w packs 64 digits (there are between 65 and 72 at this point) into groups of 4 and the auxv auxiliary is called.

\cs_new:Npn \__fp_trig_large:ww #1, #2#3#4#5#6;
\exp_after:wN \__fp_trig_large_auxi:w
\int_value:w \__fp_int_eval:w (#1 - 4) / 8 \exp_after:wN , \int_value:w #1 , ;
\cs:w
\prg_replicate:nn { #2 - #1 * 8 } { n }
\exp_after:wN \cs_end:w
\int_value:w \__kernel_intarray_item:Nn \c__fp_trig_intarray { \__fp_int_eval:w #1 + 1 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxii:w
\cs:w
\use_none:n \prg_replicate:nn { #2 - #1 * 8 } { n }
\exp_after:wN \cs_end:w
\int_value:w \__kernel_intarray_item:Nn \c__fp_trig_intarray { \__fp_int_eval:w #1 + 2 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxiii:w \int_value:w \__kernel_intarray_item:Nn \c__fp_trig_intarray { \__fp_int_eval:w #1 + 3 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxiii:w \int_value:w \__kernel_intarray_item:Nn \c__fp_trig_intarray { \__fp_int_eval:w #1 + 4 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxiii:w \int_value:w \__kernel_intarray_item:Nn \c__fp_trig_intarray { \__fp_int_eval:w #1 + 5 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxiii:w \int_value:w \__kernel_intarray_item:Nn \c__fp_trig_intarray { \__fp_int_eval:w #1 + 6 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxiii:w \int_value:w \__kernel_intarray_item:Nn \c__fp_trig_intarray { \__fp_int_eval:w #1 + 7 \scan_stop: }
First come the first 64 digits of the fractional part of $10^{#1-16}/(2\pi)$, arranged in 16 blocks of 4, and ending with a semicolon. Then a few more digits of the same fractional part, ending with a semicolon, then 4 blocks of 4 digits holding the significand of the original argument. Multiply the 16-digit significand with the 64-digit fractional part: the auxvi auxiliary receives the significand as #2#3#4#5 and 16 digits of the fractional part as #6#7#8#9, and computes one step of the usual ladder of pack functions we use for multiplication (see e.g., \__fp_fixed_mul:wwn), then discards one block of the fractional part to set things up for the next step of the ladder. We perform 13 such steps, replacing the last middle shift by the appropriate trailing shift, then discard the significand and remaining 3 blocks from the fractional part, as there are not enough digits to compute any more step in the ladder. The last semicolon closes the ladder, and we return control to the auxvii auxiliary.
The auxvii auxiliary is followed by 52 digits and a semicolon. We find the octant as the integer part of 8 times what follows, or equivalently as the integer part of \( \frac{#1#2#3}{125} \), and add it to the surrounding integer expression for the octant. We then compute 8 times the 52-digit number, with a minus sign if the octant is odd. Again, the last middle shift is converted to a trailing shift. Any integer part (including negative values which come up when the octant is odd) is discarded by \( \_\_\_fp_use_i_until_s:nw \). The resulting fractional part should then be converted to radians by multiplying by \( 2\pi / 8 \), but first, build an extended precision number by abusing \( \_\_\_fp_ep_to_ep_loop:N \) with the appropriate trailing markers. Finally, \( \_\_\_fp_trig_small:ww \) sets up the argument for the functions which compute the Taylor series.

\[
\text{\texttt{\_\_\_fp_trig_large_auxvii:ww}} \text{\_\_\_fp_trig_large_auxviii:ww} \text{\_\_\_fp_trig_large_auxix:Nw} \text{\_\_\_fp_trig_large_auxx:wNNNNN} \text{\_\_\_fp_trig_large_auxxi:w}}
\]
(End definition for \_\_fp_trig_large_auxvii:w and others.)

### 75.1.6 Computing the power series

Here we receive a conversion function \_\_fp_ep_to_float_o:wwN or \_\_fp_ep_inv_to_float_o:wwN, a \langle \text{sign} \rangle (0 or 2), a (non-negative) \langle \text{octant} \rangle delimited by a dot, a (fixed point) number delimited by a semicolon, and an extended-precision number. The auxiliary receives:

- the conversion function #1;
- the final sign, which depends on the octant #3 and the sign #2;
- the octant #3, which controls the series we use;
- the square #4 * #4 of the argument as a fixed point number, computed with \_\_fp_fixed_mul:wwn;
- the number itself as an extended-precision number.

If the octant is in \{1, 2, 5, 6, ...\}, we are near an extremum of the function and we use the series

\[
\cos(x) = 1 - x^2 \left( \frac{1}{2!} - x^2 \left( \frac{1}{4!} - x^2 \left( \cdots \right) \right) \right).
\]

Otherwise, the series

\[
\sin(x) = x \left( 1 - x^2 \left( \frac{1}{3!} - x^2 \left( \frac{1}{5!} - x^2 \left( \cdots \right) \right) \right) \right)
\]

is used. Finally, the extended-precision number is converted to a floating point number with the given sign, and \_\_fp_sanitze:Nw checks for overflow and underflow.
Contrarily to \_\_fp_sin_series_o:NNwwww which received a conversion auxiliary as #1, here, #1 is 0 for tangent and 2 for cotangent. Consider first the case of the tangent. The octant #3 starts at 1, which means that it is 1 or 2 for $|x| \in [0, \pi/2]$, it is 3 or 4 for $|x| \in [\pi/2, \pi]$, and so on: the intervals on which $\tan|x| \geq 0$ coincide with those for which $\lfloor (#3 + 1)/2 \rfloor$ is odd. We also have to take into account the original sign of $x$ to get the sign of the final result; it is straightforward to check that the first $\texttt{int_value:w}$ expansion produces 0 for a positive final result, and 2 otherwise. A similar story holds for $\cot(x)$.

The auxiliary receives the sign, the octant, the square of the (reduced) input, and the (reduced) input (an extended-precision number) as arguments. It then computes the numerator and denominator of

$$\tan(x) \approx \frac{x(1 - x^2(a_1 - x^2(a_2 - x^2(a_3 - x^2(a_4 - x^2a_5)))))}{1 - x^2(b_1 - x^2(b_2 - x^2(b_3 - x^2(b_4 - x^2b_5))))}.$$ 

The ratio is computed by \_\_fp_ep_div:wwwww, then converted to a floating point number. For octants #3 (really, quadrants) next to a pole of the functions, the fixed point
numerator and denominator are exchanged before computing the ratio. Note that this
\if_int_odd:w test relies on the fact that the octant is at least 1.

(End definition for \_\_fp_tan_series_o:NNwwww and \_\_fp_tan_series_aux_o:NNww.)

75.2 Inverse trigonometric functions

All inverse trigonometric functions (arcsine, arccosine, arctangent, arccotangent, arc-
cosecant, and arccosecant) are based on a function often denoted \texttt{atan2}. This func-

1140
tion is accessed directly by feeding two arguments to arctangent, and is defined by 
atan(y, x) = atan(y/x) for generic y and x. Its advantages over the conventional arctangent is that it takes values in \([-\pi, \pi]\) rather than \([-\pi/2, \pi/2]\), and that it is better behaved in boundary cases. Other inverse trigonometric functions are expressed in terms of atan as

\[
\begin{align*}
\text{acos } x &= \text{atan}(\sqrt{1-x^2}, x) \\
\text{asin } x &= \text{atan}(x, \sqrt{1-x^2}) \\
\text{asec } x &= \text{atan}(\sqrt{x^2-1}, 1) \\
\text{acsc } x &= \text{atan}(1, \sqrt{x^2-1}) \\
\text{atan } x &= \text{atan}(x, 1) \\
\text{acot } x &= \text{atan}(1, x).
\end{align*}
\]

Rather than introducing a new function, atan2, the arctangent function atan is overloaded: it can take one or two arguments. In the comments below, following many texts, we call the first argument \(y\) and the second \(x\), because \(\text{atan}(y, x) = \text{atan}(y/x)\) is the angular coordinate of the point \((x, y)\).

As for direct trigonometric functions, the first step in computing \(\text{atan}(y, x)\) is argument reduction. The sign of \(y\) gives that of the result. We distinguish eight regions where the point \((x, |y|)\) can lie, of angular size roughly \(\pi/8\), characterized by their "octant", between 0 and 7 included. In each region, we compute an arctangent as a Taylor series, then shift this arctangent by the appropriate multiple of \(\pi/4\) and sign to get the result. Here is a list of octants, and how we compute the arctangent (we assume \(y > 0\): otherwise replace \(y\) by \(-y\) below):

1. \(0 < y < 0.41421x\), then \(\frac{|y|}{x}\) is given by a nicely convergent Taylor series:
2. \(-0.41421x < y < x\), then \(\frac{|y|}{x} = \frac{\pi}{4} - \text{atan} \frac{x-|y|}{x+|y|};
3. \(0 < x < 0.41421|y|\), then \(\frac{|y|}{x} = \frac{\pi}{2} - \text{atan} \frac{x}{|y|};
4. \(-x < 0.41421|y|\), then \(\frac{|y|}{x} = \frac{\pi}{2} + \text{atan} \frac{x}{|y|};
5. \(-0.41421|y| < x < |y|\), then \(\frac{|y|}{x} = \frac{3\pi}{4} - \text{atan} \frac{x+|y|}{x-|y|};
6. \(-0.41421x < y < -x\), then \(\frac{|y|}{x} = \frac{3\pi}{4} + \text{atan} \frac{x-|y|}{x+|y|};
7. \(0 < y < -0.41421x\), then \(\frac{|y|}{x} = \pi - \text{atan} \frac{|y|}{x}.

In the following, we denote by \(z\) the ratio among \(\frac{x}{y}\), \(\frac{x+y}{x-y}\), \(\frac{x-y}{x+y}\) which appears in the right-hand side above.

### 75.2.1 Arctangent and arccotangent

The parsing step manipulates atan and acot like min and max, reading in an array of operands, but also leaves \use_i\nn or \use_i\ii\nn depending on whether the result

\[
\begin{align*}
\\text{\_fp_atan_o:Nw} \\
\\text{\_fp_acot_o:Nw} \\
\\text{\_fp_atan_default:w}
\end{align*}
\]
should be given in radians or in degrees. The helper \_fp_parse_function_one_two:nnw checks that the operand is one or two floating point numbers (not tuples) and leaves its second argument or its tail accordingly (its first argument is used for error messages). More precisely if we are given a single floating point number \_fp_atan_default:w places \c_one_fp (expanded) after it; otherwise \_fp_atan_default:w is omitted by \_fp_parse_function_one_two:nnw.

\cs_new:Npn \_fp_atan_o:Nw \#1 { \_fp_parse_function_one_two:nnw \{ \#1 \{ atan \} \{ atand \} \} \_fp_atan_default:w \_fp_atanii_o:Nww \#1 } \cs_new:Npn \_fp_acot_o:Nw \#1 { \_fp_parse_function_one_two:nnw \{ \#1 \{ acot \} \{ acotd \} \} \_fp_atan_default:w \_fp_acotii_o:Nww \#1 } \cs_new:Npx \_fp_atan_default:w #1#2#3 @ { #1 #2 #3 \c_one_fp @ } (End definition for \_fp_atan_o:Nw, \_fp_acot_o:Nw, and \_fp_atan_default:w.)

\_fp_atanii_o:Nww \_fp_acotii_o:Nww If either operand is nan, we return it. If both are normal, we call \_fp_atan_normal:o:NNNw:NNw. If both are zero or both infinity, we call \_fp_atan_inf:o:NNNw with argument 2, leading to a result among \{±π/4,±3π/4\} (in degrees, \{±45,±135\}). Otherwise, one is much bigger than the other, and we call \_fp_atan_inf:o:NNNw with either an argument of 4, leading to the values \{±π/2\} (in degrees, \{±90\}), or 0, leading to \{±0,±π\} (in degrees, \{±0,±180\}). Since acot(x,y) = atan(y,x), \_fp_acotii_o:ww simply reverses its two arguments.

\cs_new:Npn \_fp_atanii_o:Nww \#1 \s__fp \__fp_chk:w \#2#3#4; \s__fp \__fp_chk:w \#5 #6 @ { \if_meaning:w 3 \#2 \_fp_case_return_i_o:ww \fi: \if_meaning:w 3 \#5 \_fp_case_return_ii_o:ww \fi: \if_case:w \if_meaning:w \#2 \#5 \if_meaning:w 1 \#2 10 \else: 0 \fi: \else: \if_int_compare:w \#2 > \#5 \exp_stop_f: 1 \else: 2 \fi: \fi: \exp_stop_f: \if_stop_f: \_fp_case_return:nw \{ \_fp_atan_inf:o:NNNw \#1 \#3 \#2 \} \or: \_fp_case_return:nw \{ \_fp_atan_inf:o:NNNw \#1 \#3 \#4 \} \or: \_fp_case_return:nw \{ \_fp_atan_inf:o:NNNw \#1 \#3 \#0 \} \fi: \_fp_atan_normal:o:NNNw:NNw \#1 \s__fp \_fp_chk:w \#2#3#4; \s__fp \_fp_chk:w \#5 #6 \exp_stop_f: \exp_stop_f: \_fp_atanii_o:Nww \#1\#2; \#3; \{ \_fp_atanii_o:Nww \#1\#3; \#2; \} (End definition for \_fp_atanii_o:Nww and \_fp_acotii_o:Nww.)

1142
This auxiliary is called whenever one number is ±0 or ±∞ (and neither is NaN). Then the result only depends on the signs, and its value is a multiple of $\pi/4$. We use the same auxiliary as for normal numbers, \_\_fp_atan_combine_o:NwwwwN, with arguments the final sign #2; the octant #3; $\tan z/z = 1$ as a fixed point number; $z = 0$ as a fixed point number; and $z = 0$ as an extended-precision number. Given the values we provide, $\tan z$ is computed to be 0, and the result is $[3/2] \cdot \pi/4$ if the sign #5 of x is positive, and $[(7 - #3)/2] \cdot \pi/4$ for negative x, where the divisions are rounded up.

\_\_fp_atan_inf_o:NNNw

\begin{verbatim}
\cs_new:Npn \__fp_atan_inf_o:NNNw #1#2#3 \s__fp \__fp_chk:w #4#5#6; 
{ \exp_after:wN \__fp_atan_combine_o:NwwwwwN \exp_after:wN #2 \int_value:w \__fp_int_eval:w \if_meaning:w 2 #5 7 - \fi: \exp_after:wN ; \c__fp_one_fixed_tl {0000}{0000}{0000}{0000}{0000}{0000}; #1}
(End definition for \_\_fp_atan_inf_o:NNNw.)
\end{verbatim}

\_\_fp_atan_normal_o:NNnwNnw

\begin{verbatim}
\cs_new_protected:Npn \__fp_atan_normal_o:NNnwNnw #1 \s__fp \__fp_chk:w 1#2#3#4; \s__fp \__fp_chk:w 1#5#6#7; 
{ \__fp_atan_test_o:NwwNwwN #2 #3, #4{0000}{0000}; #5 #6, #7{0000}{0000}; #1}
(End definition for \_\_fp_atan_normal_o:NNnwNnw.)
\end{verbatim}

\_\_fp_atan_test_o:NwwNwwN

\begin{verbatim}
\cs_new:Npn \__fp_atan_test_o:NwwNwwN #1#2,#3; #4#5,#6; 
{ \exp_after:wN \__fp_atan_combine_o:NwwwwwN \exp_after:wN #1 \int_value:w \__fp_int_eval:w \if_meaning:w 2 #4 7 - \__fp_int_eval:w \fi: 1143

\end{verbatim}

\_\_fp_atan_div:wnwwnw

\begin{verbatim}
\cs_new:Npn \__fp_atan_div:wnwwnw 
{ \exp_after:wN \__fp_atan_combine_o:NwwwwwN \exp_after:wN \__fp_int_eval:w \if_meaning:w 2 \fi: }
\end{verbatim}

1143
This receives two positive numbers \(a\) and \(b\) (equal to \(|x|\) and \(|y|\) in some order), each as an exponent and 6 blocks of 4 digits, such that \(0 < a < b\). If \(0.41421b < a\), the two numbers are “near”, hence the point \((y,x)\) that we started with is closer to the diagonals \(|y| = |x|\) than to the axes \(xy = 0\). In that case, the octant is 1 (possibly combined with the \(7\) and \(3\) inserted earlier) and we wish to compute \(\tan b - a\). Otherwise, the octant is 0 (again, combined with earlier terms) and we wish to compute \(\tan b\). In any case, call \(\__fp_atan_auxi:ww\) followed by \(z\), as a comma-delimited exponent and a fixed point number.

\(\text{End definition for } \__fp_atan_test_o:NwwNwwN.\)

\(\text{End definition for } \__fp_atan_div:wnwwnw, \__fp_atan_near:wwwn, \text{ and } \__fp_atan_near_aux:wnn.\)

Convert \(z\) from a representation as an exponent and a fixed point number in \([0.01, 1)\) to a fixed point number only, then set up the call to \(\__fp_atan_Taylor_loop:www\), followed by the fixed point representation of \(z\) and the old representation.
We compute the series of \((\text{atan} z)/z\). A typical intermediate stage has 
\[ #1 = 2k - 1, \]
\[ #2 = \frac{1}{2k+3} - z^2 (\frac{1}{2k+3} - z^2 (\cdots - z^2 \frac{1}{20}))) , \]
and \[ #3 = z^2. \]
To go to the next step \(k \rightarrow k - 1\), we compute 
\[ \frac{1}{2k+1}, \]
then subtract from it \(z^2\) times \( #2\). The loop stops when \(k = 0\): then 
\[ #2 \text{ is } (\text{atan } z)/z, \]
and there is a need to clean up all the unnecessary data, end the integer 
expression computing the octant with a semicolon, and leave the result \( #2\) afterwards.

\begin{verbatim}
\cs_new:Npn \__fp_atan_Taylor_loop:www #1; #2; #3;
{ \if_int_compare:w #1 = - \c_one_int \__fp_atan_Taylor_break:w ! \fi: \exp_after:wN \__fp_fixed_div_int:wwN \c__fp_one_fixed_tl #1; \__fp_rrot:www \__fp_fixed_mul_sub_back:wwwn #2; #3; \exp_after:wN \__fp_atan_Taylor_loop:www \int_value:w \__fp_int_eval:w #1 - 2 ; #3; }
\cs_new:Npn \__fp_atan_Taylor_break:w \if: #1 \__fp_fixed_mul_sub_back:wwwn #2; #3 ! { \fi: ; #2 ; }
\end{verbatim}

(End definition for \_\_fp_atan_Taylor_loop:www and \_\_fp_atan_Taylor_break:w.)

This receives a \langle sign \rangle, an \langle octant \rangle, a fixed point value of \((\text{atan } z)/z\), a fixed point number 
\(z\), and another representation of \(z\), as an \langle exponent \rangle and the fixed point number 
\(10^{-\langle exponent \rangle} z\), followed by either \texttt{\use_i:nn} (when working in radians) or \texttt{\use_ii:nn} 
(when working in degrees). The function computes the floating point result 
\[ (\text{sign}) \left( \left( \frac{\langle octant \rangle}{2} \right) \frac{\pi}{4} + (-1)^{\langle octant \rangle} \frac{\text{atan } z}{z} \cdot z \right), \] 
multiplied by \(180/\pi\) if working in degrees, and using in any case the most appropriate 
representation of \(z\). The floating point result is passed to \_\_fp_sanitize:Nw, which 
checks for overflow or underflow. If the octant is 0, leave the exponent \#5 for \_\_fp_-
sanitize:Nw, and multiply \#3 = \frac{\text{atan } z}{z} with \#6, the adjusted \(z\). Otherwise, multiply 
\#3 = \frac{\text{atan } z}{z} with \#4 = z, then compute the appropriate multiple of \(\frac{\pi}{4}\) and add or subtract 
the product \#3 \cdot \#4. In both cases, convert to a floating point with \_\_fp_fixed_to-_float_o:wN.

\begin{verbatim}
\cs_new:Npn \_\_fp_atan_combine_o:NwwwwN \_\_fp_atan_combine_aux:ww
\{ \exp_after:wN \_\_fp_sanitize:Nw \exp_after:wN \_\_fp_fixed_to_float_o:wN \int_value:w \_\_fp_int_eval:w
\end{verbatim}

(End definition for \_\_fp_atan_Taylor_loop:www and \_\_fp_atan_Taylor_break:w.)

\[ (\text{sign}) \left( \left( \frac{\langle octant \rangle}{2} \right) \frac{\pi}{4} + (-1)^{\langle octant \rangle} \frac{\text{atan } z}{z} \cdot z \right), \] 
multiplied by \(180/\pi\) if working in degrees, and using in any case the most appropriate 
representation of \(z\). The floating point result is passed to \_\_fp_sanitize:Nw, which 
checks for overflow or underflow. If the octant is 0, leave the exponent \#5 for \_\_fp_-
sanitize:Nw, and multiply \#3 = \frac{\text{atan } z}{z} with \#6, the adjusted \(z\). Otherwise, multiply 
\#3 = \frac{\text{atan } z}{z} with \#4 = z, then compute the appropriate multiple of \(\frac{\pi}{4}\) and add or subtract 
the product \#3 \cdot \#4. In both cases, convert to a floating point with \_\_fp_fixed_to-_float_o:wN.
75.2.2 Arcsine and arccosine

\__fp_asin_o:w  Again, the first argument provided by l3fp-parse is \use_i:nn if we are to work in radians and \use_i:nn for degrees. Then comes a floating point number. The arcsine of ±0 or NaN is the same floating point number. The arcsine of ±∞ raises an invalid operation exception. Otherwise, call an auxiliary common with \__fp_acos_o:w, feeding it information about what function is being performed (for “invalid operation” exceptions).

(End definition for \__fp_atan_combine_o:NNNNNN and \__fp_atan_combine_aux:ww.)
The arccosine of ±0 is \( \pi/2 \) (in degrees, 90). The arccosine of ±\( \infty \) raises an invalid operation exception. The arccosine of NaN is itself. Otherwise, call an auxiliary common with \( \__fp_sin_o:w \), informing it that it was called by acos or acosd, and preparing to swap some arguments down the line.

\[
\text{\( \__fp_acos_o:w \)}
\]

If the exponent \( #5 \) is at most 0, the operand lies within \((-1, 1)\) and the operation is permitted: call \( \__fp_asin_auxi_o:NnNww \) with the appropriate arguments. If the number is exactly ±1 (the test works because we know that \( #5 \geq 1, #6#7 \geq 1000000, #8#9 \geq 0 \), with equality only for ±1), we also call \( \__fp_acos_o:w \). Otherwise, \( \__fp_use_i:ww \) gets rid of the asin auxiliary, and raises instead an invalid operation, because the operand is outside the domain of arcsine or arccosine.

\[
\text{\( \__fp_asin_normal_o:NfwNnnnnw \)}
\]

We compute \( x/\sqrt{1-x^2} \). This function is used by asin and acos, but also by acsc and asec after inverting the operand, thus it must manipulate extended-precision numbers. First evaluate \( 1-x^2 \) as \((1+x)(1-x)\): this behaves better near \( x = 1 \). We do the
addition/subtraction with fixed point numbers (they are not implemented for extended-
precision floats), but go back to extended-precision floats to multiply and compute the
inverse square root $1/\sqrt{1-x^2}$. Finally, multiply by the (positive) extended-precision float
$|x|$, and feed the (signed) result, and the number +1, as arguments to the arctangent
function. When computing the arccosecine, the arguments $x/\sqrt{1-x^2}$ and +1 are swapped
by \texttt{\_\_fp_reverse_args:Nww} in that case before \texttt{\_\_fp_atan_test_o:NwwNwwN} is
evaluated. Note that the arctangent function requires normalized arguments, hence the
need for \texttt{ep_to_ep} and continue after \texttt{ep_mul}.

\begin{verbatim}
cs_new:Npn \_\_fp_asin_auxi_o:NnNww #1#2#3#4,#5; 27638 \__fp_ep_to_fixed:wwn #4,#5; 27639 \__fp_asin_isqrt:wn 27640 \__fp_ep_mul:wwwwn #4,#5; 27641 \__fp_ep_to_ep:wwN 27642 \__fp_ep_to_fixed:wn { \__fp_atan_test_o:NwwNwwN #3 } 27643 \exp_after:wN \__fp_fixed_sub:wwn \c__fp_one_fixed_tl #1; 27644 \__fp_fixed_add_one:wN #1; 27645 \__fp_fixed_continue:wn \__fp_ep_mul:wwwwn 0, 0, 27646 \__fp_ep_isqrt:wwn 27647 \endverbatim

(End definition for \texttt{\_\_fp_asin_auxi_o:NnNww} and \texttt{\_\_fp_asin_isqrt:wn}.)

\subsection*{75.2.3 Arccosecant and arcsecant}

\texttt{\_\_fp_acsc_o:w} Cases are mostly labelled by \texttt{#2}, except when \texttt{#2} is \texttt{2}: then we use \texttt{#3#2}, which is \texttt{02 = 2}
when the number is \texttt{+\infty} and \texttt{22} when the number is \texttt{--\infty}. The arccosecant of \texttt{±0} raises
an invalid operation exception. The arcsecant of \texttt{±\infty} is \texttt{±0} with the same sign. The
arcsecant of \texttt{NaN} is itself. Otherwise, \texttt{\_\_fp_acsc_normal_o:NfwNnw} does some more
tests, keeping the function name (acsc or acscd) as an argument for invalid operation
exceptions.

\begin{verbatim}
cs_new:Npn \_\_fp_acsc_o:w #1 \s__fp \_\_fp_chk:w #2#3#4; @ 27872 \_\_fp_case_use:nw \_\_fp_invalid_operation_o:fw \{ #1 \{ acsc \} \{ acscd \} \} \or: \_\_fp_case_use:nw \{ \_\_fp_acsc_normal_o:NfwNnw \{ \_\_fp_acsc_normal_o:NfwNnw \} \{ \_\_fp_acsc_normal_o:NfwNnw \} \{ \_\_fp_acsc_normal_o:NfwNnw \} \} \or: \_\_fp_case_return_o:Nw \_\_fp_acsc_normal_o:NfwNnw \c_zero_fp \or: \_\_fp_case_return_o:Nw \c_minus_zero_fp \_\_fp_acsc_o:w \_\_fp_chk:w #2 #3 #4; 27880 \endverbatim

(End definition for \texttt{\_\_fp_acsc_o:w}.)
The arcsecant of ±0 raises an invalid operation exception. The arcsecant of ±∞ is π/2 (in degrees, 90). The arcsecant of NaN is itself. Otherwise, do some more tests, keeping the function name asec (or asecd) as an argument for invalid operation exceptions, and a \_\_fp_reverse_args:Nww following precisely that appearing in \_\_fp_acos_o:w.

\_\_fp_asec_o:w

\cs_new:Npn \_\_fp_asec_o:w #1 \_\_fp_chk:w #2 #3; @
{\if_case:w #2 \exp_stop_f:
\__fp_case_use:nw
{ \__fp_invalid_operation_o:fw { #1 { asec } { asecd } } }
\or:
\_\_fp_case_use:nw
\_\_fp_acsc_normal_o:NfwNnw #1 { #1 { asec } { asecd } }
\_\_fp_reverse_args:Nww
\or: \_\_fp_case_use:nw { \_\_fp_atan_inf_o:NNNw #1 0 4 }
\else: \_\_fp_case_return_same_o:w
\fi:
\s__fp \_\_fp_chk:w #2 #3; }

(End definition for \_\_fp_asec_o:w.)

\_\_fp_acsc_normal_o:NfwNnw

If the exponent is non-positive, the operand is less than 1 in absolute value, which is always an invalid operation: complain. Otherwise, compute the inverse of the operand, and feed it to \_\_fp_asin_auxi_o:Nww (with all the appropriate arguments). This computes what we want thanks to acsc(x) = asin(1/x) and asec(x) = acos(1/x).

\cs_new:Npn \_\_fp_acsc_normal_o:NfwNnw #1\_\_fp_chk:w #2 #3; @
{\int_compare:nNnTF {#2} < 1
\_\_fp_invalid_operation_o:fw {#2}
\_\_fp_chk:w 1#2 #3; }

(End definition for \_\_fp_acsc_normal_o:NfwNnw.)

//package
Chapter 76

13fp-convert implementation

76.1 Dealing with tuples

The first argument is for instance \_fp_to_tl_dispatch:w, which converts any floating point object to the appropriate representation. We loop through all items, putting ,~ between all of them and making sure to remove the leading ,~.

\cs_new:Npn \_fp_tuple_convert:Nw #1 \s__fp_tuple \_fp_tuple_chk:w #2 ;
\{\int_case:nnF { \_fp_array_count:n {#2} } { 0 } { ( ) } { 1 } { \_fp_tuple_convert_end:w \@ { #1 #2 , } }\}
\{ \_fp_tuple_convert_loop:nNw { } #1 #2 { ? \_fp_tuple_convert_end:w } ; \@ { \use_none:nn } \}
\cs_new:Npn \_fp_tuple_convert_end:w #1 \@ #2 { \exp_after:wN ( \exp:w \exp_end_continue_f:w #2 ) } (End definition for \_fp_tuple_convert:Nw, \_fp_tuple_convert_loop:nNw, and \_fp_tuple_convert_end:w.)

76.2 Trimming trailing zeros

If #1 ends with a 0, the loop auxiliary takes that zero as an end-delimiter for its first argument, and the second argument is the same loop auxiliary. Once the last trailing
zero is reached, the second argument is the dot auxiliary, which removes a trailing dot if any. We then clean-up with the end auxiliary, keeping only the number.

\begin{verbatim}
cs_new:Npn \_fp_trim_zeros:w #1 ;
    { \_fp_trim_zeros_loop:w #1 ; \_fp_trim_zeros_dot:w . ; \s__fp_stop }
cs_new:Npn \_fp_trim_zeros_loop:w #1 0; #2 { #2 #1 ; #2 }
cs_new:Npn \_fp_trim_zeros_dot:w .; { \_fp_trim_zeros_end:w #1 ; }
cs_new:Npn \_fp_trim_zeros_end:w #1 ; #2 \s__fp_stop { #1 }
\end{verbatim}

(End definition for \_fp_trim_zeros:w and others.)

### 76.3 Scientific notation

The three public functions evaluate their argument, then pass it to \_fp_to_scientific_dispatch:w.

\begin{verbatim}
cs_new:Npn \_fp_to_scientific:N #1
    { \exp_after:wN \_fp_to_scientific_dispatch:w #1 }
cs_generate_variant:Nn \_fp_to_scientific:N { c }
cs_new:Npn \_fp_to_scientific:n
    { \exp_after:wN \_fp_to_scientific_dispatch:w \exp:w \exp_end_continue_f:w \_fp_parse:n }
\end{verbatim}

(End definition for \_fp_to_scientific:N and \_fp_to_scientific:n. These functions are documented on page 241.)

We allow tuples.

\begin{verbatim}
cs_new:Npn \_fp_to_scientific_dispatch:w \_fp_to_scientific_recover:w
\_fp_tuple_to_scientific:w
    { \_fp_tuple_convert:Nw \_fp_to_scientific_dispatch:w }
\end{verbatim}

(End definition for \_fp_to_scientific_dispatch:w, \_fp_to_scientific_recover:w, and \_fp_tuple_to_scientific:w.)

Expressing an internal floating point number in scientific notation is quite easy: no rounding, and the format is very well defined. First cater for the sign: negative numbers \((#2 = 2)\) start with \(-\); we then only need to care about positive numbers and \(\text{nan}\). Then filter the special cases: \(\pm 0\) are represented as \(0\); infinities are converted to a number slightly larger than the largest after an “invalid_operation” exception; \(\text{nan}\) is represented as \(0\) after an “invalid_operation” exception. In the normal case, decrement the exponent
and unbrace the 4 brace groups, then in a second step grab the first digit (previously hidden in braces) to order the various parts correctly.

\begin{verbatim}
\cs_new:Npn \__fp_to_scientific:w \s__fp \__fp_chk:w #1#2
\{        \if_meaning:w 2 #2 \exp_after:wN - \exp:w \exp_end_continue_f:w \fi:
\if_case:w #1 \exp_stop_f:
\__fp_case_return:nw \c__fp_overflowing_fp
\or: \exp_after:wN \__fp_to_scientific_normal:wnnnn
\or:
\__fp_case_use:nw
\{ \__fp_invalid_operation:nw
\{ \fp_to_scientific:N \c__fp_overflowing_fp
\} \fp_to_scientific
\}
\or:
\__fp_case_use:nw
\{ \__fp_invalid_operation:nw
\{ \fp_to_scientific:N \c_zero_fp
\} \fp_to_scientific
\}
\fi:
\s__fp \__fp_chk:w #1 #2
\}
\cs_new:Npn \__fp_to_scientific_normal:wnnnn
\s__fp \__fp_chk:w 1 #1 #2 #3#4#5#6 ;
{ \exp_after:wN \__fp_to_scientific_normal:wNw
\exp_after:wN e \int_value:w \__fp_int_eval:w #2 - 1 ; #3 #4 #5 #6 ; }
\cs_new:Npn \__fp_to_scientific_normal:wNw #1 ; #2#3;
{ #2.#3 #1 }
\end{verbatim}

(End definition for \__fp_to_scientific:w, \__fp_to_scientific_normal:wnnnn, and \__fp_to_scientific_normal:wNw.)

76.4 Decimal representation

\texttt{\fp_to_decimal:N} All three public variants are based on the same \__fp_to_decimal_dispatch:w after evaluating their argument to an internal floating point.

\texttt{\fp_to_decimal:n}

(End definition for \fp_to_decimal:N and \fp_to_decimal:n. These functions are documented on page 240.)
We allow tuples.

The structure is similar to \_\_fp_to_scientific:w. Insert - for negative numbers. Zero gives 0, ±∞ and NaN yield an “invalid operation” exception; note that ±∞ produces a very large output, which we don’t expand now since it most likely won’t be needed. Normal numbers with an exponent in the range [1, 15] have that number of digits before the decimal separator: “decimate” them, and remove leading zeros with \int_value:w, then trim trailing zeros and dot. Normal numbers with an exponent 16 or larger have no decimal separator, we only need to add trailing zeros. When the exponent is non-positive, the result should be 0.(zeros)(digits), trimmed.
\int_compare:nNnTF {#2} < \c__fp_prec_int
  { \__fp_decimate:nNnnn { \c__fp_prec_int - #2 } \__fp_to_decimal_large:Nnnw
  }
  { \exp_after:wN \exp_after:wN \exp_after:wN \__fp_to_decimal_large:wnnn \prg_replicate:nn { #2 - \c__fp_prec_int } { 0 } ; }
\{#3} {#4} {#5} {#6}
\}
\exp_after:wN \exp_after:wN \__fp_to_decimal_huge:wnnnn
\prg_replicate:nn { #2 - \c__fp_prec_int } { 0 } ;
\}
\exp_after:wN \__fp_trim_zeros:w \int_value:w \if_int_compare:w #2 > \c_zero_int
  #2 \fi:\exp_stop_f: #3.#4 ;
\}
\cs_new:Npn \__fp_to_decimal_large:Nnnw #1#2#3#4; \cs_new:Npn \__fp_to_decimal_huge:wnnnn #1; #2#3#4#5 { #2#3#4#5 #1 }
(End definition for \__fp_to_decimal:w and others.)

76.5 Token list representation

These three public functions evaluate their argument, then pass it to \__fp_to_tl_dispatch:w.
\cs_new:Npn \fp_to_tl:N #1 { \exp_after:wN \__fp_to_tl_dispatch:w #1 }
\cs_generate_variant:Nn \fp_to_tl:N { c }
\cs_new:Npn \fp_to_tl:n { \exp_after:wN \__fp_to_tl_dispatch:w \exp:w \exp_end_continue_f:w \__fp_parse:n }
(End definition for \fp_to_tl:N and \fp_to_tl:n. These functions are documented on page 241.)

We allow tuples.
\cs_new:Npn \__fp_to_tl_dispatch:w #1
  { \__fp_change_func_type:NNN \__fp_to_tl:w \__fp_to_tl:recover:w \__fp_tuple_to_tl:w #1 #1 #2 ; }
\cs_new:Npn \__fp_to_tl:recover:w #1 #2 ;
\{ \__fp_error:nffn { unknown-type } { \tl_to_str:n { #2 ; } } { } { } }
1154
A structure similar to \_\_fp_to_scientific\_dispatch:w\ and \_\_fp_to_decimal\_\_dispatch:w, but without the “invalid operation” exception. First filter special cases. We express normal numbers in decimal notation if the exponent is in the range \([-2, 16]\), and otherwise use scientific notation.

\cs_new:Npn \_\_fp_to_tl:w \s__fp \_\_fp_chk:w #1#2
\{ 
\if_meaning:w 2 #2 \exp_after:wN - \exp:w \exp_end_continue_f:w \fi:
\if_case:w #1 \exp_stop_f:
\__fp_case_return:nw { 0 }
\or: \exp_after:wN \_\_fp_to_tl_normal:nnnnn
\or: \__fp_case_return:nw { inf }
\else: \__fp_case_return:nw { nan }
\fi:
\}
\cs_new:Npn \_\_fp_to_tl_normal:nnnnn #1
\{ 
\int_compare:nTF \{-2 <= #1 <= \c__fp_prec_int\} 
\{ \_\_fp_to_decimal_normal:wnnnn \}
\{ \_\_fp_to_tl_scientific:wnnnn \}
\s__fp \_\_fp_chk:w 1 0 {#1}
\}
\cs_new:Npn \_\_fp_to_tl_scientific:wnnnn \s__fp \_\_fp_chk:w 1 #1 #2 #3 #4 #5 #6 ;
\{ 
\exp_after:wN \_\_fp_to_tl_scientific:wnW
\exp_after:wN e
\int_value:w \_\_fp_int_eval:w #2 - 1
; #3 #4 #5 #6 ;
\}
\cs_new:Npn \_\_fp_to_tl_scientific:wnW #1 ; #2 #3 ;
\{ \_\_fp_trim_zeros:w #2.3 ; #1 \}
(End definition for \_\_fp_to_tl:w and others.)

## 76.6 Formatting

This is not implemented yet, as it is not yet clear what a correct interface would be, for this kind of structured conversion from a floating point (or other types of variables) to a string. Ideas welcome.

## 76.7 Convert to dimension or integer

All three public variants are based on the same \_\_fp_to_dim\_\_dispatch:w after evaluating their argument to an internal floating point. We only allow floating point numbers,
not tuples.

\cs_new:Npn \fp_to_dim:N \ #1
\{ \exp_after:wN \__fp_to_dim_dispatch:w \ #1 \}
\cs_generate_variant:Nn \fp_to_dim:N { c }
\cs_new:Npn \fp_to_dim:n
\{ \exp:w \exp_end_continue_f:w \__fp_parse:n \}
\cs_new:Npn \__fp_to_dim_dispatch:w \ #1#2 ;
\{ \__fp_change_func_type:NNN \ #1 \ __fp_to_dim:w \ __fp_to_dim_recover:w \ #1 \ #2 ; \}
\cs_new:Npn \__fp_to_dim_recover:w \ #1
\{ \__fp_invalid_operation:nnw \{ 0 pt \} \}
\cs_new:Npn \__fp_to_dim:w \ #1; { \__fp_to_decimal:w \ #1 ; pt }

(End definition for \fp_to_dim:N and others. These functions are documented on page 240.)

\fp_to_int:N
\fp_to_int:c
\fp_to_int:n
\__fp_to_int_dispatch:w
\__fp_to_int_recover:w
For the most part identical to \fp_to_dim:N but without pt, and where \__fp_to_int:w does more work. To convert to an integer, first round to 0 places (to the nearest integer), then express the result as a decimal number: the definition of \__fp_to_decimal_dispatch:w is such that there are no trailing dot nor zero.

\cs_new:Npn \fp_to_int:N \ #1 \{ \exp_after:wN \__fp_to_int_dispatch:w \ #1 \}
\cs_generate_variant:Nn \fp_to_int:N { c }
\cs_new:Npn \fp_to_int:n
\{ \exp:w \exp_end_continue_f:w \__fp_parse:n \}
\cs_new:Npn \__fp_to_int_dispatch:w \ #1#2 ;
\{ \__fp_change_func_type:NNN \ #1 \ __fp_to_int:w \ __fp_to_int_recover:w \ #1 \ #2 ; \}
\cs_new:Npn \__fp_to_int_recover:w \ #1
\{ \__fp_invalid_operation:nnw \{ 0 \} \}
\cs_new:Npn \__fp_to_int:w \ #1; { \__fp_to_decimal:w \ #1 ; pt }

(End definition for \fp_to_int:N and others. These functions are documented on page 240.)

\dim_to_fp:n
\__fp_from_dim_test:ww
\__fp_from_dim:wW
\__fp_from_dim:wW
\__fp_from_dim:wNNnnnn
\__fp_from_dim:wNNnnnWw

76.8 Convert from a dimension

The dimension expression (which can in fact be a glue expression) is evaluated, converted to a number (i.e., expressed in scaled points), then multiplied by $2^{-16} = 0.0000152587890625$ to give a value expressed in points. The auxiliary \__fp_mul_\npos_o:Nww expects the desired (final sign) and two floating point operands (of the form
\s_{\fp \ldots ;}\) as arguments. This set of functions is also used to convert dimension registers to floating points while parsing expressions: in this context there is an additional exponent, which is the first argument of \s_{\fp_{\text{from dim test}}:ww}, and is combined with the exponent \s_{-4} \text{ of } 2^{-16}. There is also a need to expand afterwards: this is performed by \s_{\fp_{\text{mul npos o}:Nw}}, and cancelled by \s_{\text{prg do nothing}}: here.

(End definition for \s_{\dim{to fp:n}} and others. This function is documented on page 213.)

### 76.9 Use and eval

\s_{\text{fp\_use:}}N

Those public functions are simple copies of the decimal conversions.

\s_{\text{fp\_use:}}c

\s_{\text{fp\_eval:n}}

(End definition for \s_{\text{fp\_use:}}N and \s_{\text{fp\_eval:n}}. These functions are documented on page 241.)
\fp_sign:n Trivial but useful. See the implementation of \fp_add:Nn for an explanation of why to use \__fp_parse:n, namely, for better error reporting.
\cs_new:Npn \fp_sign:n #1
\fp_to_decimal:n { sign \__fp_parse:n {#1} } 
(End definition for \fp_sign:n. This function is documented on page 240.)

\fp_abs:n Trivial but useful. See the implementation of \fp_add:Nn for an explanation of why to use \__fp_parse:n, namely, for better error reporting.
\cs_new:Npn \fp_abs:n #1
\fp_to_decimal:n { abs \__fp_parse:n {#1} } 
(End definition for \fp_abs:n. This function is documented on page 256.)

\fp_max:nn \fp_min:nn Similar to \fp_abs:n, for consistency with \int_max:nn, etc.
\cs_new:Npn \fp_max:nn #1#2
\fp_to_decimal:n { max ( \__fp_parse:n {#1} , \__fp_parse:n {#2} ) } 
\cs_new:Npn \fp_min:nn #1#2
\fp_to_decimal:n { min ( \__fp_parse:n {#1} , \__fp_parse:n {#2} ) } 
(End definition for \fp_max:nn and \fp_min:nn. These functions are documented on page 256.)

76.10 Convert an array of floating points to a comma list
\__fp_array_to_clist:n \__fp_array_to_clist_loop:Nw
Converts an array of floating point numbers to a comma-list. If speed here ends up irrelevant, we can simplify the code for the auxiliary to become
\cs_new:Npn \__fp_array_to_clist_loop:Nw #1#2;
{ 
 \use_none:n #1 
 \{ , ~ \} \fp_to_tl:n { #1 #2 ; } 
 \__fp_array_to_clist_loop:Nw 
 }

The \use_ii:nn function is expanded after \__fp_expand:n is done, and it removes ,~ from the start of the representation.
\exp_not:f { \_fp_to_tl_dispatch:w #1 #2 ; }
\_fp_array_to_clist_loop:Nw
\}

(End definition for \_fp_array_to_clist:n and \_fp_array_to_clist_loop:Nw.)

\langle/package\rangle
Chapter 77

l3fp-random Implementation

Those functions may receive a variable number of arguments. We won’t use the argument ?.

(End definition for \__fp_parse_word_rand:N and \__fp_parse_word_randint:N.)

77.1 Engine support

Most engines provide random numbers, but not all. We write the test twice simply in order to write the false branch first.

(End definition for \__fp_parse_word_rand:N and \__fp_parse_word_randint:N.)
The primitive random number generator (RNG) is provided as \texttt{\textbackslash \texttt{\texttt{\texttt{tex_uniformdeviate}}}}:D. Under the hood, it maintains an array of 55 28-bit numbers, updated with a linear recursion relation (similar to Fibonacci numbers) modulo \(2^{28}\). When \texttt{\texttt{\texttt{\texttt{tex_uniformdeviate}}}}:D \texttt{\texttt{\texttt{\texttt{\langle\texttt{integer}\rangle}}} is called (for brevity denote by \(N\) the \texttt{\texttt{\texttt{\texttt{\langle\texttt{integer}\rangle}}}}, the next 28-bit number is read from the array, scaled by \(N/2^{28}\), and rounded. To prevent 0 and \(N\) from appearing half as often as other numbers, they are both mapped to the result 0.

This process means that \texttt{\texttt{\texttt{\texttt{tex_uniformdeviate}}}}:D only gives a uniform distribution from 0 to \(N-1\) if \(N\) is a divisor of \(2^{28}\), so we will mostly call the RNG with such power of 2 arguments. If \(N\) does not divide \(2^{28}\), then the relative non-uniformity (difference between probabilities of getting different numbers) is about \(N/2^{28}\). This implies that detecting deviation from \(1/N\) of the probability of a fixed value \(x\) requires about \(2^{56}/N\) random trials. But collective patterns can reduce this to about \(2^{56}/N^2\). For instance with \(N = 3 \times 2^k\), the modulo 3 repartition of such random numbers is biased with a non-uniformity about \(2^k/2^{28}\) (which is much worse than the circa \(3/2^{28}\) non-uniformity from taking directly \(N = 3\)). This is detectable after about \(2^{56}/2^{2k} = 9 \cdot 2^{56}/N^2\) random numbers. For \(k = 15, N = 98304\), this means roughly \(2^{20}\) calls to the RNG (experimentally this takes at the very least 16 seconds on a 2 giga-hertz processor). While this bias is not quite problematic, it is uncomfortably close to being so, and it becomes worse as \(N\) is increased.

In our code, we shall thus combine several results from the RNG.

The RNG has three types of unexpected correlations. First, everything is linear modulo \(2^{28}\), hence the lowest \(k\) bits of the random numbers only depend on the lowest \(k\) bits of the seed (and of course the number of times the RNG was called since setting the seed). The recommended way to get a number from 0 to \(N-1\) is thus to scale the raw 28-bit integer, as the engine’s RNG does. We will go further and in fact typically we discard some of the lowest bits.

Second, suppose that we call the RNG with the same argument \(N\) to get a set of \(K\) integers in \([0,N-1]\) (throwing away repeats), and suppose that \(N > K^3\) and \(K > 55\). The recursion used to construct more 28-bit numbers from previous ones is linear: \(x_n = x_{n-55} - x_{n-24}\) or \(x_n = x_{n-55} - x_{n-24} + 2^{28}\). After rescaling and rounding we find that the result \(N_n \in [0,N-1]\) is among \(N_{n-55} - N_{n-24} + \{-1,0,1\}\) modulo \(N\) (a more detailed analysis shows that 0 appears with frequency close to \(3/4\)). The resulting set thus has more triplets \((a,b,c)\) than expected obeying \(a = b + c\) modulo \(N\). Namely it will have of order \((K - 55) \times 3/4\) such triplets, when one would expect \(K^3/(6N)\). This starts to be detectable around \(N = 2^{18} > 55^3\) (earlier if one keeps track of positions too, but this is more subtle than it looks because the array of 28-bit integers is read backwards by the engine). Hopefully the correlation is subtle enough to not affect realistic documents so we do not specifically mitigate against this. Since we typically use two calls to the RNG per \texttt{\texttt{\texttt{\texttt{\texttt{\texttt{\\\\texttt{int_rand:nn}}}}} we would need to investigate linear relations between the \(x_{2n}\) on the one hand and between the \(x_{2n+1}\) on the other hand. Such relations will have more complicated coefficients than \(\pm 1\), which alleviates the issue.

Third, consider successive batches of 165 calls to the RNG (with argument \(2^{28}\) or with argument 2 for instance), then most batches have more odd than even numbers. Note
that this does not mean that there are more odd than even numbers overall. Similar
issues are discussed in Knuth’s TAOCP volume 2 near exercise 3.3.2-31. We do not have
any mitigation strategy for this.
Ideally, our algorithm should be:

- Uniform. The result should be as uniform as possible assuming that the RNG’s
  underlying 28-bit integers are uniform.

- Uncorrelated. The result should not have detectable correlations between different
  seeds, similar to the lowest-bit ones mentioned earlier.

- Quick. The algorithm should be fast in \( \TeX \), so no “bit twiddling”, but “digit
  twiddling” is ok.

- Simple. The behaviour must be documentable precisely.

- Predictable. The number of calls to the RNG should be the same for any \( \texttt{int-}
  \texttt{rand:nn} \), because then the algorithm can be modified later without changing
  the result of other uses of the RNG.

- Robust. It should work even for \( \texttt{int\_rand:nn \{ - \texttt{c\_max\_int} \} \{ \texttt{c\_max\_int} \} } \)
  where the range is not representable as an integer. In fact, we also provide later a
  floating-point \( \texttt{randint} \) whose range can go all the way up to \( 2 \times 10^{16} - 1 \) possible
  values.

Some of these requirements conflict. For instance, uniformity cannot be achieved with a
fixed number of calls to the RNG.

Denote by \( \text{random}(N) \) one call to \( \texttt{tex\_uniform\_deviate:D} \) with argument \( N \), and
by \( \text{ediv}(p, q) \) the \( \varepsilon\)-\( \TeX \) rounding division giving \( [p/q + 1/2] \). Denote by \( \langle \text{min} \rangle, \langle \text{max} \rangle \)
and \( R = \langle \text{max} \rangle - \langle \text{min} \rangle + 1 \) the arguments of \( \texttt{int\_min:nn} \) and the number of possible
outcomes. Note that \( R \in [1, 2^{42} - 1] \) cannot necessarily be represented as an integer
(however, \( R - 2^{31} \) can). Our strategy is to get two 28-bit integers \( X \) and \( Y \) from the
RNG, split each into 14-bit integers, as \( X = X_1 \times 2^{14} + X_0 \) and \( Y = Y_1 \times 2^{14} + Y_0 \) then
return essentially \( \langle \text{min} \rangle + |R(X_1 \times 2^{-14} + Y_1 \times 2^{-28} + Y_0 \times 2^{-42} + X_0 \times 2^{-56}]|. \) For
small \( R \) the \( X_0 \) term has a tiny effect so we ignore it and we can compute \( R \times Y / 2^{28} \)
much more directly by \( \text{random}(R) \).

- If \( R \leq 2^{17} - 1 \) then return \( \text{ediv}(R \text{random}(2^{14}) + \text{random}(R) + 2^{13} ; 2^{14}) - 1 + \langle \text{min} \rangle \). The
  shifts by \( 2^{13} \) and \( -1 \) convert \( \varepsilon\)-\( \TeX \) division to truncated division. The bound
  on \( R \) ensures that the number obtained after the shift is less than \( \texttt{c\_max\_int} \). The
  non-uniformity is at most of order \( 2^{17} / 2^{42} = 2^{-25} \).

- Split \( R = R_2 \times 2^{28} + R_1 \times 2^{14} + R_0 \), where \( R_2 \in [0, 15] \). Compute
  \( \langle \text{min} \rangle + R_2 X_1 2^{14} + (R_2 Y_1 + R_1 X_1) + \text{ediv}(R_2 Y_0 + R_1 Y_1 + R_0 X_1 + \text{ediv}(R_2 X_0 + R_0 Y_1 + \text{ediv}((2^{14} R_1 + R_0)(2^{14} Y_0 + X_0), 2^{28}), 2^{14}), 2^{14}) \) then map a result of \( \langle \text{max} \rangle + 1 \) to \( \langle \text{min} \rangle \). Writing
  each \text{ediv} in terms of truncated division with a shift, and using \( [(p + [r/s])]/q] = \lfloor (ps + r)/(sq) \rfloor \), what we compute is equal to \( [(\text{exact}) + 2^{-29} + 2^{-15} + 2^{-1}] \) with
  \( \langle \text{exact} \rangle = \langle \text{min} \rangle + R \times 0.X_1 Y_1 Y_0 X_0 \). Given we map \( \langle \text{max} \rangle + 1 \) to \( \langle \text{min} \rangle \), the shift
  has no effect on uniformity. The non-uniformity is bounded by \( R / 2^{56} < 2^{-24} \). It
  may be possible to speed up the code by dropping tiny terms such as \( R_0 X_0 \), but
  the analysis of non-uniformity proves too difficult.

To avoid the overflow when the computation yields \( \langle \text{max} \rangle + 1 \) with \( \langle \text{max} \rangle = 2^{31} - 1 \)
(note that \( R \) is then arbitrary), we compute the result in two pieces. Compute
\( \langle \text{first} \rangle = \langle \text{min} \rangle + R_2 X_1^{14} \) if \( R_2 < 8 \) or \( \langle \text{min} \rangle + 8X_1^{14} + (R_2 - 8)X_1^{14} \) if \( R_2 \geq 8 \), the expressions being chosen to avoid overflow. Compute \( \langle \text{second} \rangle = R_0 Y_1 + R_1 X_1 + \text{e} \text{div} (\ldots) \), at most \( R_2 2^{14} + R_2 2^{14} + R_0 \leq 2^{28} + 15 \times 2^{14} - 1 \), not at risk of overflowing. We have \( \langle \text{first} \rangle + \langle \text{second} \rangle = \langle \text{max} \rangle + 1 = \langle \text{min} \rangle + R \) if and only if \( \langle \text{second} \rangle = R_1 2^{14} + R_0 + R_2 2^{14} \) and \( 2^{14} R_2 X_1 = 2^{28} R_2 - 2^{14} R_2 \) (namely \( R_2 = 0 \) or \( X_1 = 2^{14} - 1 \)). In that case, return \( \langle \text{min} \rangle \), otherwise return \( \langle \text{first} \rangle + \langle \text{second} \rangle \), which is safe because it is at most \( \langle \text{max} \rangle \). Note that the decision of what to return does not need \( \langle \text{first} \rangle \) explicitly so we don’t actually compute it, just put it in an integer expression in which \( \langle \text{second} \rangle \) is eventually added (or not).

- To get a floating point number in \([0, 1)\) just call the \( R = 10000 \leq 2^{17} - 1 \) procedure above to produce four blocks of four digits.

- To get an integer floating point number in a range (whose size can be up to \( 2 \times 10^{16} - 1 \)), work with fixed-point numbers: get six times four digits to build a fixed point number, multiply by \( R \) and add \( \langle \text{min} \rangle \). This requires some care because l3fp-extended only supports non-negative numbers.

\( \text{\c__kernel_randint_max_int} \) Constant equal to \( 2^{17} - 1 \), the maximal size of a range that \( \text{\int_range:nn} \) can do with its “simple” algorithm.

\( \text{\c__kernel_randint:n} \) Used in an integer expression, \( \text{\c__kernel_randint:n} \{ R \} \) gives a random number \( 1 + \frac{[(R \text{random}(2^{14}) + \text{random}(R)) / 2^{14}]}{2^{14}} \) that is in \([1, R] \). Previous code was computing \( \frac{p}{2^{14}} \) as \( \text{e} \text{div}(p - 2^{13}, 2^{14}) \) but that wrongly gives \(-1\) for \( p = 0 \).

\( \text{\c__fp_rand_myriads:n} \) Used as \( \text{\c__fp_rand_myriads:n} \{ XXX \} \) with one letter \( X \) (specifically) per block of four digit we want; it expands to \( X \) followed by the requested number of brace groups, each containing four (pseudo-random) digits. Digits are produced as a random number in \([10000, 19999]\) for the usual reason of preserving leading zeros.
77.2 Random floating point

\__fp_rand_o:Nw \__fp_rand_o:w

First we check that \texttt{random} was called without argument. Then get four blocks of four digits and convert that fixed point number to a floating point number (this correctly sets the exponent). This has a minor bug: if all of the random numbers are zero then the result is correctly 0 but it raises the underflow flag; it should not do that.

\begin{verbatim}
cs_new:Npn \__fp_rand_o:Nw ? #1 @
  { tl_if_empty:nTF {#1}
    { \exp_after:wN \__fp_rand_o:w 
      \exp:w \exp_end_continue_f:w 
      \__fp_rand_myriads:n { XXXX } { 0000 } { 0000 } ; 0 }
     }
    \msg_expandable_error:nnnnn
    { fp } { num-args } { rand() } { 0 } { 0 }
  \exp_after:wN \c_nan_fp }
\cs_new:Npn \__fp_rand_o:w ;
  { \exp_after:wN \__fp_sanitize:Nw \exp_after:wN 0 \int_value:w \__fp_int_eval:w \c_zero_int \__fp_fixed_to_float_o:wN }
\end{verbatim}

(End definition for \__fp_rand_o:Nw and \__fp_rand_o:w.)

77.3 Random integer


Enforce that there is one argument (then add first argument 1) or two arguments. Call \__fp_randint_badarg:w on each; this function inserts 1 \exp_stop_f: to end the \if_case:w statement if either the argument is not an integer or if its absolute value is \geq 10^{16}. Also bail out if \__fp_compare_back:ww yields 1, meaning that the bounds are not in the right order. Otherwise an auxiliary converts each argument times 10^{-16} (hence the shift in exponent) to a 24-digit fixed point number (see \texttt{l3fp-extended}). Then compute the number of choices, \langle \max \rangle + 1 - \langle \min \rangle. Create a random 24-digit fixed-point number with \__fp_randint_myriads:n, then use a fused multiply-add instruction to multiply the number of choices to that random number and add it to \langle \min \rangle. Then truncate to 16 digits (namely select the integer part of 10^{16} times the result) before converting back to a floating point number (\__fp_sanitize:Nw takes care of zero). To avoid issues with negative numbers, add 1 to all fixed point numbers (namely 10^{16} to the integers they represent), except of course when it is time to convert back to a float.

\begin{verbatim}
cs_new:Npn \__fp_randint_o:Nw ?
  { \__fp_parse_function_one_two:nnw
    { randint } 
    { \__fp_randint_default:w \__fp_randint_o:w }
  }
\cs_new:Npn \__fp_randint_default:w #1 { \exp_after:wN #1 \c_one_fp }
\end{verbatim}
\int_rand:nn \_fp_randint:ww Evaluate the argument and filter out the case where the lower bound #1 is more than the upper bound #2. Then determine whether the range is narrower than \_c\_kernel\_-randint_max_int; #2-#1 may overflow for very large positive #2 and negative #1. If the
range is narrow, call \texttt{\_\_kernel\_randint:n \langle \text{choices} \rangle} where \langle \text{choices} \rangle is the number of possible outcomes. If the range is wide, use somewhat slower code.

\begin{verbatim}
\cs_new:Npn \int_rand:nn #1#2
\{\int_eval:n
  \exp_after:wN \_\_fp\_randint:ww
  \int_value:w \int_eval:n \{#1\} \exp_after:wN ;
  \int_value:w \int_eval:n \{#2\};
\}
\cs_new:Npn \_\_fp\_randint:ww #1; #2;
\{\if_int_compare:w #1 > #2 \exp_stop_f:
  \msg_expandable_error:nnnn { kernel } { randint-backward-range } {#1} {#2}
\else:
  \if_int_compare:w \_\_fp\_int_eval:w #2
    \if_int_compare:w #1 > \c_zero_int
      \_\_kernel\_randint_max_int
      \_\_fp\_int_eval_end:
      \_\_kernel\_randint:n
      \{ \_\_fp\_int_eval:w #2 - #1 + 1 \_\_fp\_int_eval_end: \}
      - 1 + #1
    \else:
      \_\_kernel\_randint:nn \{#1\} \{#2\}
      \_\_kernel\_randint:n.nn \{#1\} \{#2\}
      \_\_kernel\_randint:nn \{#1\} \{#2\}
      \fi:
  \fi:
\fi:
\}
\end{verbatim}

(End definition for \texttt{\_\_kernel\_randint:nn} and \texttt{\_\_fp\_randint:ww}. This function is documented on page 164.)

\begin{verbatim}
\_\_kernel\_randint:nn
\_\_fp\_randint\_split\_o:Nw
\_\_fp\_randint\_split\_aux:w
\_\_fp\_randint\_wide\_aux:w
\_\_fp\_randint\_wide\_auxii:w
\end{verbatim}

Any \( n \in [-2^{31} + 1, 2^{31} - 1] \) is uniquely written as \( 2^{14} n_1 + n_2 \) with \( n_1 \in [-2^{17}, 2^{17} - 1] \) and \( n_2 \in [0, 2^{14} - 1] \). Calling \texttt{\_\_fp\_randint\_split\_o:Nw n} gives \( n_1 \), \( n_2 \); and expands the next token once. We do this for two random numbers and apply \texttt{\_\_fp\_randint\_split\_o:Nw} twice to fully decompose the range \( R \). One subtlety is that we compute \( R - 2^{31} = \langle \text{max} \rangle - \langle \text{min} \rangle - (2^{31} - 1) \in [-2^{31} + 1, 2^{31} - 1] \) rather than \( R \) to avoid overflow.

Then we have \texttt{\_\_fp\_randint\_wide\_aux:w \langle X_1 \rangle; \langle X_0 \rangle; \langle Y_1 \rangle; \langle Y_0 \rangle; \langle R_2 \rangle; \langle R_1 \rangle; \langle R_0 \rangle;}. and we apply the algorithm described earlier.

\begin{verbatim}
\cs_new:Npn \_\_kernel\_randint:nn #1#2
\{\exp_after:wN \_\_fp\_randint\_wide\_aux:w
  \int_value:w
  \exp_after:wN \_\_fp\_randint\_split\_o:Nw
  \tex_uniformdeviate:D 268435456 ;
  \int_value:w
  \exp_after:wN \_\_fp\_randint\_split\_o:Nw
\}
\end{verbatim}

1167
\tex_uniformdeviate:D 268435456 ;
\int_value:w
\exp_after:wN \__fp_randint_split_o:Nw
\int_value:w \__fp_int_eval:w 131072 +
\exp_after:wN \__fp_randint_split_o:Nw
\int_value:w
\__kernel_int_add:nnn {#2} { -#1 } { -\c_max_int } ;
\)
\cs_new:Npn \__fp_randint_split_o:Nw #1#2 ;
{\if_meaning:w 0 #1
0 \exp_after:wN ; \int_value:w 0
\else:
\exp_after:wN \__fp_randint_split_aux:w
\int_value:w \__fp_int_eval:w (#1#2 - 8192) / 16384 ;
+ #1#2
\fi:
\exp_after:wN ;
}
\cs_new:Npn \__fp_randint_split_aux:w #1 ;
{#1 \exp_after:wN ;
\int_value:w \__fp_int_eval:w - #1 * 16384
}
\cs_new:Npn \__fp_randint_wide_aux:w #1;#2; #3;#4; #5;#6;#7 ;
{\exp_after:wN \__fp_randint_wide_auxii:w
\int_value:w \__fp_int_eval:w #5 * #3 + #6 * #1 +
(#5 * #4 + #6 * #3 + #7 * #1 +
(#5 * #2 +
#7 * #3 +
(16384 * #6 + #7) * (16384 * #4 + #2) / 268435456) / 16384
) / 16384 \exp_after:wN ;
\int_value:w \__fp_int_eval:w (#5 + #6) * 16384 + #7 ;
#1 ; #5 ;
}
\cs_new:Npn \__fp_randint_wide_auxii:w #1; #2; #3; #4;
{\if_int_odd:w 0
\if_int_compare:w #1 = #2 \else: \exp_stop_f: \fi:
\if_int_compare:w #4 = \c_zero_int 1 \fi:
\if_int_compare:w #3 = 16383 - 1 \fi:
\exp_stop_f:
\exp_after:wN \prg_break:
\fi:
\if_int_compare:w #4 < 8 \exp_stop_f:
+ #4 * #3 * 16384
\else:
+ 8 * #3 * 16384 + (#4 - 8) * #3 * 16384
\fi:
+ #1
\prg_break_point:
}

(End definition for \_kernel_randint:nn and others.)
\texttt{\textbackslash int\_rand:n} Similar to \texttt{\textbackslash int\_rand:nn}, but needs fewer checks.

\begin{verbatim}
\cs_new:Npn \int_rand:n #1
\{
  \int_eval:n
  \{ \exp_args:Nf \__fp_randint:n \{ \int_eval:n \{#1\} \} \}
\}
\cs_new:Npn \__fp_randint:n #1
\{
  \if_int_compare:w #1 < \c_one_int
    \msg_expandable_error:nnnn
    { kernel } { randint-backward-range } { 1 } {#1}
  \__fp_randint:ww #1; 1;
  \else:
    \if_int_compare:w #1 > \c__kernel_randint_max_int
      \__kernel_randint:nn { 1 } {#1}
    \else:
      \__kernel_randint:n {#1}
    \fi:
  \fi:
\}
\end{verbatim}

(\textit{\textbf{End definition for} \texttt{\textbackslash int\_rand:n} \textbf{and} \texttt{\textbackslash __fp\_randint:n}. This function is documented on page 164.})

End the initial conditional that ensures these commands are only defined in engines that support random numbers.
Chapter 78

\l3fparray implementation

In analogy to \l3intarray it would make sense to have \<\*=fparray\>, but we need direct access to \__fp_parse:n from \l3fp-parse, and a few other (less crucial) internals of the \l3fp family.

78.1 Allocating arrays

There are somewhat more than \((2^{31} - 1)^2\) floating point numbers so we store each floating point number as three entries in integer arrays. To avoid having to multiply indices by three or to add 1 etc, a floating point array is just a token list consisting of three tokens: integer arrays of the same size.

\g__fp_array_int

Used to generate unique names for the three integer arrays.

\int_new:N \g__fp_array_int

(End definition for \g__fp_array_int.)

\l__fp_array_loop_int

Used to loop in \__fp_array_gzero:N.

\int_new:N \l__fp_array_loop_int

(End definition for \l__fp_array_loop_int.)

\fparray_new:Nn
\fparray_new:cn
\__fp_array_new:nNNN

Build a three-token token list, then define all three tokens to be integer arrays of the same size. No need to initialize the data: the integer arrays start with zeros, and three zeros denote precisely \c_zero_fp, as we want.

\cs_new_protected:Npn \fparray_new:Nn #1#2
\tl_new:N #1
\prg_replicate:nn { 3 }
\int_gincr:N \g__fp_array_int
\exp_args:NNc \tl_gput_right:Nn #1 { g__fp_array_ \__fp_int_to_roman:w \g__fp_array_int _intarray }
\exp_last_unbraced:Nfo \__fp_array_new:nNNNN
{ \int_eval:n {#2} } #1 #1

1170
\begin{verbatim}
28318  }
28319 \cs_generate_variant:Nn \fparray_new:Nn { c }
28320 \cs_new_protected:Npn \__fp_array_new:nNNN #1\#2\#3\#4\#5
28321 {
28322  \int_compare:nNnTF {#1} < 0
28323   {
28324    \msg_error:nnn { kernel } { negative-array-size } {#1}
28325    \cs_undefine:N #1
28326  }
28327  {
28328   \intarray_new:Nn #2 {#1}
28329   \intarray_new:Nn #3 {#1}
28330   \intarray_new:Nn #4 {#1}
28331  }
28332 }
\end{verbatim}

(End definition for \fparray_new:Nn and \__fp_array_new:nNNN. This function is documented on page 259.)

\texttt{\__fp_array_bounds:NNnTF} \texttt{\__fp_array_bounds_error:NNn}

See the \texttt{l3intarray} analogue: only names change. The functions \texttt{\fparray_gset:Nnn} and \texttt{\fparray_item:Nn} share bounds checking. The \texttt{T} branch is used if \#3 is within bounds of the array \#2.

\begin{verbatim}
28340 \cs_new:Npn \__fp_array_bounds:NNnTF #1#2#3#4#5
28341 {
28342   \if_int_compare:w 1 > #3 \exp_stop_f:
28343     \__fp_array_bounds_error:NNn #1 #2 (#3)
28344     #5
28345   \else:
28346     \if_int_compare:w #3 > \fparray_count:N #2 \exp_stop_f:
28347       \__fp_array_bounds_error:NNn #1 #2 (#3)
28348       #5
28349     \else:
28350       #4
28351     \fi:
28352     \fi:
28353 }
\end{verbatim}

\texttt{\__fp_array_bounds_error:NNn}

(End definition for \fparray_count:N. This function is documented on page 259.)

78.2 Array items

See the \texttt{l3intarray} analogue: only names change. The functions \texttt{\fparray_gset:Nnn} and \texttt{\fparray_item:Nn} share bounds checking. The \texttt{T} branch is used if \#3 is within bounds of the array \#2.

\begin{verbatim}
28354 \cs_new:Npn \__fp_array_bounds:NNnTF #1\#2\#3\#4\#5
28355 {
28356   \if_int_compare:w 1 > #3 \exp_stop_f:
28357     \__fp_array_bounds_error:NNn #1 #2 #3
28358   \else:
28359     #4
28360   \fi:
28361 \end{verbatim}

\texttt{\__fp_array_bounds_error:NNn}
\texttt{\__fp_array_gset:Nnn} \texttt{\__fp_array_gset:cppn} \texttt{\_fp_array_gset:NNNNww} \texttt{\_fp_array_gset_recover:Nw} \texttt{\_fp_array_gset_special:nnNNN} \texttt{\_fp_array_gset_normal:w}

Evaluate, then store exponent in one intarray, sign and 8 digits of mantissa in the next, and 8 trailing digits in the last.

\begin{verbatim}
\cs_new_protected:Npn \fparray_gset:Nnn #1#2#3
\exp_after:wN \exp_after:wN
\exp_after:wN \__fp_array_gset:NNNNww \exp_after:wN #1 \exp_after:wN #1
\int_value:w \int_eval:n {#2} \exp_after:wN ;
\exp:w \exp_end_continue_f:w \__fp_parse:n {#3}
}
\cs_generate_variant:Nn \fparray_gset:Nnn { c }
\cs_new_protected:Npn \__fp_array_gset:NNNNww #1#2#3#4#5 ; #6 ;
\__fp_array_bounds:NNnTF \msg_error:nnxxx #4 {#5}
\exp_after:wN \__fp_change_func_type:NNN
\__fp_use_i_until_s:nw #6 ;
\__fp_array_gset:w
\__fp_array_gset_recover:Nw
\int_value:w \int_eval:n {#2} \exp_after:wN ;
\exp:w \exp_end_continue_f:w \__fp_parse:n {#3}
}
\cs_new_protected:Npn \__fp_array_gset_recover:Nw #1#2 ;
\__fp_error:nffn { unknown-type } { \tl_to_str:n { #2 ; } } { } { }
\exp_after:wN #1 \c_nan_fp
\cs_new_protected:Npn \__fp_array_gset:w \s__fp \__fp_chk:w #1#2#3#4#5 ; #6#7#8#9
\__kernel_intarray_gset:Nnn #7 {#6} {#2}
\__kernel_intarray_gset:Nnn #8 {#6}
\__kernel_intarray_gset:Nnn #9 {#6}
\if_meaning:w 2 #1 3 \else: \fi: #3#4
\__kernel_intarray_gset:Nnn #10 {#6} #5
\cs_new_protected:Npn \__fparray_gset_special:nnNNN
\_fp_array_gset_special:nnNNN #1#2#3#4#5
\s__fp \__fp_chk:w 1 #1
\cs_new_protected:Npn \__fparray_gset_normal:w
\__kernel_intarray_gset:Nnn #7 {#6} {#2}
\__kernel_intarray_gset:Nnn #8 {#6}
\__kernel_intarray_gset:Nnn #9 {#6}
\__kernel_intarray_gset:Nnn #10 {#6} #5
\cs_new_protected:Npn \__fparray_gset_special:nnNNN
\_fp_array_gset_special:nnNNN #1#2#3#4#5
\end{verbatim}
End definition for \fparray_gset:Nnn and others. This function is documented on page 259.

\fparray_gzero:N
\fparray_gzero:c

\cs_new_protected:Npn \fparray_gzero:N #1
\int_zero:N \l__fp_array_loop_int
\prg_replicate:nn { \fparray_count:N #1 }
\int_incr:N \l__fp_array_loop_int
\exp_after:wN \__fp_array_gset_special:nnNNN
\exp_after:wN 0
\exp_after:wN \l__fp_array_loop_int
#1
\cs_generate_variant:Nn \fparray_gzero:N { c }

End definition for \fparray_gzero:N. This function is documented on page 259.

\fparray_item:Nn
\fparray_item:cn
\fparray_item_to_tl:Nn
\fparray_item_to_tl:cn
\__fp_array_item:NwN
\__fp_array_item:NNNnN
\__fp_array_item:N
\__fp_array_item:w
\__fp_array_item_special:w
\__fp_array_item_normal:w
\cs_new:Npn \fparray_item:Nn #1#2
\exp_after:wN \__fp_array_item:NwN\exp_after:wN #1\int_value:w \int_eval:n {#2} ;\__fp_to_decimal:w
\cs_generate_variant:Nn \fparray_item:Nn { c }
\cs_new:Npn \fparray_item_to_tl:Nn #1#2
\exp_after:wN \__fp_array_item:NwN\exp_after:wN #1\int_value:w \int_eval:n {#2} ;\__fp_to_tl:w
\cs_generate_variant:Nn \fparray_item_to_tl:Nn { c }
\cs_new:Npn \__fp_array_item:NwN #1#2 ;#3
\__fp_array_bounds:NNnTF \msg_expandable_error:nnfff #1 {#2}
{ \exp_after:wN \__fp_array_item:NNNnN #1 {#2} {#3} }
{ \exp_after:wN #3 \c_nan_fp }
\cs_new:Npn \__fp_array_item:NNNnN #1#2#3#4
{ \exp_after:wN \__fp_array_item:N\int_value:w \__kernel_intarray_item:Nn #2 {#4} \exp_after:wN ;\int_value:w \__kernel_intarray_item:Nn #3 {#4} \exp_after:wN ;\int_value:w \__kernel_intarray_item:Nn #1 {#4} ;
\cs_new:Npn \__fp_array_item:N #1
{ \if_meaning:w 0 #1 \exp_after:wN \__fp_array_item_special:w \fi:\__fp_array_item:w #1 }
\cs_new:Npn \_fp_array_item:w #1 #2#3#4#5 #6 ; 1 #7 ;
\exp_after:wN \_fp_array_item_normal:w
\int_value:w \if_meaning:w #1 1 0 \else: 2 \fi: \exp_stop_f:
#7 ; {#2#3#4#5} {#6} ;
\}
\cs_new:Npn \_fp_array_item_special:w #1 ; #2 ; #3 ; #4
\exp_after:wN #4
\exp:w \exp_end_continue_f:w
\if_case:w #3 \exp_stop_f:
\exp_after:wN \c_zero_fp
\or: \exp_after:wN \c_nan_fp
\or: \exp_after:wN \c_minus_zero_fp
\or: \exp_after:wN \c_inf_fp
\else: \exp_after:wN \c_minus_inf_fp
\fi:
\}
\cs_new:Npn \_fp_array_item_normal:w #1 #2#3#4#5 #6 ; #7 ; #8 ; #9
{ #9 \a__fp \_fp_chk:w 1 #1 {#8} #7 {#2#3#4#5} {#6} ; }

(End definition for \fparray_item:Nn and others. These functions are documented on page 259.)
Chapter 79

\textbf{13cctab implementation}

As \textsc{LuaTeX} offers engine support for category code tables, and this is entirely lacking from the other engines, we need two complementary approaches. (Some future \textsc{XeTeX} may add support, at which point the conditionals below would be different.)

\section{Variables}

\begin{itemize}
\item \texttt{\g__cctab_stack_seq} List of catcode tables saved by nested \texttt{\cctab\textunderscore begin:N}, to restore catcodes at the matching \texttt{\cctab\textunderscore end:}. When popped from the \texttt{\g__cctab_stack_seq} the table numbers are stored in \texttt{\g__cctab_unused_seq} for later reuse.
\item \texttt{\g__cctab_group_seq} A stack to store the group level when a catcode table started.
\item \texttt{\g__cctab_allocate_int} Integer to keep track of what category code table to allocate. In \textsc{LuaTeX} it is only used in format mode to implement \texttt{\cctab\textunderscore new:N}. In other engines it is used to make csnames for dynamic tables.
\item \texttt{\l__cctab\textunderscore internal\_a\_tl} Scratch space. For instance, when popping \texttt{\g__cctab_stack_seq/\g__cctab\textunderscore unused\_seq}, consists of the catcode table number (integer denotation) in \textsc{LuaTeX}, or of an intarray variable (as a single token) in other engines.
\end{itemize}

\section*{Appendix}

\begin{itemize}
\item \texttt{\l__cctab\textunderscore internal\_b\_tl}
\end{itemize}
In LuaTeX we store the endlinechar associated to each catcodetable in a property list, unless it is the default value 13.

```latex
\prop_new:N \g__cctab_endlinechar_prop
```

(End definition for \g__cctab_endlinechar_prop.)

## 79.2 Allocating category code tables

The \__cctab_new:N auxiliary allocates a new catcode table but does not attempt to set its value consistently across engines. It is used both in \cctab_new:N, which sets catcodes to \iniTeX values, and in \cctab_begin:N/\cctab_end: for dynamically allocated tables.

First, the LuaTeX case. Creating a new category code table is done like other registers. In ConTeXt, newcatcodetable does not include the initialisation, so that is added explicitly.

```latex
\sys_if_engine_luatex:TF
\cs_new_protected:Npn \cctab_new:N #1
{ \__kernel_chk_if_free_cs:N #1
  \__cctab_new:N #1
}
\cs_new_protected:Npn \__cctab_new:N #1
{ \newcatcodetable #1
  \tex_initcatcodetable:D #1
}
```

Now the case for other engines. Here, each table is an integer array. Following the LuaTeX pattern, a new table starts with ini\TeX\ codes. The index base is out-by-one, so we have an internal function to handle that. The ini\TeX\ \endlinechar is 13.

```latex
\sys_if_engine_luatex:TF
\cs_new_protected:Npn \cctab_new:N #1
{ \__kernel_chk_if_free_cs:N #1
  \__cctab_new:N #1
  \int_step_inline:nn { 256 }
  \__cctab_gstore:Nnn #1 {##1} { 12 }
  \__cctab_gstore:Nnn #1 { 257 } { 13 }
  \__cctab_gstore:Nnn #1 { 0 } { 9 }
  \__cctab_gstore:Nnn #1 { 13 } { 5 }
  \__cctab_gstore:Nnn #1 { 32 } { 10 }
  \__cctab_gstore:Nnn #1 { 37 } { 14 }
  \int_step_inline:nnn { 65 } { 90 }
  \__cctab_gstore:Nnn #1 {##1} { 11 }
  \__cctab_gstore:Nnn #1 { 92 } { 0 }
  \int_step_inline:nnn { 97 } { 122 }
  \__cctab_gstore:Nnn #1 {##1} { 11 }
  \__cctab_gstore:Nnn #1 { 127 } { 15 }
```

1176
79.3 Saving category code tables

In various functions we need to save the current catcodes (globally) in a table. In LuaTeX, saving the catcodes is a primitive, but the \endlinechar needs more work: to avoid filling \g__cctab_endlinechar_prop with many entries we special-case the default value 13. In other engines we store 256 current catcodes and the \endlinechar in an intarray variable.

\cs_new_protected:Npn \__cctab_gset:n #1
\exp_args:Nf \__cctab_gset_aux:n { \int_eval:n {#1} } }
\cs_new_protected:Npn \__cctab_gset_aux:n #1
\tex_savecatcodetable:D #1 \scan_stop:
\int_compare:nNnTF { \tex_endlinechar:D } = { 13 }
{ \prop_gremove:Nn \g__cctab_endlinechar_prop {#1} }
{ \prop_gput:NnV \g__cctab_endlinechar_prop {#1} \tex_endlinechar:D }
\sys_if_engine_luatex:TF
{ \cs_new_protected:Npn \__cctab_gset:n #1
{ \int_step_inline:nn { 256 }
{ \__kernel_intarray_gset:Nnn #1 {##1} { \char_value_catcode:n { ##1 - 1 } } }
\__kernel_intarray_gset:Nnn #1 { 257 } \tex_endlinechar:D }
\sys_if_engine_luatex:FN
\cs_new_protected:Npn \__cctab_gset:n #1
{ \__cctab_chk_if_valid:NT #1
{ \group_begin:
\cs_set_eq:NN \cctab_gset:cn \cctab_gset:Nn
\cctab_gset:Nn
Category code tables are always global, so only one version of assignments is needed. Simply run the setup in a group and save the result in a category code table \#1, provided it is valid. The internal function is defined above depending on the engine.

\cs_new_protected:Npn \cctab_gset:Nn #1#2
{ \__cctab_chk_if_valid:NT #1
{ \group_begin:
\cctab_select:N \c_initex_cctab
}
79.4 Using category code tables

In LuaTEX, we must ensure that the saved tables are read-only. This is done by applying the saved table, then switching immediately to a scratch table. Any later catcode assignment will affect that scratch table rather than the saved one. If we simply switched to the saved tables, then \char_set_catcode_other:N in the example below would change \c_document_cctab and a later use of that table would give the wrong category code to _.

\use:n
{
  \cctab_begin:N \c_document_cctab
  \char_set_catcode_other:N _
  \cctab_end:
  \cctab_begin:N \c_document_cctab
  \int_compare:nTF { \char_value_catcode:n { _ } = 8 } { \TRUE } { \ERROR }
  \cctab_end:
}

We must also make sure that a scratch table is never reused in a nested group: in the following example, the scratch table used by the first \cctab_begin:N would be changed globally by the second one issuing \savecatcodetable, and after \group_end: the wrong category codes (those of \c_str_cctab) would be imposed. Note that the inner \cctab_end: restores the correct catcodes only locally, so the problem really comes up because of the different grouping level. The simplest is to use a scratch table labeled by the \currentgrouplevel. We initialize one of them as an example.

\use:n
{
  \cctab_begin:N \c_document_cctab
  \group_begin:
   \cctab_begin:N \c_str_cctab
   \cctab_end:
  \group_end:
  \cctab_end:
}
\sys_if_engine_luatex:T
{
  \__cctab_new:N \g__cctab_internal_cctab
  \cs_new:Npn \__cctab_internal_cctab_name:
    {
}
The public function simply checks the \langle cctab var \rangle exists before using the engine-dependent \_cctab_select:N. Skipping these checks would result in low-level engine-dependent errors. First, the \texttt{LuaTeX} case. In other engines, selecting a catcode table is a matter of doing 256 catcode assignments and setting the \texttt{\texttt{endlinechar}}.

\begin{verbatim}
\cs_new_protected:Npn \cctab_select:N #1
 { \__cctab_chk_if_valid:NT #1 { \__cctab_select:N #1 } }
\cs_generate_variant:Nn \cctab_select:N { c }
\sys_if_engine_luatex:TF
 { \cs_new_protected:Npn \__cctab_select:N #1
   { \tex_catcodetable:D #1
     \prop_get:NVNTF \g__cctab_endlinechar_prop #1 \l__cctab_internal_a_tl
     { \int_set:Nn \tex_endlinechar:D { \l__cctab_internal_a_tl } }
     { \int_set:Nn \tex_endlinechar:D { 13 } }
     \cs_if_exist:cF { \__cctab_internal_cctab_name: }
     { \exp_args:Nc \__cctab_new:N { \__cctab_internal_cctab_name: } }
     \exp_args:Nc \tex_savecatcodetable:D { \__cctab_internal_cctab_name: }
     \exp_args:Nc \tex_catcodetable:D { \__cctab_internal_cctab_name: } }
   { \int_step_inline:nn { 256 }
     { \char_set_catcode:nn { ##1 - 1 } { \__kernel_intarray_item:Nn #1 {##1} } }
     \int_set:Nn \tex_endlinechar:D { \__kernel_intarray_item:Nn #1 { 257 } }
   }
 }
\end{verbatim}

(End definition for \cctab_select:N and \_cctab_select:N. This function is documented on page \pageref{ctab_select:N}.)

For \cctab_begin:N/\cctab_end: we will need to allocate dynamic tables. This is done here by \_cctab_begin_aux:, which puts a table number (in \texttt{LuaTeX}) or name (in other engines) into \l__cctab_internal_a_tl. In \texttt{LuaTeX} this simply calls \_cctab_new:N and uses the resulting catcodetable number; in other engines we need to give a name to the intarray variable and use that. In \texttt{LuaTeX}, to restore catcodes at \cctab_end: we cannot just set \texttt{\texttt{catcodetable}} to its value before \cctab_begin:N, because that table may have been altered by other code in the mean time. So we must make sure to save the catcodes in a table we control and restore them at \cctab_end:.

\begin{verbatim}
\sys_if_engine_luatex:TF
 \end{verbatim}

(End definition for \g__cctab_next_cctab and \_cctab_begin_aux:)
\begin{verbatim}
28604 \cs_new_protected:Npn \_cctab_begin_aux:
28605 \{ \_cctab_new:N \g__cctab_next_cctab
28606 \tl_set:NV \l__cctab_internal_a_tl \g__cctab_next_cctab
28607 \cs_undefine:N \g__cctab_next_cctab
28608 \}
28609 \}
28610 \}
28611 \cs_new_protected:Npn \_cctab_begin_aux:
28612 \{ \int_gincr:N \g__cctab_allocate_int
28613 \exp_args:Nc \__cctab_new:N { \g__cctab_ \int_use:N \g__cctab_allocate_int_cctab }
28614 \exp_args:NNc \tl_set:Nn \l__cctab_internal_a_tl { \g__cctab_ \int_use:N \g__cctab_allocate_int_cctab }
28615 \}
28616 \}
28617 (End definition for \g__cctab_next_cctab and \_cctab_begin_aux:.)
28618 \cctab_begin:N \cctab_begin:c
28619 Check the \langle cctab var \rangle exists, to avoid low-level errors. Get in \l__cctab_internal_a_tl the number/name of a dynamic table, either from \g__cctab_unused_seq where we save tables that are not currently in use, or from \_cctab_begin_aux: if none are available. Then save the current catcodes into the table (pointed to by) \l__cctab_internal_a_tl and save that table number in a stack before selecting the desired catcodes.
28620 \cs_new_protected:Npn \cctab_begin:N #1
28621 \{ \__cctab_chk_if_valid:NT #1
28622 \{ \seq_gpop:NNF \g__cctab_unused_seq \l__cctab_internal_a_tl
28623 \{ \_cctab_begin_aux: \}
28624 \exp_args:Nx \__cctab_chk_group_begin:n \{ \__cctab_nesting_number:N \l__cctab_internal_a_tl
28625 \seq_gpush:NV \g__cctab_stack_seq \l__cctab_internal_a_tl
28626 \exp_args:NV \__cctab_gset:n \l__cctab_internal_a_tl
28627 \__cctab_select:N #1
28628 \}
28629 \}
28630 \cs_generate_variant:Nn \cctab_begin:N { c }
28631 (End definition for \cctab_begin:N. This function is documented on page 260.)
28632 \cctab_end:
28633 Make sure a \cctab_begin:N was used some time earlier, get in \l__cctab_internal_a_tl the catcode table number/name in which the prevailing catcodes were stored, then restore these catcodes. The dynamic table is now unused hence stored in \g__cctab_unused_seq for recycling by later \cctab_begin:N.
28634 \cs_new_protected:Npn \cctab_end:
28635 \{ \seq_gpop:NNTF \g__cctab_unused_seq \l__cctab_internal_a_tl
28636 \{ \seq_gpush:NV \g__cctab_stack_seq \l__cctab_internal_a_tl
28637 \exp_args:Nx \__cctab_chk_group_end:n \{ \__cctab_nesting_number:N \l__cctab_internal_a_tl
28638 \}
28639 \}
28640 \}
28641 \cs_generate_variant:Nn \cctab_end: { c }
\end{verbatim}
\__cctab_select:N \l__cctab_internal_a_tl
\__cctabchk_group_begin:n \__cctabchk_group_end:n

Catcode tables are not allowed to be intermixed with groups, so here we check that they are properly nested regarding \TeX{} groups. \__cctabchk_group_begin:n stores the current group level in a stack, and locally defines a dummy control sequence \__cctabchk_group_{\langle cctab-level \rangle}chk::
\__cctabchk_group_end:n pops the stack, and compares the returned value with \tex_currentgrouplevel:D. If they differ, \cctab_end: is in a different grouping level than the matching \cctab_begin:N. If they are the same, both happened at the same level, however a group might have ended and another started between \cctab_begin:N and \cctab_end:.

\begin{verbatim}
\group_begin:
  \cctab_begin:N \c_document_cctab
\group_end:
\group_begin:
  \cctab_end:
\group_end:
\end{verbatim}

In this case checking \tex_currentgrouplevel:D is not enough, so we locally define \__cctabchk_group_{\langle cctab-level \rangle}chk::, and then check if it exist in \cctab_end:. If it doesn’t, we know there was a group end where it shouldn’t.

The \langle cctab-level \rangle in the sentinel macro above cannot be replaced by the more convenient \tex_currentgrouplevel:D because with the latter we might be tricked. Suppose:

\begin{verbatim}
\group_begin:
  \cctab_begin:N \c_code_cctab % A
\group_end:
\group_begin:
  \cctab_begin:N \c_code_cctab % B
  \cctab_end: % C
  \cctab_end: % D
\group_end:
\end{verbatim}

The line marked with A would start a \cctab with a sentinel token named \__cctabchk_group_1_chk:, which would disappear at the \group_end: that follows. But B would create the same sentinel token, since both are at the same group level. Line C would end the \cctab from line B correctly, but so would line D because line B created the same sentinel token. Using \langle cctab-level \rangle works correctly because it signals that certain \cctab level was activated somewhere, but if it doesn’t exist when the \cctab_end: is reached, we had a problem.

Unfortunately these tests only flag the wrong usage at the \cctab_end:, which might be far from the \cctab_begin:N. However it isn’t possible to signal the wrong usage at the \group_end: without using \tex_aftergroup:D, which is unsafe in certain types of groups.

The three cases checked here just raise an error, and no recovery is attempted: usually interleaving groups and catcode tables will work predictably.

(End definition for \cctab_end:. This function is documented on page 261.)
\cs_new_protected:Npn \__cctab_chk_group_begin:n #1
{
  \seq_gpush:Nx \g__cctab_group_seq { \int_use:N \tex_currentgrouplevel:D }
  \cs_set_eq:cN { __cctab_group_ #1 _chk: } \prg_do_nothing:
}
\cs_new_protected:Npn \__cctab_chk_group_end:n #1
{
  \seq_gpop:NN \g__cctab_group_seq \l__cctab_internal_b_tl
  \bool_lazy_and:nnF
    { \int_compare_p:nNn { \tex_currentgrouplevel:D } = { \l__cctab_internal_b_tl } }
    { \cs_if_exist_p:c { __cctab_group_ #1 _chk: } }
  { \msg_error:nnx { cctab } { group-mismatch } { \int_sign:n { \tex_currentgrouplevel:D - \l__cctab_internal_b_tl } } }
  \cs_undefine:c { __cctab_group_ #1 _chk: }
} (End definition for \__cctab_chk_group_begin:n and \__cctab_chk_group_end:n.)
\__cctab_nesting_number:N
\__cctab_nesting_number:w
This macro returns the numeric index of the current catcode table. In Lua\TeX{} this is just
the argument, which is a count reference to a \texttt{catcodetable} register. In other engines,
the number is extracted from the \texttt{cctab} variable.
\sys_if_engine_luatex:TF
  { \cs_new:Npn \__cctab_nesting_number:N #1 {#1} }
  { \cs_new:Npn \__cctab_nesting_number:N #1
    { \exp_after:wN \exp_after:wN \exp_after:wN \__cctab_nesting_number:w
      \exp_after:wN \token_to_str:N #1 }
  } \use:x
{ \cs_new:Npn \exp_not:N \__cctab_nesting_number:w
  ##1 \tl_to_str:n { g__cctab_ } ##2 \tl_to_str:n { \cctab } {##2} }
(End definition for \__cctab_nesting_number:N and \__cctab_nesting_number:w.)
Finally, install some code at the end of the \TeX{} run to check that all \texttt{cctab_begin:N}
were ended by some \texttt{cctab_end:}:
\cs_if_exist:NT \hook_gput_code:nnn
  { \hook_gput_code:nnn { enddocument/end } { cctab } }
  { \seq_if_empty:NT \g__cctab_stack_seq
    { \msg_error:nn { cctab } { missing-end } } }

1182
\cctab_item:Nn
\cctab_item:cn

Evaluate the integer argument only once. In most engines the \cctab variable only has 256 entries so we only look up the catcode for these entries, otherwise we use the current catcode. In particular, for out-of-range values we use whatever fall-back \char_value_\catcode:n. In \LaTeX, we use the \texttt{tex.getcatcode} function.

\begin{verbatim}
\cs_new:Npn \cctab_item:Nn \l_{\_\_cctab_item:nn} \int_eval:n \{ \#2 \} \#1 \\
\sys_if_engine_luatex:TF \\
\cs_new:Npn \l_{\_\_cctab_item:nn} \int_eval:n \{ \#2 \} \#1 \\
\sys_if_engine_luatex:TF \\
\cs_new:Npn \l_{\_\_cctab_item:nn} \intarray_item:Nn \#1 + 1 \#2 \\
\char_value_catcode:n \#1 \\
\end{verbatim}

(End definition for \cctab_item:Nn. This function is documented on page 261.)

\subsection*{79.5 Category code table conditionals}

\cctab_if_exist_p:N
\cctab_if_exist_p:c
\cctab_if_exist:NTF
\cctab_if_exist:ctF

Checks whether a \langle\texttt{cctab var}\rangle is defined.

\begin{verbatim}
\prg_new_eq_conditional:Nn \cs_if_exist:N \\
\prg_new_eq_conditional:Nn \cs_if_exist:c \\
\prg_new_protected_conditional:Npnn \\_\_cctab_chk_if_valid:N \#1 \\
\{ \#1, \#1 \}
\end{verbatim}

(End definition for \cctab_if_exist:NTF. This function is documented on page 261.)

\_\_cctab_chk_if_valid:NTF
\_\_cctab_chk_if_valid_aux:NTF

Checks whether the argument is defined and whether it is a valid \langle\texttt{cctab var}\rangle. In \LaTeX\ the validity of the \langle\texttt{cctab var}\rangle is checked by the engine, which complains if the argument is not a \texttt{chardef}ed constant. In other engines, check if the given command is an intarray variable (the underlying definition is a copy of the \texttt{cmr10} font).

\begin{verbatim}
\prg_new_protected_conditional:Nn \l_{\_\_cctab_chk_if_valid:N} \\_\_cctab_chk_if_valid:N \#1 \\
\{ \#1 \}
\end{verbatim}

1183
\begin{verbatim}
{ \msg_error:nxx { kernel } { command-not-defined }
  { \token_to_str:N #1 }
  \prg_return_false:
}
\sys_if_engine_luatex:TF
{
  \cs_new_protected:Npn \__cctab_chk_if_valid_aux:NTF #1
  {
    \int_compare:nNnTF {#1-1} < { \alloc@ccodetable@count }
  }
  \cs_if_exist:NT \c_syst_catcodes_n
  {
    \cs_gset_protected:Npn \__cctab_chk_if_valid_aux:NTF #1
    {
      \int_compare:nTF { #1 <= \c_syst_catcodes_n }
    }
  }
  {
    \cs_new_protected:Npn \__cctab_chk_if_valid_aux:NTF #1
    {
      \exp_args:Nf \str_if_in:nnTF
      { \cs_meaning:N #1 }
      { select~font~cmr10~at~ }
    }
  }
}\end{verbatim}

(End definition for \_\_cctab_chk_if_valid:NTF and \_\_cctab_chk_if_valid_aux:NTF.)

\subsection{Constant category code tables}

\begin{itemize}
\item \cctab_const:Nn
\item \cctab_const:cn
\end{itemize}

Creates a new \texttt{\langle cctab var \rangle} then sets it with the current and user-supplied codes.

\begin{verbatim}
\cs_new_protected:Npn \_\_cctab_chk_if_valid:NTF #1
{
  \int_compare:nNnTF \str_if_in:nTF
  { \cs_meaning:N #1 }
  { select-font-cmr10-at- }
}
\end{verbatim}

(End definition for \texttt{\_\_cctab_chk_if_valid:NTF}. This function is documented on page 260.)

\begin{itemize}
\item \cinitex_cctab
\item \c_other_cctab
\item \c_str_cctab
\end{itemize}

Creating category code tables means thinking starting from \textsc{in}\TeX. For all-other and the standard “string” tables that’s easy.

\begin{verbatim}
\cctab_new:N \cinitex_cctab
\cctab_const:Nn \c_other_cctab
{
  \cctab_select:N \cinitex_cctab
  \int_set:Nn \tex_endlinechar:D { -1 }
  \int_step_inline:nnn { 0 } { 127 }
  \char_set_catcode_other:n {#1 }
}
\end{verbatim}

(End definition for \texttt{\_\_cctab_chk_if_valid:NTF} and \texttt{\_\_cctab_chk_if_valid_aux:NTF}.)

79.6 Constant category code tables
To pick up document-level category codes, we need to delay setup to the end of the format, where that’s possible. Also, as there are a lot of category codes to set, we avoid using the official interface and store the document codes using internal code. Depending on whether we are in the hook or not, the catcodes may be code or document, so we explicitly set up both correctly:

\texttt{\cs_if_exist:NTF \@expl@finalise@setup@@}{\tl_gput_right:Nn \@expl@finalise@setup@@} \texttt{\use:n}{\__cctab_new:N \c_code_cctab\group_begin:\int_set:Nn \tex_endlinechar:D { 32 }\bool_lazy_or:nnTF \{ \sys_if_engine_xetex_p: \} \{ \sys_if_engine_luatex_p: \}{\int_step_function:nN \int_set:Nn \char_set_catcode_active:n} \int_step_function:nnN \int_set:Nn \char_set_catcode_letter:n \int_step_function:nnN \int_set:Nn \char_set_catcode_letter:n \char_set_catcode_ignore:n { 9 } % tab\char_set_catcode_other:n { 10 } % if\char_set_catcode_active:n { 12 } % ff\char_set_catcode_end_line:n { 13 } % cr\char_set_catcode_ignore:n { 32 } % space\char_set_catcode_other:n { 33 } % hash\char_set_catcode_math_toggle:n { 36 } % dollar\char_set_catcode_comment:n { 37 } % percent\char_set_catcode_alignment:n { 38 } % ampersand\char_set_catcode_letter:n { 58 } % colon\char_set_catcode_escape:n { 92 } % backslash\char_set_catcode_math_superscript:n { 94 } % circumflex\char_set_catcode_letter:n { 95 } % underscore\char_set_catcode_group_begin:n { 123 } % left brace\char_set_catcode_other:n { 124 } % pipe\char_set_catcode_group_end:n { 125 } % right brace\char_set_catcode_space:n { 126 } % tild\char_set_catcode_invalid:n { 127 } % ^\char_set_catcode_invalid:n { 128 } % \texttt{\bool_lazy_or:nnF}{\int_step_function:nN \int_set:Nn \char_set_catcode_active:n}{\__cctab_gset:n \c_code_cctab}\group_end:\cctab_const:Nn \c_document_cctab}
\begin{verbatim}
{ \cctab_select:N \c_code_cctab \\
\int_set:Nn \tex_endlinechar:D { 13 } \\
\char_set_catcode_space:n { 9 } \\
\char_set_catcode_space:n { 32 } \\
\char_set_catcode_other:n { 58 } \\
\char_set_catcode_math_subscript:n { 95 } \\
\char_set_catcode_active:n { 126 } 
}
\end{verbatim}

(End definition for \c_code_cctab and \c_document_cctab. These variables are documented on page 261.)

79.7 Messages

\begin{verbatim}
\msg_new:nnnn { cctab } { stack-full } { The-category-code-table-stack-is-exhausted. }

\msg_new:nnnn { cctab } { extra-end } { Extra-\iow_char:N\cctab_end: ignored-\msg_line_context:. }

\msg_new:nnnn { cctab } { missing-end } { Missing-\iow_char:N\cctab_end: before-end-of-TeX-run. }

\msg_new:nnnn { cctab } { invalid-cctab } { Invalid-\iow_char:N\catcode-table. }

\msg_new:nnnn { cctab } { group-mismatch } { Catcode-tables-and-groups-must-be-properly-nested,-but-
\end{verbatim}

1186
you-tried-to-interleave-them.-LaTeX-will-try-to-proceed,-
but-results-may-be-unexpected.

\prop_gput:Nnn \g_msg_module_name_prop { cctab } { LaTeX3 }
\prop_gput:Nnn \g_msg_module_type_prop { cctab } { }

(/package)
Chapter 80

\texttt{l3unicode} implementation

Case changing both for strings and “text” requires data from the Unicode Consortium. Some of this is built into the format (as \texttt{"lccode} and \texttt{"uccode} values) but this covers only the simple one-to-one situations and does not fully handle for example case folding.

As only the data needs to remain at the end of this process, everything is set up inside a group. The only thing that is outside is creating a stream: they are global anyway and it is best to force a stream for all engines. For performance reasons, some of the code here is very low-level: the material is read during loading expl3 in package mode.

\begin{verbatim}
\ior_new:N \g__char_data_ior
\bool_lazy_or:nnTF { \sys_if_engine_luatex_p: } { \sys_if_engine_xetex_p: } {
\group_begin:
Access the primitive but suppress further expansion: active chars are otherwise an issue.
\cs_set:Npn \__char_generate_char:n #1 { \tex_detokenize:D \tex_expandafter:D { \tex_Uchar:D " #1 } }
A fast local implementation for generating characters; the chars may be active, so we prevent further expansion.
\cs_set:Npx \__char_generate:n #1 {
    \exp_not:N \tex_unexpanded:D \exp_not:N \exp_after:wN
    \exp_not:N \tex_Ucharcat:D
    \exp_not:N \tex_catcode:D #1 ~
    \text_catcode:D #1 -
}
\ior_open:Nn \g__char_data_ior { UnicodeData.txt }
\cs_set_protected:Npn \__char_data_auxi:w
\end{verbatim}

Parse the main Unicode data file for two things. First, we want the titlecase exceptions: the one-to-one lower- and uppercase mappings it contains are all be covered by the \TeX data. Second, we need normalization data: at present, just the canonical NFD mappings. Those all yield either one or two codepoints, so the split is relatively easy.

\begin{verbatim}
\ior_open:Nn \g__char_data_ior { UnicodeData.txt }
\cs_set_protected:Npn \__char_data_auxi:w
\end{verbatim}
The other data files all use C-style comments so we have to worry about # tokens (and reading as strings). The set up for case folding is in two parts. For the basic (core) mappings, folding is the same as lower casing in most positions so only store the differences. For the more complex foldings, always store the result, splitting up the two or three code points in the input as required.
For upper- and lowercasing special situations, there is a bit more to do as we also have title casing to consider, plus we need to stop part-way through the file.

\ior_open:Nn \g__char_data_iior \ SpecialCasing.txt
\ior_set_protected:Npn \__char_data_auxii:w #1 \_#2 \_#3 \_#4 \#5 \q_stop
\use:n \__char_data_auxii:w #1 - #2 - #3 - #4 \q_stop
\use:n \__char_data_auxii:w #1 - #2 - #3 - #4 \q_stop
\str_if_eq:nnF {#3} {#4}
{ \use:n \__char_data_auxii:w #1 - title - #3 - \q_stop }
For the 8-bit engines, the above is skipped but there is still some set up required. As case changing can only be applied to bytes, and they have to be in the ASCII range, we define a series of data stores to represent them, and the data are used such that only these are ever case-changed. We do open and close one file to force allocation of a read; this keeps all engines in line.

\begin{verbatim}
\cs_set_protected:Npn \__char_tmp:NN #1#2
\quark_if_recursion_tail_stop:N #2
\tl_const:cn { c__char_uppercase_ #2 _tl } {#1}
\tl_const:cn { c__char_lowercase_ #1 _tl } {#2}
\tl_const:cn { c__char_foldcase_ #1 #2 _tl } {#1 #2}
\ior_open:Nn \g__char_data_ior { UnicodeData.txt }
\ior_close:N \g__char_data_ior
\end{verbatim}
Chapter 81

l3text implementation

81.1 Internal auxiliaries

\s__text_stop Internal scan marks.
\scan_new:N \s__text_stop
(End definition for \s__text_stop.)

\q__text_nil Internal quarks.
\quark_new:N \q__text_nil
(End definition for \q__text_nil.)

\__text_quark_if_nil_p:n Branching quark conditional.
\__text_quark_if_nil:nTF
\__kernel_quark_new_conditional:Nn \__text_quark_if_nil:n { TF }
(End definition for \__text_quark_if_nil:nTF.)

\q__text_recursion_tail Internal recursion quarks.
\q__text_recursion_stop
\quark_new:N \q__text_recursion_tail
\quark_new:N \q__text_recursion_stop
(End definition for \q__text_recursion_tail and \q__text_recursion_stop.)

\text_use_i_delimit_by_q_recursion_stop:nw Functions to gobble up to a quark.
\cs_new:Npn \__text_use_i_delimit_by_q_recursion_stop:nw
#1 #2 \q__text_recursion_stop {#1}
(End definition for \__text_use_i_delimit_by_q_recursion_stop:nw.)

\__text_if_recursion_tail_stop_do:Nn Functions to query recursion quarks.
\__kernel_quark_new_test:N \__text_if_recursion_tail_stop_do:Nn
(End definition for \__text_if_recursion_tail_stop_do:Nn.)
The idea here is to take a token and ensure that if it's an implicit char, we output the explicit version. Otherwise, the token needs to be unchanged. First, we have to split between control sequences and everything else.

For control sequences, we can check for macros versus other cases using \if_meaning:w, then explicitly check for \chardef and \mathchardef.

For character tokens, we need to filter out the implicit characters from those that are explicit. That's done here, then if necessary we work out the category code and generate the char. To avoid issues with alignment tabs, that one is done by elimination rather than looking up the code explicitly. The trick with finding the charcode is that the \TeX messages are either the ⟨something⟩ character ⟨char⟩ or the ⟨type⟩ ⟨char⟩.
An idea from \texttt{l3char}: we need to get the category code of a specific token, not the general case.
\else: \
\if_catcode:w \exp_not:N #1 \c_alignment_token 
 4 
\else: \
\if_catcode:w \exp_not:N #1 \c_math_superscript_token 
 7 
\else: \
\if_catcode:w \exp_not:N #1 \c_math_subscript_token 
 8 
\else: \
\if_catcode:w \exp_not:N #1 \c_space_token 
 10 
\else: \
\if_catcode:w \exp_not:N #1 \c_catcode_letter_token 
 11 
\else: \
\if_catcode:w \exp_not:N #1 \c_catcode_other_token 
 12 
\else: 
13 
\fi: 
\fi: 
\fi: 
\fi: 
\fi: 
\fi: 
\fi: 
{ \fi: }

(End definition for \_text_char_catcode:N.)

\_text_if_expandable:NTF Test for tokens that make sense to expand here: that is more restrictive than the engine view.
\prg_new_conditional:Nnn \_text_if_expandable:N #1 { T , F , TF } 
{ 
\token_if_expandable:NTF #1 
1 
\bool_lazy_any:nTF 
2 
{ \token_if_protected_macro_p:N #1 } 
3 
{ \token_if_protected_long_macro_p:N #1 } 
4 
{ \token_if_eq_meaning_p:NN \q__text_recursion_tail #1 } 
5 
{ \prg_return_false: } 
6 
{ \prg_return_true: } 
7 
{ \prg_return_false: }
8
(End definition for \_text_if_expandable:NTF.)

1195
81.3 Configuration variables

$\ltextaccents$ \texttt{l_text_accents_tl}$
Special cases for accents and letter-like symbols, which in some cases will need to be converted further.

\begin{verbatim}
\tl_new:N \l_text_accents_tl
\tl_set:Nn \l_text_accents_tl { \^ \~ \= \u \. " \r \H \v \d \c \k \b \t }
\end{verbatim}

$\ltextletterlike$ \texttt{l_text_letterlike_tl}$
\begin{verbatim}
\tl_new:N \l_text_letterlike_tl
\tl_set:Nn \l_text_letterlike_tl { \AA \aa \\AE \ae \\DH \dh \\IJ \ij \\L \l \\NG \ng \\O \o \\OE \oe \\SS \ss \\TH \th }
\end{verbatim}
(End definition for $\ltextaccents$ \texttt{l_text_accents_tl}$ and $\ltextletterlike$ \texttt{l_text_letterlike_tl}$.$ These variables are documented on page 266.)

$\ltextcaseexclude$ \texttt{l_text_case_exclude_arg_tl}$
Non-text arguments.

\begin{verbatim}
\tl_new:N \l_text_case_exclude_arg_tl
\tl_set:Nn \l_text_case_exclude_arg_tl { \begin \cite \end \label \ref }
\end{verbatim}
(End definition for $\ltextcaseexclude$ \texttt{l_text_case_exclude_arg_tl}$.$ This variable is documented on page 267.)

$\ltextmath$ \texttt{l_text_math_arg_tl}$
Math mode as arguments.

\begin{verbatim}
\tl_new:N \l_text_math_arg_tl
\tl_set:Nn \l_text_math_arg_tl { \ensuremath }
\end{verbatim}
(End definition for $\ltextmath$ \texttt{l_text_math_arg_tl}$.$ This variable is documented on page 267.)

$\ltextmathdelims$ \texttt{l_text_math_delims_tl}$
Paired math mode delimiters.

\begin{verbatim}
\tl_new:N \l_text_math_delims_tl
\tl_set:Nn \l_text_math_delims_tl { $ $ \( \) }
\end{verbatim}
(End definition for $\ltextmathdelims$ \texttt{l_text_math_delims_tl}$.$ This variable is documented on page 267.)

$\ltextexpandexclude$ \texttt{l_text_expand_exclude_tl}$
Commands which need not to expand.

\begin{verbatim}
\tl_new:N \l_text_expand_exclude_tl
\tl_set:Nn \l_text_expand_exclude_tl { \begin \cite \end \label \ref }
\end{verbatim}
(End definition for $\ltextexpandexclude$ \texttt{l_text_expand_exclude_tl}$.$ This variable is documented on page 267.)

$\ltextmathmode$ \texttt{l_text_math_mode_tl}$
Used to control math mode output: internal as there is a dedicated setter.

\begin{verbatim}
\tl_new:N \l_text_math_mode_tl
\end{verbatim}
(End definition for $\ltextmathmode$ \texttt{l_text_math_mode_tl}$.$)
81.4 Expansion to formatted text

Markers for implicit char handling.
\text_chardef:D \c__text_chardef_space_token = '\ %'
\text_chardef:D \c__text_mathchardef_space_token = '\ %'
\text_chardef:D \c__text_chardef_group_begin_token = '{ % '
\text_chardef:D \c__text_mathchardef_group_begin_token = '{ % '
\text_chardef:D \c__text_chardef_group_end_token = '% } '
\text_chardef:D \c__text_mathchardef_group_end_token = '% } '

(End definition for \c__text_chardef_space_token and others.)

After precautions against & tokens, start a simple loop: that of course means that “text” cannot contain the two recursion quarks. The loop here must be \textit{f}-type expandable; we have arbitrary user commands which might be protected \textit{and} take arguments, and if the expansion code is used in a typesetting context, that will otherwise explode. (The same issue applies more clearly to case changing: see the example there.)

\cs_new:Npn \text_expand:n #1
\__text_expand_result:n { }
Before we get into the real work, we have to watch out for problematic implicit characters: spaces and grouping tokens. Converting these to explicit characters later would lead to real issues as they are not \(\text{N}\)-type. A space is the easy case, so it’s dealt with first: just insert the explicit token and continue the loop.

Implicit \(/\) offer two issues. First, the token could be an implicit brace character: we need to avoid turning that into a brace group, so filter out the cases manually. Then we handle the case where an implicit group is present. That is done in an “open-ended” way: there’s the possibility the closing token is hidden somewhere.
The first step in dealing with $N$-type tokens is to look for math mode material: that needs to be left alone. The starting function has to be split into two as we need \verb|\quark-\if_recursion_tail_stop:N| first before we can trigger the search. We then look for matching pairs of delimiters, allowing for the case where math mode starts but does not end. Within math mode, we simply pass all the tokens through unchanged, just checking the $N$-type ones against the end marker.
At this stage, either we have a control sequence or a simple character: split and handle.

Next we exclude math commands: this is mainly as there might be an \emph{\texttt{ensuremath}}.
Accents.

Another list of exceptions: these ones take no arguments so are easier to handle.
\texttt{\LaTeX2e}'s \texttt{protect} makes life interesting. Where possible, we simply remove it and replace with the "parent" command; of course, the \texttt{protect} might be explicit, in which case we need to leave it alone if it's required. There is also the case of a straight \texttt{@protected@testopt} to cover.

\begin{verbatim}
\cs_new:Npn \__text_expand_letterlike:NN #1#2
{\__text_if_recursion_tail_stop_do:Nn #2
 { \__text_expand_cs:N #1 }
 \cs_if_eq:NNTF #2 #1
 { \__text_use_i_delimit_by_q_recursion_stop:nw
  { \__text_expand_store:n {#1}
   \__text_expand_loop:w
  }
 }
 { \__text_expand_letterlike:NN #1 }
}
\end{verbatim}
Deal with encoding-specific commands

\cs_new:Npn \__text_expand_encoding:N #1
{\bool_lazy_or:nnTF{\cs_if_eq_p:NN #1 \@current@cmd}{\cs_if_eq_p:NN #1 \@changed@cmd}{\exp_after:wN \__text_expand_loop:w \__text_expand_encoding_escape:NN}{\__text_expand_replace:N #1}}

\cs_new:Npn \__text_expand_encoding_escape:NN #1#2 {\exp_not:n {#1}}

See if there is a dedicated replacement, and if there is, insert it.

\cs_new:Npn \__text_expand_replace:N #1
{\bool_lazy_and:nnTF{\cs_if_exist_p:c { l__text_expand_ \token_to_str:N #1 _tl }}{\bool_lazy_or_p:nn{\token_if_cs_p:N #1}{\token_if_active_p:N #1}}{\exp_args:Nv \__text_expand_replace:n { l__text_expand_ \token_to_str:N #1 _tl } }{ \__text_expand_store:n {#1} \__text_expand_loop:w}}

Finally, expand any macros which can be: this then loops back around to deal with what they produce. The only issue is if the token is \exp_not:n, as that must apply to the following balanced text.

\cs_new:Npn \__text_expand_replace:n #1 { \__text_expand_loop:w #1}
Since `\exp_not:n` is actually a primitive, it allows a strange syntax and it particular the primitive expands what follows and discards spaces and `\scan_stop:` until finding a braced argument (the opening brace can be implicit but we will not support this here). Here, we repeatedly `f`-expand after such an `\exp_not:n`, and test what follows. If it is a brace group, then we found the intended argument of `\exp_not:n`. If it is a space, then the next `f`-expansion will eliminate it. If it is an `N`-type token then `\__text_=expand_unexpanded:N` leaves the token to be expanded if it is expandable, and otherwise removes it, assuming that it is `\scan_stop:`. This silently hides errors when `\exp_not:n` is incorrectly followed by some non-expandable token other than `\scan_stop:`, but this should be pretty rare, and there is no good error recovery anyways.

```
\cs_new:Npn \__text_expand_unexpanded:w
\exp_after:wN \__text_expand_unexpanded_test:w
\exp:w \exp_end_continue_f:w
\cs_new:Npn \__text_expand_unexpanded_test:w \#1 \#_text_recursion_stop
\tl_if_head_is_group:nTF {#1}
{ \__text_expand_unexpanded:n }
{
\__text_expand_unexpanded:w
\tl_if_head_is_N_type:nT {#1} { \__text_expand_unexpanded:N }
}
#1 \#_text_recursion_stop
\cs_new:Npn \__text_expand_unexpanded:N #1
\exp_after:wN \if_meaning:w \exp_not:N #1 #1
\else:
\exp_after:wN #1
\fi:
\cs_new:Npn \__text_expand_unexpanded:n #1
\__text_expand_store:n {#1}
\__text_expand_loop:w
```

(`End definition for `\text-expand:n` and others. This function is documented on page 264.)

Create equivalents to allow replacement.

```
\cs_new_protected:Npn \text_declare_expand_equivalent:Nn #1#2
\tl_clear_new:c { l__text_expand_ \token_to_str:N #1 _tl }
\tl_set:cn { l__text_expand_ \token_to_str:N #1 _tl } {#2}
\cs_generate_variant:Nn \text_declare_expand_equivalent:Nn { c }
```

(`End definition for `\text_declare_expand_equivalent:Nn`. This function is documented on page 264.)

(//package)
Chapter 82

\text-case implementation

82.1 Case changing

The user level functions here are all wrappers around the internal functions for case changing.

As for the expansion code, the business end of case changing is the handling of \textit{N}-type tokens. First, we expand the input fully (so the loops here don’t need to worry about awkward look-aheads and the like). Then we split into the different paths.

The code here needs to be \textit{f}-type expandable to deal with the situation where case changing is applied in running text. There, we might have case changing as a document command and the text containing other non-expandable document commands.
If we use an e-type expansion and wrap each token in \exp_not:n, that would explode: the document command grabs \exp_not:n as an argument, and things go badly wrong. So we have to wrap the entire result in exactly one \exp_not:n, or rather in the kernel version.

As for expansion, collect up the tokens for future use.

The main loop is the standard tl action type.
For a group, we could worry about whether this contains a character or not. However, that would make life very complex for little gain: exactly what a first character is is rather weakly-defined anyway. So if there is a group, we simply assume that a character has been seen, and for title case we switch to the "rest of the tokens" situation. To avoid having too much testing, we use a two-step process here to allow the titlecase functions to be separate.

\cs_new:Npn \__text_change_case_group_lower:nnn #1#2#3
\__text_change_case_store:o
\exp_after:wN
\exp:w
\__text_change_case_aux:nnn {#3} {#1} {#2}
\__text_change_case_loop:nnw { lower } {#2}
\__text_change_case_loop:nnw {#1} {#2}
\__text_change_case_group_upper:nnn
\cs_new_eq:NN \__text_change_case_group_upper:nnn \__text_change_case_group_lower:nnn
\cs_new:Npn \__text_change_case_group_title:nnn #1#2#3
\__text_change_case_store:o
\exp_after:wN
\exp:w
\__text_change_case_aux:nnn {#3} {#1} {#2}
\__text_change_case_break:w
\use:x
\use:x
\cs_new:Npn \exp_not:N \__text_change_case_space:nnw ##1##2 \c_space_tl
\__text_change_case_store:n { ~ }
The first step of handling N-type tokens is to filter out the end-of-loop. That has to be done separately from the first real step as otherwise we pick up the wrong delimiter. The loop here is the same as the expand one, just passing the additional data long. If no close-math token is found then the final clean-up is forced (i.e. there is no assumption of “well-behaved” input in terms of math mode).
Once potential math-mode cases are filtered out the next stage is to test if the token grabbed is a control sequence: the two routes the code may take are then very different.

To deal with a control sequence there is first a need to test if it is on the list which indicate that case changing should be skipped. That’s done using a loop as for the other special cases. If a hit is found then the argument is grabbed and passed through as-is.
Letter-like commands may still be present: they are set up using a simple lookup approach, so can easily be handled with no loop. If there is no hit, we are at the end of the process: we loop around. Letter-like chars are all available only in upper- and lowercase, so titlecasing maps to the uppercase version.

For upper- and lowercase changes, once we get to this stage there are only a couple of questions remaining: is there a language-specific mapping and is there the special case of a terminal sigma. If not, then we pass to a simple character mapping.
If the current character is an uppercase sigma, a check is made on the next item in
the input. If it is \textit{N}-type and not a control sequence then there is a look-ahead phase:
the logic here is simply based on letters. The one exception is Dutch: see below.

In the 8-bit engines, we have to look ahead once we find the first byte of the possible hit.
For titlecasing, we need to fully expand the new character to see if it is a letter (or active) But that means looking ahead in the 8-bit case, so we have to grab the required tokens up-front. Life is a lot easier for Unicode TeX’s, where we just have one token to worry about. The one wrinkle here is that for look-ahead we’d get into trouble: luckily, only Dutch has that issue.
For Unicode engines we can handle all characters directly. However, for the 8-bit engines the aim is to deal with (a subset of) Unicode (UTF-8) input. They deal with that by making the upper half of the range active, so we look for that and if found work out how many UTF-8 octets there are to deal with. Those can then be grabbed to reconstruct the full Unicode character, which is then used in a lookup. (As will become obvious below, there is no intention here of covering all of Unicode.) For (u)p-TEX there are a limited number of tokens we can touch.
\cs_new:Npn \__text_change_case_char_auxii:nnnN #1#2#3#4
  \__text_change_case_store:n {#4}
  \use:c { \__text_change_case_char_next_:nn } {#2} {#3}
\endinput

\cs_new:Npn \__text_change_case_char_UTFviii:nnnNN #1#2#3#4#5
  \__text_change_case_char_UTFviii:nnnn {#1} {#2} {#3} {#4#5#6}
\cs_new:Npn \__text_change_case_char_UTFviii:nnnnNN #1#2#3#4#5#6#7
  \__text_change_case_char_UTFviii:nnnn {#1} {#2} {#3} {#4#5#6#7}
\cs_new:Npn \__text_change_case_char_UTFviii:nnnnNNN #1#2#3#4#5#6#7#8
  \__text_change_case_char_UTFviii:nnnn {#1} {#2} {#3} {#4#5#6#7#8}
\cs_new:Npn \__text_change_case_char_UTFviii:nnnn #1#2#3#4
  \cs_if_exist:cTF { c__text_#1 case_ \tl_to_str:n {#4} _tl }
  \__text_change_case_store:v { c__text_#1 case_ \tl_to_str:n {#4} _tl }
  \use:c { \__text_change_case_char_next_:nn } {#2} {#3}
\endinput
\__text_change_case_upper_de-alt:nnnN \__text_change_case_upper_de-alt:nnnNN
\bool_lazy_or:nnTF
  \sys_if_engine_luatex_p: 
  \sys_if_engine_xetex_p: 
  \cs_new:cpn { \__text_change_case_upper_de-alt:nnnN } #1#2#3#4
  \int_compare:nNnTF { '#4 } = { "00DF }
  \__text_change_case_store:e
    \__text_char_catcode:N #4 
  \use:c { \__text_change_case_char_next_:nn } {#2} {#3}
\endinput

A simple alternative version for German.
\begin{verbatim}
\text_change_case_upper_de-alt:nnnN
\text_change_case_upper_de-alt:nnnNN\end{verbatim}
For Greek uppercasing, we need to know if characters in the Greek range have accents. That means doing a NFD conversion first, then starting a search. As described by the Unicode CLDR, Greek accents need to be found after any U+0308 (diaeresis) and are done in two groups to allow for the canonical ordering. The implementation here follows the data and examples from ICU (https://sites.google.com/site/icusite/design/case/greek-upper), although necessarily the implementation is somewhat different.

At this stage we have the first NFD codepoint as #3. What we need to know is whether after that we have another character token, either from the NFD or directly in the input.

If not, we store the changed character at this stage.

\textchangeupperel:nNnN
\textchangeupperel:dialytika:nNnNn
\textchangeupperel:hiatus:nNnN
\textchangeupperel:gobble:nN
\textchangeupperel:if_greek:nTF
\textchangeupperel:if_greek:accent:nTF
\textchangeupperel:if_greek:diacritic:nTF
\textchangeupperel:if_takes_dialytika:nTF
\textchangeupperel_if_greek:n
\textchangeupperel_if_greek_accent:n
\textchangeupperel_if_greek_diacritic:n
\textchangeupperel_if_takes_dialytika:n
\textchangeupperel_if_greek:nTF
\textchangeupperel_if_greek:accent:nTF
\textchangeupperel_if_greek_diacritic:nTF
\textchangeupperel_if_takes_dialytika:nTF
\textchangeupperel_if_greek:n
\textchangeupperel_if_greek_accent:n
\textchangeupperel_if_greek_diacritic:n
\textchangeupperel_if_takes_dialytika:n
\textchangeupperel_if_greek:n
\textchangeupperel_if_greek_accent:n
\textchangeupperel_if_greek_diacritic:n
\textchangeupperel_if_takes_dialytika:n
\textchangeupperel_if_greek:n
\textchangeupperel_if_greek_accent:n
\textchangeupperel_if_greek_diacritic:n
\textchangeupperel_if_takes_dialytika:n
\textchangeupperel_if_greek:n
Now, we check the detail of the next codepoint: again we filter out the not-a-char cases, before checking if it’s an dialytika, accent or diacritic. (The latter do not have the same hiatus behavior as accents.)

\begin{verbatim}
\cs_new:Npn \__text_change_case_upper_el:NnnN #1#2#3#4
{\token_if_cs:NTF #4
\__text_change_case_store:e { \char_uppercase:N #1 }
\__text_change_case_loop:nnw {#2} {#3} #4
}\int_compare:nNnTF { '#4 } = { "0308 }
{ \__text_change_case_upper_el_dialytika:nnN {#2} {#3} #1 }
{ \__text_change_case_if_greek_accent:nTF { '#4 }
{ \__text_change_case_upper_el_hiatus:nnNw {#2} {#3} #1 }
{ \__text_change_case_if_greek_diacritic:nTF { '#4 }
{ \__text_change_case_store:e { \char_uppercase:N #1 }
\__text_change_case_loop:nnw {#2} {#3} #4
}
{ \__text_change_case_store:e { \char_uppercase:N #3 } }
\__text_change_case_upper_el_gobble:nnw {#1} #4
}
}\__text_change_case_upper_el_dialytika:nnN {#1} {#2} {#3} #4
\end{verbatim}

We handle dialytika in parts as it’s also needed for the hiatus. We know only two letters take it, so we can shortcut here on the second part of the tests.

\begin{verbatim}
\cs_new:Npn \__text_change_case_upper_el_dialytika:nnN #1#2#3
{\__text_change_case_if_takes_dialytika:nTF { '#3 }
\__text_change_case_upper_el_dialytika:nnN {#1} {#2} #3
\__text_change_case_upper_el_gobble:nnw {#1} #2
\__text_change_case_upper_el_dialytika:nnN {#1} #1
{ \bool_lazy_or:nnTF
{ \int_compare_p:nNn { '#1 } = { "0399 } }
{ \int_compare_p:nNn { '#1 } = { "03B9 } }
{ \char_generate:nn { "03AA } { \__text_char_catcode:N #1 } }
}\end{verbatim}
Adding a hiatus needs some of the same ideas, but if there is not one we skip this code point, hence needing a separate function.

\cs_new:Npn \__text_change_case_upper_el_hiatus:nnNw #1#2#3#4 \q__text_recursion_stop
{ \__text_change_case_store:e { \char_uppercase:N #3 }
\tl_if_head_is_N_type:nTF {#4}
{ \__text_change_case_upper_el_hiatus:nnN }
\__text_change_case_loop:nnw }
\__text_change_case_upper_el_gobble:nnw {#1} {#2} #4 \q__text_recursion_stop
\cs_new:Npn \__text_change_case_upper_el_hiatus:nnN #1#2#3
{ \token_if_cs:NTF #3
{ \__text_change_case_upper_el_dialytika:N #3
{ \__text_change_case_upper_el_gobble:nnw {#1} {#2} }
}
\__text_change_case_upper_el_gobble:nnw {#1} {#2} #3 \q__text_recursion_stop
\cs_new:Npn \__text_change_case_upper_el_dialytika:N #1
{ \bool_lazy_or:nnTF
{ \token_if_cs_p:N #1 }
{ \__text_change_case_if_greek_accent_p:n { #1 }
{ \__text_change_case_if_greek_diacritic_p:n { #1 }
{ \__text_change_case_loop:nnw {#1} {#2} #3 }
\__text_change_case_upper_el_gobble:nnw {#1} {#2} }
}
\__text_change_case_upper_el_gobble:nnw {#1} {#2} #3 \q__text_recursion_stop
\bool_lazy_or:nnTF
{ \token_if_cs_p:N #3 }
{ \__text_change_case_if_greek_accent_p:n { #3 }
{ \__text_change_case_if_greek_diacritic_p:n { #3 }
{ \__text_change_case_loop:nnw {#1} {#2} #3 }
\__text_change_case_upper_el_gobble:nnw {#1} {#2} }
}
\__text_change_case_loop:nnw {#1} {#2} #3 \q__text_recursion_stop
\prg_new_conditional:Nppnn \__text_change_case_if_greek:n #1 { TF }
{ }

For clearing out trailing combining marks after we have dealt with the first one.

\cs_new:Npn \__text_change_case_upper_el_gobble:nnw #1#2#3 \q__text_recursion_stop
{ \tl_if_head_is_N_type:nTF {#3}
{ \__text_change_case_upper_el_dialytika:N #3
{ \__text_change_case_upper_el_gobble:nnN }
\__text_change_case_loop:nnw }
\__text_change_case_upper_el_gobble:nnw {#1} {#2} #3 \q__text_recursion_stop
\cs_new:Npn \__text_change_case_upper_el_dialytika:N #1
{ \bool_lazy_or:nnTF
{ \token_if_cs_p:N #1 }
{ \__text_change_case_if_greek_accent_p:n { #1 }
{ \__text_change_case_if_greek_diacritic_p:n { #1 }
{ \__text_change_case_loop:nnw {#1} {#2} #3 }
\__text_change_case_upper_el_gobble:nnw {#1} {#2} }
}
\__text_change_case_upper_el_gobble:nnw {#1} {#2} #3 \q__text_recursion_stop
\bool_lazy_or:nnTF
{ \token_if_cs_p:N #3 }
{ \__text_change_case_if_greek_accent_p:n { #3 }
{ \__text_change_case_if_greek_diacritic_p:n { #3 }
{ \__text_change_case_loop:nnw {#1} {#2} #3 }
\__text_change_case_upper_el_gobble:nnw {#1} {#2} }
}
\__text_change_case_loop:nnw {#1} {#2} #3 \q__text_recursion_stop
\prg_new_conditional:Nppnn \__text_change_case_if_greek:n #1 { TF }
{ }

Luckily the Greek range is limited and clear.

\prg_new_conditional:Nppnn \__text_change_case_if_greek:n #1 { TF }
{ }
We follow ICU in adding a few extras to the accent list here.

\prg_new_conditional:Npnn \__text_change_case_if_greek_accent:n #1 \{ TF , p \}
\prg_new_conditional:Npnn \__text_change_case_if_greek_diacritic:n #1 \{ TF , p \}
There is one special case in Greek that needs to be picked up based on being an isolated letter. We do that using a test similar to final sigma, but it has to fire off from the space grabber.

(End definition for \text_change_case_upper_el:nnn and others.)


\cs_new:Npn \_text_change_case_boundary_upper_el:nnN #1#2#3
{
\bool_lazy_or:nnTF
{ \token_if_cs_p:N #3 }
{
! \bool_lazy_or_p:nn
{ \int_compare_p:nNn { '#3 } = { "03AE } }
{ \int_compare_p:nNn { '#3 } = { "1F22 } }
}
{ \_text_change_case_loop:nnw }
{ \_text_change_case_boundary_upper_el:nnw }
{#1} {#2} #3
}
\cs_new:Npn \_text_change_case_boundary_upper_el:nnNw#1#2#3#4
{\_text_change_case_loop:nnw {#2} {#3} #1#4 }
\cs_new:Npn \_text_change_case_title_el:nnnN#1#2#3#4
{ \_text_change_case_char:nnnN {#1} {#2} {#3} #4 }
\cs_new:Npn \_text_change_case_lower_lt:nnnN#1#2#3#4
{\_text_change_case_char:nnnN #1#2#3#4 }
\cs_new:Npn \_text_change_case_lower_lt:nnw#1#2#3#4
{ \_text_change_case_char:nnw {#1} {#2} {#3} #4 }

(End definition for \_text_change_case_boundary_upper_el:nnw and others.)

Titlecasing retains accents, but to prevent the uppercasing code from kicking in, there has to be an explicit function here.

\bool_lazy_or:nnT
{ \sys_if_engine_luatex_p: }
{ \sys_if_engine_xetex_p: }
{ \cs_new:Npn \_text_change_case_title_el:nnnN #1#2#3#4
{ \_text_change_case_char:nnnN #1} {#2} {#3} #4 }

(End definition for \_text_change_case_title_el:nnnN.)

For Lithuanian, the issue to be dealt with is dots over lower case letters: these should be present if there is another accent. The first step is a simple match attempt: look for the
three uppercase accented letters which should gain a dot-above char in their lowercase form.

\bool_lazy_or:nnT
\sys_if_engine_luatex_p: }
\sys_if_engine_xetex_p: }
\cs_new:Npn \_text_change_case_lower_lt:nnnN #1#2#3#4
  \exp_args:Ne \_text_change_case_lower_lt_auxi:nnnN
  \int_case:nn { '#4 }
  { "0049 } { "0069 }
  { "004A } { "006A }
  { "012E } { "012F }
  }
  {#2} {#3} #4
\__text_change_case_loop:nnw {#2} {#3}
\__text_change_case_store:e
\char_generate:nn { "0069 } { \__text_char_catcode:N #4 }
\char_generate:nn { "0307 } { \__text_char_catcode:N #4 }
\char_generate:nn { #1 } { \_text_char_catcode:N #4 }
\_text_change_case_loop:nnw {#2} {#3}
\_text_change_case_store:e
\char_generate:nn { "0069 } { \_text_char_catcode:N #4 }
\char_generate:nn { "0307 } { \_text_char_catcode:N #4 }
\char_generate:nn { #1 } { \_text_char_catcode:N #4 }
\_text_change_case_loop:nnw {#2} {#3}

If there was a hit, output the result with the dot-above and move on. Otherwise, look for one of the three letters that can take a combining accent: I, J and I-ogonek.

\cs_new:Npn \_text_change_case_lower_lt:nnnN #1#2#3#4
  \tl_if_blank:nTF {#1}
  { \_text_change_case_lower_lt_auxi:nnnN
    \int_case:nn { '#4 }
    { "0049 } { "0069 }
    { "004A } { "006A }
    { "012E } { "012F }
  }
  {#2} {#3} #4
\__text_change_case_loop:nnw {#2} {#3}
\_text_change_case_store:e
\char_generate:nn { "0069 } { \__text_char_catcode:N #4 }
\char_generate:nn { "0307 } { \__text_char_catcode:N #4 }
\char_generate:nn { #1 } { \_text_char_catcode:N #4 }
\_text_change_case_loop:nnw {#2} {#3}

Again, branch depending on a hit. If there is one, we output the character then need to look for a combining accent: as usual, we need to be aware of the loop situation.

\cs_new:Npn \_text_change_case_lower_lt:nnnN #1#2#3#4
  \tl_if_blank:nTF {#1}
  { \_text_change_case_lower_sigma:nnnN {#2} {#2} {#3} #4 }
\_text_change_case_loop:nnw {#2} {#3} {#4}
The uppercasing version: first find i/j/i-ogonek, then look for the combining char: drop it if present.

The uppercasing version: first find i/j/i-ogonek, then look for the combining char: drop it if present.
\cs_new:Npn \__text_change_case_upper.lt_aux:nnnN #1#2#3#4
{\tl_if_blank:nTF{#1}
 { \__text_change_case_char:nnnN{ upper }{#2}{#3}#4 }
 { \__text_change_case_store:e
   \__text_char_catcode:N#4
   \__text_change_case_upper_lt:nw{#2}{#3} }
 }
\cs_new:Npn \__text_change_case_upper_lt:nnw #1#2#3 \\q__text_recursion_stop
{\tl_if_head_is_N_type:nTF{#3}
 { \__text_change_case_upper_lt:nnN }
 { \use:c{ \__text_change_case_char_next_:nn}{#1}{#2}#3\q__text_recursion_stop }
}
\cs_new:Npn \__text_change_case_upper_lt:nnN #1#2#3
{\bool_lazy_and:nnTF
 { ! \token_if_cs_p:N #3 }
 { \int_compare_p:nNn { '#3 } = { "0307 } }
 { \int_compare_p:nNn { '#3 } = { "0049 } }
 { \int_compare_p:nNn { '#3 } = { "0069 } }
 { \__text_change_case_store:e
   \char_generate:n { "0049 }\__text_change_case_upper.lt:nw{#2}{#3} }
 { \__text_change_case_char:nnnN{#1}{#2}#3 }
 }
\cs_new:Npn \__text_change_case_title_nl:nnnN #1#2#3#4
{\bool_lazy_or:nnTF
 { ! \token_if_cs_p:N #3 }
 { \int_compare_p:nNn { '#4 } = { "0049 } }
 { \int_compare_p:nNn { '#4 } = { "0069 } }
 { \__text_change_case_store:e
   \char_generate:n { "0049 }\__text_change_case_title.nl:nw{#2}{#3} }
 { \__text_change_case_title_nl:nnw{#1}{#2}{#3}#4 }
 }
\cs_new:Npn \__text_change_case_title_nl:nnw #1#2#3 \q__text_recursion_stop
{\tl_if_head_is_N_type:nTF{#3}
 { \__text_change_case_title_nl:nnN }
 { \use:c{ \__text_change_case_char_next_:nn}{#1}{#2}#3\q__text_recursion_stop }
}
\cs_new:Npn \__text_change_case_title_nl:nnN #1#2#3
{\bool_lazy_and:nnTF
 { ! \token_if_cs_p:N #3 }
 { \int_compare_p:nNn { '#3 } = { "0049 } }
 { \int_compare_p:nNn { '#3 } = { "0069 } }
 { \__text_change_case_store:e
   \char_generate:n { "0049 }\__text_change_case_title_nl:nw{#2}{#3} }
 { \__text_change_case_title_nl:nw{#1}{#2}{#3}#4 }
 }
\cs_new:Npn \__text_change_case_title_nl:nw #1#2#3 \q__text_recursion_stop
{\tl_if_head_is_N_type:nTF{#3}
 { \__text_change_case_title_nl:nw{#1}{#2}{#3} }
 { \use:c{ \__text_change_case_char_next_:nn}{#1}{#2}#3\q__text_recursion_stop }
}
\cs_new:Npn \__text_change_case_title_nl:nw #1#2#3
{\bool_lazy_and:nnTF
 { ! \token_if_cs_p:N #3 }

(End definition for \__text_change_cases_upper.lt:nnnN and others.)

For Dutch, there is a single look-ahead test for ij when title casing. If the appropriate letters are found, produce IJ and gobble the j/J.

\cs_new:Npn \__text_change_case_title_nl:nw #1#2#3#4
{\bool_lazy_or:nnTF
 { ! \token_if_cs_p:N #3 }
 { \int_compare_p:nNn { '#4 } = { "0049 } }
 { \int_compare_p:nNn { '#4 } = { "0069 } }
 { \__text_change_case_store:e
   \char_generate:n { "0049 }\__text_change_case_title_nl:nw{#2}{#3} }
 { \__text_change_case_title_nl:nw{#1}{#2}{#3}#4 }
 }
\cs_new:Npn \__text_change_case_title_nl:nw #1#2#3 \q__text_recursion_stop
{\tl_if_head_is_N_type:nTF{#3}
 { \__text_change_case_title_nl:nw{#1}{#2}{#3} }
 { \use:c{ \__text_change_case_char_next_:nn}{#1}{#2}#3\q__text_recursion_stop }
}
\cs_new:Npn \__text_change_case_title_nl:nw #1#2#3
{\bool_lazy_and:nnTF
 { ! \token_if_cs_p:N #3 }
The Turkic languages need special treatment for dotted-i and dotless-i. The lower casing rule can be expressed in terms of searching first for either a dotless-I or a dotted-I. In the latter case the mapping is easy, but in the former there is a second stage search.

After a dotless-I there may be a dot-above character. If there is then a dotted-i should be produced, otherwise output a dotless-i. When the combination is found both the dotless-I and the dot-above char have to be removed from the input.
\bool_lazy_or:nTF
  \token_if_cs_p:N #4
  \int_compare_p:nNn { ‘#4 } = { “0307 } }
  \__text_change_case_store:e
  \char_generate:nn { “0131 } { \__text_char_catcode:N #1 }
  \__text_change_case_loop:nnw {#2} {#3} #4
  \__text_change_case_store:e
  \char_generate:nn { “0069 } { \__text_char_catcode:N #1 }
  \__text_change_case_loop:nnw {#2} {#3}
\__text_change_case_upper_tr:nNn
For 8-bit engines, dot-above is not available so there is a simple test for an upper-case I. Then we can look for the UTF-8 representation of an upper case dotted-I without the combining char. If it’s not there, preserve the UTF-8 sequence as-is. With 8bit engines, we cannot completely preserve category codes, so we have to make some assumptions: output a “normal” i for the dotted case. As the original character here is catcode-13, we have to make a choice about handling of i: generate a “normal” one.

\cs_new:Npn \__text_change_case_lower_tr:nnnN #1#2#3#4
  \int_compare:nNnTF { ‘#4 } = { “0049 }
  \__text_change_case_store:V \c__text_dotless_i_tl
  \__text_change_case_loop:nnw {#1} {#3}
  \int_compare:nNnTF { ‘#4 } = { “00C4 }
  \__text_change_case_lower_tr:nnnNN {#1} {#2} {#3} {#4}
  \__text_change_case_char:nnnN {#1} {#2} {#3} #4
\cs_new:Npn \__text_change_case_lower_tr:nnnNN #1#2#3#4#5
  \int_compare:nNnTF { ‘#5 } = { “00B0 }
  \__text_change_case_store:e
  \char_generate:nn { “0069 }
  \char_value_catcode:n { “0069 }
  \__text_change_case_loop:nnw {#1} {#3}
  \__text_change_case_char:nnnN {#1} {#2} {#3} {#4} {#5}
(End definition for \__text_change_case_lower_tr:nNn and others.)

\text_change_case_upper_tr:nNn Uppercasing is easier: just one exception with no context.
82.2 Case changing data for 8-bit engines

For cases where there is an 8-bit option in the T1 font set up, a variant is provided in both cases. There are also a few extras for LGR.
For 8-bit engines we now need to define the case-change data for the multi-octet mappings. This data is here not in the char module as the multi-byte nature means they are never N-type. These need a list of what code points are doable in T1 so the list is hard coded (there’s no saving in loading the mappings dynamically). All of the straight-forward ones have two octets, so that is taken as read.
\exp_not:N \char_generate:nn {#4} { 13 }
\exp_after:wN \exp_after:wN \exp_after:wN
\exp_not:N \char_generate:nn {#5} { 13 }
\cs_set_protected:Npn \__text_tmp:w #1#2#3#4#5#6#7#8
{\tl_const:cx
  \char_generate:nn {#1} { 12 }
  \char_generate:nn {#2} { 12 }
  \_tl
}{
  \exp_after:wN \exp_after:wN \exp_after:wN
  \exp_not:N \char_generate:nn {#5} { 13 }
  \exp_after:wN \exp_after:wN \exp_after:wN
  \exp_not:N \char_generate:nn {#6} { 13 }
}{\__text_tmp:nnnn { upper } {#5} {#6} {#1} {#2}}
\__text_tmp:nnnn { title } {#5} {#6} {#1} {#2}
\__text_loop:nn
  { 00C0 } { 00E0 }
  { 00C1 } { 00E1 }
  { 00C2 } { 00E2 }
  { 00C3 } { 00E3 }
  { 00C4 } { 00E4 }
  { 00C5 } { 00E5 }
  { 00C6 } { 00E6 }
  { 00C7 } { 00E7 }
  { 00C8 } { 00E8 }
  { 00C9 } { 00E9 }
  { 00CA } { 00EA }
  { 00CB } { 00EB }
  { 00CC } { 00EC }
  { 00CD } { 00ED }
  { 00CE } { 00EE }
  { 00CF } { 00EF }
  { 00D0 } { 00F0 }
  { 00D1 } { 00F1 }
  { 00D2 } { 00F2 }
  { 00D3 } { 00F3 }
  { 00D4 } { 00F4 }
  { 00D5 } { 00F5 }
  { 00D6 } { 00F6 }
  { 00D8 } { 00F8 }
  { 00D9 } { 00F9 }
  { 00DA } { 00FA }
  { 00DB } { 00FB }
  { 00DC } { 00FC }
  { 00DD } { 00FD }
  { 00DE } { 00FE }

1228
Add T2 (Cyrillic) as this is doable using a classical `\texttt{MakeUppercase}` approach.
{ 0425 } { 0445 }
{ 0426 } { 0446 }
{ 0427 } { 0447 }
{ 0428 } { 0448 }
{ 0429 } { 0449 }
{ 042A } { 044A }
{ 042B } { 044B }
{ 042C } { 044C }
{ 042D } { 044D }
{ 042E } { 044E }
{ 042F } { 044F }

Greek support: everything in the two-octet range.

{ 0370 } { 0371 }
{ 0372 } { 0373 }
{ 0376 } { 0377 }
{ 037B } { 037D }
{ 037C } { 037D }
{ 0386 } { 03AC }
{ 0388 } { 03AD }
{ 0389 } { 03AE }
{ 038A } { 03AF }
{ 0391 } { 03B1 }
{ 0392 } { 03B2 }
{ 0393 } { 03B3 }
{ 0394 } { 03B4 }
{ 0395 } { 03B5 }
{ 0396 } { 03B6 }
{ 0397 } { 03B7 }
{ 0398 } { 03B8 }
{ 0399 } { 03B9 }
{ 039A } { 03BA }
{ 039B } { 03BB }
{ 039C } { 03BC }
{ 039D } { 03BD }
{ 039E } { 03BE }
{ 039F } { 03BF }
{ 03A0 } { 03CC }
{ 03A1 } { 03CD }
{ 03A2 } { 03CE }
{ 03A3 } { 03CF }
{ 03A4 } { 03D7 }
{ 03A5 } { 03D9 }
{ 03A6 } { 03DA }
{ 03A7 } { 03DB }
Odds and ends for Greek; mainly symbols that are for compatibility, but also things like the terminal sigma. Almost all are uppercase mappings, but there is one that is not!

```
\cs_set_protected:Npn \__text_tmp:w #1#2#3
\group_begin:
  \cs_set_protected:Npn \__text_tmp:w ##1##2##3##4##5##6##7##8
  {\tl_const:cx {\char_generate:nn {##1} { 12 } \char_generate:nn {##2} { 12 } _tl}}
  \use:x
  {\__text_tmp:w \char_to_utfviii_bytes:n { "#1 } \char_to_utfviii_bytes:n { "#2 } }
\group_end:
\__text_tmp:w { 0345 } { 0399 } { upper }
\__text_tmp:w { 03C2 } { 03A3 } { upper }
\__text_tmp:w { 03D0 } { 0392 } { upper }
\__text_tmp:w { 03D1 } { 0398 } { upper }
\__text_tmp:w { 03D5 } { 03A6 } { upper }
\__text_tmp:w { 03D6 } { 03A0 } { upper }
\__text_tmp:w { 03F0 } { 039A } { upper }
\__text_tmp:w { 03F1 } { 03A1 } { upper }
```
Odds and ends that are not simple one-to-one mappings. These are still two-octet code points.

\cs_set_protected:Npn \__text_tmp:w #1#2#3
\group_begin:
\cs_set_protected:Npn \__text_tmp:w ##1##2##3##4
\tl_const:cn
\group_end:
\__text_tmp:w #1#2#3#4#5#6#7
\cs_set_protected:Npn \__text_tmp:w #1#2#3#4#5#6#7#8
\tl_const:cx
\__text_tmp:w { 00DF } { SS } { upper }
\__text_tmp:w { 00DF } { Ss } { title }
\__text_tmp:w { 0131 } { I } { upper }

Greek support: the three-octet code points.

\cs_set_protected:Npn \__text_tmp:nnnnnn #1#2#3#4#5#6#7
\group_end:
\__text_tmp:w { 00DF } { SS } { upper }
\__text_tmp:w { 00DF } { Ss } { title }
\__text_tmp:w { 0131 } { I } { upper }

\cs_set_protected:Npn \__text_tmp:w #1#2#3#4#5#6#7#8
\tl_const:cx
\__text_tmp:w { 00DF } { SS } { upper }
\__text_tmp:w { 00DF } { Ss } { title }
\__text_tmp:w { 0131 } { I } { upper }
One three-octet special case for Greek: it also moves to two-octets!

```
\cs_set_protected:Npn \_text_tmp:w #1#2#3
\group_begin:
  \cs_set_protected:Npn \_text_tmp:w ##1##2##3##4##5##6##7##8
  \group_begin:
    \tl_const:cx
    { c__text__ #3 case_
      \char_generate:nn {##1} { 12 }
      \char_generate:nn {##2} { 12 }
      \char_generate:nn {##3} { 12 }
      _tl
    }
    \exp_after:wN \exp_after:wN \exp_after:wN
      \exp_not:N \char_generate:nn {##5} { 13 }
    \exp_after:wN \exp_after:wN \exp_after:wN
      \exp_not:N \char_generate:nn {##6} { 13 }
  \group_end:
  \use:x
  { \_text_tmp:w
    \char_to_utfviii_bytes:n { "#1 } 
    \char_to_utfviii_bytes:n { "#2 } 
  }
\group_end:
\_text_tmp:w { 1FBE } { 0399 } { upper }
\group_end:
```

The (fixed) look-up mappings for letter-like control sequences.

```
\group_begin:
  \cs_set_protected:Npn \_text_change_case_setup:NN #1#2
  { \quark_if_recursion_tail_stop:N \_text_change_case_setup:NN #1#2
    \tl_const:cn { c__text_lowercase_ \token_to_str:N #1 _tl } { #2 }
    \tl_const:cn { c__text_uppercase_ \token_to_str:N #2 _tl } { #1 }
  }
\_text_change_case_setup:NN
AA \aa
AE \ae
DH \dh
DJ \dj
IJ \ij
```

1236
To deal with possible encoding-specific extensions to \@uclclist, we check at the end of the preamble. This will therefore only apply to \LaTeX\ package mode.

\cs_if_exist:cT { \@uclclist }
{
  \AtBeginDocument
  {
    \group_begin:
      \cs_set_protected:Npn \__text_change_case_setup:Nn #1#2
        {
          \quark_if_recursion_tail_stop:N #1
          \tl_if_single_token:nT {#2}
            {
              \cs_if_exist:cF
                { \c__text_uppercase_ \token_to_str:N #1 _tl }
                { \tl_const:cn
                    { \c__text_uppercase_ \token_to_str:N #1 _tl }
                    { #2 }
                }
              \cs_if_exist:cF
                { \c__text_lowercase_ \token_to_str:N #2 _tl }
                { \tl_const:cn
                    { \c__text_lowercase_ \token_to_str:N #2 _tl }
                    { #1 }
                }
            }
        }
    \__text_change_case_setup:Nn \@uclclist
  }
}
Chapter 83

l3text-purify implementation

83.1 Purifying text

Functions to query recursion quarks.

As in the other parts of the module, we start off with a standard “action” loop, with expansion applied up-front.

As for expansion, collect up the tokens for future use.
The main loop is a standard “tl action”. Unlike the expansion or case changing, here any groups have to be run inline. Most of the business end is as before in the N-type token processing.

```latex
\cs_new:Npn \__text_purify_loop:w #1 \q__text_recursion_stop
\{\tl_if_head_is_N_type:nTF {#1}\{\__text_purify_N_type:N \}
\{\tl_if_head_is_group:nTF {#1}\{\__text_purify_group:n \}
\{\__text_purify_space:w \}
\}
\tl_if_head_is_N_type:nTF {#1}\{\__text_purify_end:w \}
\__text_purify_math_search:NNN
\exp_after:wN \__text_purify_math_start:NNw \l_text_math_delims_tl\q__text_recursion_tail ?
\__text_purify_math_result:n \{}\__text_purify_loop:w
\}
```

The first part of handling math mode is exactly the same as in the other functions: look for a start-of-math mode token and if found start a new loop tracking the closing token.

```latex
\cs_new:Npn \__text_purify_math_start:NNw #1#2#3 \q__text_recursion_stop
\{\__text_purify_math_loop:NNw #1#2#3 \q__text_recursion_stop
\__text_purify_math_result:n \{}\__text_purify_math_store:n #1
\}
```

1239
Then handle math mode as an argument: same outcomes, different input syntax.
For \( N \)-type tokens, we first look for a string-context replacement before anything else: this can therefore cover anything. Assuming we don’t find one, check to see if we can expand control sequences: if not, they have to be dropped. We also allow for LaTeX \texttt{\textbackslash protect}: there’s an assumption that we don’t have \texttt{\textbackslash protect} \{ \textbackslash oops \} or similar, but that’s also in the expansion code and seems like a reasonable balance.

Handle encoding commands, as detailed for expansion.
Now pre-define a range of standard commands that need dedicated definitions in purified text. First handle font-related stuff: all of this needs to be disabled.

```
\tl_map_inline:nn
{\fontencoding\fontfamily\fontseries\fontshape}
{\text\textnormal\textsf\texttt\textbf\textmd\textit\textsl}
```

(End definition for `\text_purify:n` and others. This function is documented on page 266.)
Environments have to be handled by pure expansion.

\__text_end_env:n

(End definition for \__text_end_env:n.)

Some common symbols and similar ideas.

Cross-referencing.

Spaces.
83.2 Accent and letter-like data for purifying text

In contrast to case changing, both 8-bit and Unicode engines need information for text purification to handle accents and letter-like functions: these all need to be removed. However, the results are of course engine-dependent.

For the letter-like commands, life is relatively easy: they are all simply added as standard exceptions. The only oddity is \SS, which gets converted to two letters. (At some stage an alternative version can presumably be added to babel or similar.)
Accent handling is a little more complex. Accents may exist as pre-composed codepoints or as independent glyphs. The former are all saved as single token lists, whilst for the latter the combining accent needs to be re-ordered compared to the character it applies to.

First set up the combining accents.

\__text_purify:NN

\cs_new:Npn \__text_purify:NN #1#2
{\cs_if_exist:cTF
{c__text_purify_ \token_to_str:N #1 _ \token_to_str:N #2 _tl}
{\exp_not:v
{c__text_purify_ \token_to_str:N #1 _ \token_to_str:N #2 _tl}}
\exp_not:n {#2}
\exp_not:v {c__text_purify_ \token_to_str:N #1 _tl}
}
\tl_map_inline:Nn \l_text_accents_tl
{\text_declare_purify_equivalent:Nn #1 \SS \SS}
Now we handle the pre-composed accents: the list here is taken from puenc.def. All of the precomposed cases take a single letter as their second argument. We do not try to cover the case where an accent is added to a “real” dotless-i or -j, or a æ/Æ. Rather, we assume that if the UTF-8 character is used, it will have the real accent character too.
\- o \00F5
\" o \00F6
\' u \00F9
\^ u \00FA
\^ u \00FB
\" u \00FC
\^ y \00FD
\" y \00FF
\= A \0100
\= a \0101
\u A \0102
\u a \0103
\k A \0104
\k a \0105
\^ C \0106
\^ c \0107
\^ C \0108
\^ c \0109
\. C \010A
\. c \010B
\^ C \010C
\^ c \010D
\v D \010E
\v d \010F
\= E \0112
\= e \0113
\u E \0114
\u e \0115
\. E \0116
\. e \0117
\k E \0118
\k e \0119
\v E \011A
\v e \011B
\^ G \011C
\^ g \011D
\u G \011E
\u g \011F
\. G \0120
\. g \0121
\c G \0122
\c g \0123
\^ H \0124
\h \0125
\i \0126
\\i \0127
\= I \012A
\= i \012B
\= \i \012B
\u I \012C
\u i \012D
\u \i \012D
\k I \012E
\H U { 0170 }
\H u { 0171 }
\k U { 0172 }
\k u { 0173 }
\^ W { 0174 }
\^ w { 0175 }
\^ Y { 0176 }
\^ y { 0177 }
\^ Y { 0178 }
\^ Z { 0179 }
\^ z { 017A }
\^ A { 01CD }
\^ a { 01CE }
\v I { 01D1 }
\v i { 01D2 }
\v U { 01D3 }
\v u { 01D4 }
\v G { 01E6 }
\v g { 01E7 }
\v K { 01E8 }
\v k { 01E9 }
\v O { 01EA }
\v o { 01EB }
\v j { 01F0 }
\v j { 01F0 }
\v G { 01F4 }
\v g { 01F5 }
\v N { 01F8 }
\v n { 01F9 }
\v \AE { 01FC }
\v \ae { 01FD }
\v \O { 01FE }
\v \o { 01FF }
\v H { 021E }
\v h { 021F }
. A { 0226 }
. a { 0227 }
\c E { 0228 }
\c e { 0229 }
. O { 022E }
. o { 022F }
= Y { 0232 }
= y { 0233 }
q_recursion_tail ? { }
q_recursion_stop
\group_end:
(End definition for \_text_purifyAccent:NN.)

3173: (</package>
Chapter 84

\texttt{l3box} implementation

84.1 Support code

\begin{verbatim}
\__box_dim_eval:w
\__box_dim_eval:n
Evaluating a dimension expression expandably. The only difference with \texttt{\dim_eval:n}
is the lack of \texttt{\dim_use:N}, to produce an internal dimension rather than expand it intocharacters.
\cs_new_eq:NN \__box_dim_eval:w \tex_dimexpr:D
\cs_new:Npn \__box_dim_eval:n #1
\{ \__box_dim_eval:w #1 \scan_stop: \}
\end{verbatim}

(End definition for \texttt{\__box_dim_eval:w} and \texttt{\__box_dim_eval:n}.)

\begin{verbatim}
\__kernel_kern:n
We need kerns in a few places. At present, we don't have a module for this concept, so
it goes in at first use: here. The idea is to avoid repeated use of the bare primitive.
\cs_new_protected:Npn \__kernel_kern:n #1
\{ \tex_kern:D \__box_dim_eval:n {#1} \}
\end{verbatim}

(End definition for \texttt{\__kernel_kern:n}.)

84.2 Creating and initialising boxes

The following test files are used for this code: \texttt{m3box001.lvt}.

\begin{verbatim}
\box_new:N
\box_new:c
Defining a new \texttt{(box)} register: remember that box 255 is not generally available.
\cs_new_protected:Npn \box_new:N #1
\{ \__kernel_chk_if_free_cs:N #1 \cs:w newbox \cs_end: \#1 \}
\cs_generate_variant:Nn \box_new:N{c}
\end{verbatim}

Clear a ⟨box⟩ register.
\begin{verbatim}
\cs_new_protected:Npn \box_clear:N #1
\{
\box_set_eq:NN #1 \c_empty_box
\}
\cs_new_protected:Npn \box_gclear:N #1
\{
\box_gset_eq:NN #1 \c_empty_box
\}
\cs_generate_variant:Nn \box_clear:N { c }
\cs_generate_variant:Nn \box_gclear:N { c }
\end{verbatim}
Clear or new.
\begin{verbatim}
\cs_new_protected:Npn \box_clear_new:N #1
\{ \box_if_exist:NTF #1 \{ \box_clear:N #1 \} \{ \box_new:N #1 \} \}
\cs_new_protected:Npn \box_gclear_new:N #1
\{ \box_if_exist:NTF #1 \{ \box_gclear:N #1 \} \{ \box_new:N #1 \} \}
\cs_generate_variant:Nn \box_clear_new:N { c }
\cs_generate_variant:Nn \box_gclear_new:N { c }
\end{verbatim}
Assigning the contents of a box to be another box.
\begin{verbatim}
\cs_new_protected:Npn \box_set_eq:NN #1#2
\{ \tex_setbox:D #1 \tex_copy:D #2 \}
\cs_new_protected:Npn \box_gset_eq:NN #1#2
\{ \tex_global:D \tex_setbox:D #1 \tex_copy:D #2 \}
\cs_generate_variant:Nn \box_set_eq:NN { c , Nc , cc }
\cs_generate_variant:Nn \box_gset_eq:NN { c , Nc , cc }
\end{verbatim}
Assigning the contents of a box to be another box, then drops the original box.
\begin{verbatim}
\cs_new_protected:Npn \box_set_eq_drop:NN #1#2
\{ \tex_setbox:D #1 \tex_box:D #2 \}
\cs_new_protected:Npn \box_gset_eq_drop:NN #1#2
\{ \tex_global:D \tex_setbox:D #1 \tex_box:D #2 \}
\cs_generate_variant:Nn \box_set_eq_drop:NN { c , Nc , cc }
\cs_generate_variant:Nn \box_gset_eq_drop:NN { c , Nc , cc }
\end{verbatim}
Copies of the \texttt{cs} functions defined in \texttt{l3basics}.
\begin{verbatim}
\prg_new_eq_conditional:NNn \box_if_exist:N \cs_if_exist:N \{
\prg_new_eq_conditional:NNn \box_if_exist:c \cs_if_exist:c \{
\end{verbatim}
84.3 Measuring and setting box dimensions

Accessing the height, depth, and width of a ⟨box⟩ register.
\begin{verbatim}
\cs_new_protected:Nn \box_ht:N \tex_ht:D
\cs_new_protected:Nn \box_dp:N \tex_dp:D
\cs_new_protected:Nn \box_wd:N \tex_wd:D
\end{verbatim}
The \texttt{box_ht:N} and \texttt{box_dp:N} primitives do not expand but rather are suitable for use after \texttt{\the} or inside dimension expressions. Here we obtain the same behaviour by using \texttt{\_\_box_dim_eval:n} (basically \texttt{\dimexpr}) rather than \texttt{\dim_eval:n} (basically \texttt{\the\dimexpr}).
\cs_new_protected:Npn \box_ht_plus_dp:N #1
\{ \__box_dim_eval:n { \box_ht:N #1 + \box_dp:N #1 } \}
\cs_generate_variant:Nn \box_ht_plus_dp:N { c }

\box_set_ht:Nn
\box_set_ht:cn
\box_gset_ht:Nn
\box_gset_ht:cn
\box_set_dp:Nn
\box_set_dp:cn
\box_gset_dp:Nn
\box_gset_dp:cn
\box_set_wd:Nn
\box_set_wd:cn
\box_gset_wd:Nn
\box_gset_wd:cn

Setting the size whilst respecting local scope requires copying; the same issue does not come up when working globally. When debugging, the dimension expression \texttt{#2} is surrounded by parentheses to catch early termination.

\cs_new_protected:Npn \box_set_dp:Nn #1#2
\{ \tex_setbox:D #1 = \tex_copy:D #1 \box_dp:N #1 \__box_dim_eval:n \{#2\}\}
\cs_generate_variant:Nn \box_set_dp:Nn { c }
\cs_new_protected:Npn \box_gset_dp:Nn #1#2
\{ \box_dp:N #1 \__box_dim_eval:n \{#2\} \}
\cs_generate_variant:Nn \box_gset_dp:Nn { c }
\cs_new_protected:Npn \box_set_ht:Nn #1#2
\{ \tex_setbox:D #1 = \tex_copy:D #1 \box_ht:N #1 \__box_dim_eval:n \{#2\}\}
\cs_generate_variant:Nn \box_set_ht:Nn { c }
\cs_new_protected:Npn \box_gset_ht:Nn #1#2
\{ \box_ht:N #1 \__box_dim_eval:n \{#2\} \}
\cs_generate_variant:Nn \box_gset_ht:Nn { c }
\cs_new_protected:Npn \box_set_wd:Nn #1#2
\{ \tex_setbox:D #1 = \tex_copy:D #1 \box_wd:N #1 \__box_dim_eval:n \{#2\}\}
\cs_generate_variant:Nn \box_set_wd:Nn { c }
\cs_new_protected:Npn \box_gset_wd:Nn #1#2
\{ \box_wd:N #1 \__box_dim_eval:n \{#2\} \}
\cs_generate_variant:Nn \box_gset_wd:Nn { c }

84.4 Using boxes

Using a \texttt{⟨box⟩}. These are just \TeX\ primitives with meaningful names.

\cs_new_eq:NN \box_use_drop:N \tex_box:D
\cs_new_eq:NN \box_use:N \tex_copy:D
\cs_generate_variant:Nn \box_use_drop:N { c }
\cs_generate_variant:Nn \box_use:N { c }
\box_move_left:nn
\box_move_right:nn
\box_move_up:nn
\box_move_down:nn

Move box material in different directions. When debugging, the dimension expression \texttt{#1} is surrounded by parentheses to catch early termination.

\cs_new_protected:Npm \box_move_left:nn \#1#2
\{ \tex_moveleft:D \__box_dim_eval:n \{#1\} \#2 \}
\cs_new_protected:Npm \box_move_right:nn \#1#2
\{ \tex_moveright:D \__box_dim_eval:n \{#1\} \#2 \}
\cs_new_protected:Npm \box_move_up:nn \#1#2
\{ \tex_raise:D \__box_dim_eval:n \{#1\} \#2 \}
\cs_new_protected:Npm \box_move_down:nn \#1#2
\{ \tex_lower:D \__box_dim_eval:n \{#1\} \#2 \}
84.5 Box conditionals

The primitives for testing if a ⟨box⟩ is empty/void or which type of box it is.

\texttt{\if hbox:N \if vbox:N \if_box_empty:N}

\texttt{\box_if_horizontal_p:N \box_if_horizontal_p:c \box_if_horizontal:N \box_if_horizontal:c}

\texttt{\box_if_vertical_p:N \box_if_vertical_p:c \box_if_vertical:N \box_if_vertical:c}

\texttt{\box_if_empty_p:N \box_if_empty_p:c \box_if_empty:N \box_if_empty:c}

Testing if a ⟨box⟩ is empty/void.

\texttt{\prg_new_conditional:Nppnn \box_if_horizontal:N #1 { p , T , F , TF }}
\texttt{\prg_new_conditional:Nppnn \box_if_vertical:N #1 { p , T , F , TF }}
\texttt{\prg_new_conditional:Nppnn \box_if_empty:N #1 { p , T , F , TF }}

(End definition for \texttt{\box_new:N} and others. These functions are documented on page 269.)

84.6 The last box inserted

\texttt{\box_set_to_last:N \box_set_to_last:c \box_gset_to_last:N \box_gset_to_last:c}

\texttt{\cs_new_protected:Npn \box_set_to_last:N #1} \texttt{\cs_new_protected:Npn \box_gset_to_last:N #1}

\texttt{\cs_generate_variant:Nn \box_set_to_last:N { c }} \texttt{\cs_generate_variant:Nn \box_gset_to_last:N { c }}

(End definition for \texttt{\box_set_to_last:N} and \texttt{\box_gset_to_last:N}. These functions are documented on page 272.)

84.7 Constant boxes

\texttt{\c_empty_box}

A box we never use.

\texttt{\box_new:N \c_empty_box}

(End definition for \texttt{\c_empty_box}. This variable is documented on page 272.)
84.8 Scratch boxes

Scratch boxes.

(End definition for \l_tmpa_box and others. These variables are documented on page 272.)

84.9 Viewing box contents

\TeX{}’s \showbox{} is not really that helpful in many cases, and it is also inconsistent with other \LTref{show} functions as it does not actually shows material in the terminal. So we provide a richer set of functionality.

Essentially a wrapper around the internal function, but evaluating the breadth and depth arguments now outside the group.

(End definition for \box_show:N and \box_show:Nnn. These functions are documented on page 273.)

Getting \TeX{} to write to the log without interruption the run is done by altering the interaction mode. For that, the \LTref{\epsilon-TeX} extensions are needed.

The internal auxiliary to actually do the output uses a group to deal with breadth and depth values. The use:n here gives better output appearance. Setting \long\tracingonline and \long\errorcontextlines is used to control what appears in the terminal.

(End definition for \box_log:N, \box_log:Nnn, and \__box_log:nNnn. These functions are documented on page 273.)
84.10 Horizontal mode boxes

\hbox:n

(The test suite for this command, and others in this file, is m3box002.ltx.)

Put a horizontal box directly into the input stream.

\hbox_set:Nn
\hbox_set:cn
\hbox_gset:Nn
\hbox_gset:cn

(End definition for \hbox_set:Nn and \hbox_gset:Nn. These functions are documented on page 273.)

\hbox_set_to_wd:Nnn
\hbox_set_to_wd:cnn
\hbox_gset_to_wd:Nnn
\hbox_gset_to_wd:cnn

Storing material in a horizontal box with a specified width. Again, put the dimension expression in parentheses when debugging.
\section*{Horizontal Boxes}

Storing material in a horizontal box. This type is useful in environment definitions.

\begin{verbatim}
\cs_new_protected:Npm \hbox_set:Nw #1
\{ \tex_setbox:D #1 \tex_hbox:D \c_group_begin_token \color_group_begin:
\}
\cs_new_protected:Npm \hbox_gset:Nw #1
\{ \tex_global:D \tex_setbox:D #1 \tex_hbox:D \c_group_begin_token \color_group_begin:
\}
\cs_generate_variant:Nn \hbox_set:Nw { c }
\cs_generate_variant:Nn \hbox_gset:Nw { c }
\cs_new_protected:Npn \hbox_set_end:
\{ \color_group_end: \c_group_end_token \}
\cs_new_eq:NN \hbox_gset_end: \hbox_set_end:
\end{verbatim}

(End definition for \hbox_set:Nw and others. These functions are documented on page 274.)

Combining the above ideas.

\begin{verbatim}
\cs_new_protected:Npm \hbox_set_to_wd:Nnw #1#2
\{ \tex_hbox:D to \__box_dim_eval:n {#1} {#2} \color_group_begin: #2 \color_group_end: \}
\cs_new_protected:Npm \hbox_gset_to_wd:Nnw #1#2
\{ \tex_global:D \tex_setbox:D #1 \tex_hbox:D to \__box_dim_eval:n {#1} {#2} \color_group_begin: \color_group_end: \}
\cs_generate_variant:Nn \hbox_set_to_wd:Nnw { c }
\cs_generate_variant:Nn \hbox_gset_to_wd:Nnw { c }
\end{verbatim}

(End definition for \hbox_set_to_wd:Nnw and \hbox_gset_to_wd:Nnw. These functions are documented on page 274.)

Put a horizontal box directly into the input stream.

\begin{verbatim}
\cs_new_protected:Npm \hbox_to_wd:nn #1#2
\{ \tex_hbox:D to \__box_dim_eval:n {#1} \color_group_begin: #2 \color_group_end: \}
\cs_new_protected:Npm \hbox_to_zero:n #1
\{ \tex_hbox:D to \c_zero_dim \}
\end{verbatim}

(End definition for \hbox_to_wd:nn and \hbox_to_zero:n. These functions are documented on page 274.)
Put a zero-sized box with the contents pushed against one side (which makes it stick out on the other) directly into the input stream.

\hbox_overlap_center:n
\hbox_overlap_left:n
\hbox_overlap_right:n

\hbox_overlap_center:n
\hbox_overlap_left:n
\hbox_overlap_right:n

Unpacking a box and if requested also clear it.

\hbox_unpack:N
\hbox_unpack:c
\hbox_unpack_drop:N
\hbox_unpack_drop:c

84.11 Vertical mode boxes

\vbox:n
The following test files are used for this code: m3box003.lvt.

\vbox_top:n
The following test files are used for this code: m3box003.lvt.

\vbox_to_ht:nn
\vbox_to_zero:n
\vbox_to_ht:nn
\vbox_to_zero:n

Put a vertical box directly into the input stream.

End definition for \hbox_overlap_center:n, \hbox_overlap_left:n, and \hbox_overlap_right:n. These functions are documented on page 274.

End definition for \hbox_unpack:N and \hbox_unpack_drop:N. These functions are documented on page 274.

End definition for \hbox_to_ht:nn and \hbox_to_zero:n. These functions are documented on page 275.
Storing material in a vertical box with a natural height.
\begin{Verbatim}
\cs_new_protected:Npn \vbox_set:Nn #1#2
{ \tex_setbox:D #1 \tex_vbox:D
  \{ \color_group_begin: #2 \par \color_group_end: \}
}
\cs_new_protected:Npn \vbox_gset:Nn #1#2
{ \tex_global:D \tex_setbox:D #1 \tex_vbox:D
  \{ \color_group_begin: #2 \par \color_group_end: \}
}
\cs_generate_variant:Nn \vbox_set:Nn { c }
\cs_generate_variant:Nn \vbox_gset:Nn { c }
\end{Verbatim}

(End definition for \vbox_set:Nn and \vbox_gset:Nn. These functions are documented on page 275.)

Storing material in a vertical box with a natural height and reference point at the baseline of the first object in the box.
\begin{Verbatim}
\cs_new_protected:Npn \vbox_set_top:Nn #1#2
{ \tex_setbox:D #1 \tex_vtop:D
  \{ \color_group_begin: #2 \par \color_group_end: \}
}
\cs_new_protected:Npn \vbox_gset_top:Nn #1#2
{ \tex_global:D \tex_setbox:D #1 \tex_vtop:D
  \{ \color_group_begin: #2 \par \color_group_end: \}
}
\cs_generate_variant:Nn \vbox_set_top:Nn { c }
\cs_generate_variant:Nn \vbox_gset_top:Nn { c }
\end{Verbatim}

(End definition for \vbox_set_top:Nn and \vbox_gset_top:Nn. These functions are documented on page 275.)

Storing material in a vertical box with a specified height.
\begin{Verbatim}
\cs_new_protected:Npn \vbox_set_to_ht:Nnn #1#2#3
{ \tex_setbox:D #1 \tex_vbox:D to \__box_dim_eval:n {#2}
  \{ \color_group_begin: #3 \par \color_group_end: \}
}
\cs_new_protected:Npn \vbox_gset_to_ht:Nnn #1#2#3
{ \tex_global:D \tex_setbox:D #1 \tex_vbox:D to \__box_dim_eval:n {#2}
  \{ \color_group_begin: #3 \par \color_group_end: \}
}
\cs_generate_variant:Nn \vbox_set_to_ht:Nnn { c }
\cs_generate_variant:Nn \vbox_gset_to_ht:Nnn { c }
\end{Verbatim}

(End definition for \vbox_set_to_ht:Nnn and \vbox_gset_to_ht:Nnn. These functions are documented on page 275.)
Storing material in a vertical box. This type is useful in environment definitions.

\begin{verbatim}
\cs_new_protected:Npn \vbox_set:Nw #1
\{
  \tex_setbox:D #1 \tex_vbox:D
  \c_group_begin_token
  \color_group_begin:
\}
\cs_new_protected:Npn \vbox_gset:Nw #1
\{
  \tex_global:D \tex_setbox:D #1 \tex_vbox:D
  \c_group_begin_token
  \color_group_begin:
\}
\cs_generate_variant:Nn \vbox_set:Nw { c }
\cs_generate_variant:Nn \vbox_gset:Nw { c }
\cs_new_protected:Npn \vbox_set_end:
\{
  \par
  \color_group_end:
  \c_group_end_token
\}
\cs_new_eq:NN \vbox_gset_end: \vbox_set_end:
\end{verbatim}

(End definition for \vbox_set:Nw and others. These functions are documented on page 275.)

A combination of the above ideas.

\begin{verbatim}
\cs_new_protected:Npn \vbox_set_to_ht:Nnw #1#2
\{
  \tex_setbox:D #1 \tex_vbox:D to \__box_dim_eval:n {#2}
  \c_group_begin_token
  \color_group_begin:
\}
\cs_new_protected:Npn \vbox_gset_to_ht:Nnw #1#2
\{
  \tex_global:D \tex_setbox:D #1 \tex_vbox:D to \__box_dim_eval:n {#2}
  \c_group_begin_token
  \color_group_begin:
\}
\cs_generate_variant:Nn \vbox_set_to_ht:Nnw { c }
\cs_generate_variant:Nn \vbox_gset_to_ht:Nnw { c }
\end{verbatim}

(End definition for \vbox_set_to_ht:Nnw and \vbox_gset_to_ht:Nnw. These functions are documented on page 276.)

Unpacking a box and if requested also clear it.

\begin{verbatim}
\cs_new_eq:NN \vbox_unpack:N \tex_unvcopy:D
\cs_new_eq:NN \vbox_unpack_drop:N \tex_unvbox:D
\cs_generate_variant:Nn \vbox_unpack:N { c }
\cs_generate_variant:Nn \vbox_unpack_drop:N { c }
\end{verbatim}

(End definition for \vbox_unpack:N and \vbox_unpack_drop:N. These functions are documented on page 276.)
\vbox_set_split_to_ht:NNn
\vbox_set_split_to_ht:cNn
\vbox_set_split_to_ht:Ncn
\vbox_set_split_to_ht:ccn
\vbox_gset_split_to_ht:NNn
\vbox_gset_split_to_ht:cNn
\vbox_gset_split_to_ht:Ncn
\vbox_gset_split_to_ht:ccn

Splitting a vertical box in two.

\cs_new_protected:Npn \vbox_set_split_to_ht:NNn #1#2#3
{ \tex_setbox:D #1 \tex_vsplit:D #2 to \__box_dim_eval:n {#3} }
\cs_generate_variant:Nn \vbox_set_split_to_ht:NNn { c , Nc , cc }

\cs_new_protected:Npn \vbox_gset_split_to_ht:NNn #1#2#3
{ \tex_global:D \tex_setbox:D #1 \tex_vsplit:D #2 to \__box_dim_eval:n {#3} }
\cs_generate_variant:Nn \vbox_gset_split_to_ht:NNn { c , Nc , cc }

(End definition for \vbox_set_split_to_ht:NNn and \vbox_gset_split_to_ht:NNn. These functions are documented on page 276.)

84.12 Affine transformations

\l__box_angle_fp When rotating boxes, the angle itself may be needed by the engine-dependent code. This
is done using the \textbf{fp} module so that the value is tidied up properly.
\fp_new:N \l__box_angle_fp
(End definition for \l__box_angle_fp.)

\l__box_cos_fp \l__box_sin_fp These are used to hold the calculated sine and cosine values while carrying out a rotation.
\fp_new:N \l__box_cos_fp
\fp_new:N \l__box_sin_fp
(End definition for \l__box_cos_fp and \l__box_sin_fp.)

\l__box_top_dim These are the positions of the four edges of a box before manipulation.
\l__box_bottom_dim \l__box_left_dim \l__box_right_dim
\dim_new:N \l__box_top_dim
\dim_new:N \l__box_bottom_dim
\dim_new:N \l__box_left_dim
\dim_new:N \l__box_right_dim
(End definition for \l__box_top_dim and others.)

\l__box_top_new_dim These are the positions of the four edges of a box after manipulation.
\l__box_bottom_new_dim \l__box_left_new_dim \l__box_right_new_dim
\dim_new:N \l__box_top_new_dim
\dim_new:N \l__box_bottom_new_dim
\dim_new:N \l__box_left_new_dim
\dim_new:N \l__box_right_new_dim
(End definition for \l__box_top_new_dim and others.)

\l__box_internal_box Scratch space, but also needed by some parts of the driver.
\box_new:N \l__box_internal_box
(End definition for \l__box_internal_box.)
Figure 1: Co-ordinates of a box prior to rotation.

Rotation of a box starts with working out the relevant sine and cosine. The actual rotation is in an auxiliary to keep the flow slightly clearer:

\begin{verbatim}
\box_rotate:Nn \box_rotate:cn \box_grotate:Nn \box_grotate:cn
\__box_rotate:Nn \__box_rotate:N \__box_rotate_xdir:nn \__box_rotate_ydir:nn
\__box_rotate_quadrant_one: \__box_rotate_quadrant_two:
\__box_rotate_quadrant_three: \__box_rotate_quadrant_four:

\cs_new_protected:Npn \box_rotate:Nn #1#2 { \__box_rotate:NnN #1 {#2} \hbox_set:Nn }
\cs_generate_variant:Nn \box_rotate:Nn { c }
\cs_new_protected:Npn \box_grotate:Nn #1#2 { \__box_rotate:NnN #1 {#2} \hbox_gset:Nn }
\cs_generate_variant:Nn \box_grotate:Nn { c }
\cs_new_protected:Npn \__box_rotate:NnN #1#2#3 {
   #3 #1 {
      \fp_set:Nn \l__box_angle_fp {#2}
      \fp_set:Nn \l__box_sin_fp {sind \l__box_angle_fp}
      \fp_set:Nn \l__box_cos_fp {cosd \l__box_angle_fp}
      \__box_rotate:N #1
   }
}
\end{verbatim}

The edges of the box are then recorded: the left edge is always at zero. Rotation of the four edges then takes place: this is most efficiently done on a quadrant by quadrant basis:

\begin{verbatim}
\cs_new_protected:Npn \__box_rotate:N { \dim_set:Nn \l__box_top_dim { \box_ht:N #1 }
\dim_set:Nn \l__box_bottom_dim { -\box_dp:N #1 }
\dim_set:Nn \l__box_right_dim { \box_wd:N #1 }
\dim_zero:N \l__box_left_dim }
\cs_new_protected:Npm \__box_rotate:Nm #1
\end{verbatim}

The next step is to work out the \( x \) and \( y \) coordinates of vertices of the rotated box in relation to its original coordinates. The box can be visualized with vertices \( B, C, D \) and \( E \) is illustrated (Figure 1). The vertex \( O \) is the reference point on the baseline, and in this implementation is also the centre of rotation. The formulae are, for a point \( P \) and angle \( \alpha \):

\[
\begin{align*}
P'_x &= P_x - O_x \\
P'_y &= P_y - O_y \\
P''_x &= (P'_x \cos(\alpha)) - (P'_y \sin(\alpha)) \\
P''_y &= (P'_x \sin(\alpha)) + (P'_y \cos(\alpha)) \\
P'''_x &= P''_x + O_x + L_x \\
P'''_y &= P''_y + O_y
\end{align*}
\]
The “extra” horizontal translation $L_x$ at the end is calculated so that the leftmost point of the resulting box has $x$-coordinate 0. This is desirable as TeX boxes must have the reference point at the left edge of the box. (As $O$ is always $(0,0)$, this part of the calculation is omitted here.)

```latex
\fp_compare:nNnTF \l__box_sin_fp > \c_zero_fp
\{ \fp_compare:nNnTF \l__box_cos_fp > \c_zero_fp
\{ \__box_rotate_quadrant_one: \}
\{ \__box_rotate_quadrant_two: \}
\}
\}
\fp_compare:nNnTF \l__box_sine_fp < \c_zero_fp
\{ \__box_rotate_quadrant_three: \}
\{ \__box_rotate_quadrant_four: \}
\}
```

The position of the box edges are now known, but the box at this stage be misplaced relative to the current TeX reference point. So the content of the box is moved such that the reference point of the rotated box is in the same place as the original.

```latex
\hbox_set:Nn \l__box_internal_box \{ \box_use:N \#1 \}
\hbox_set:Nn \l__box_internal_box
\{ \__kernel_kern:n \{ -\l__box_left_new_dim \}
\hbox:n
\{ \__box_backend_rotate:Nn
\l__box_internal_box
\l__box_angle_fp
\}
\}
```

Tidy up the size of the box so that the material is actually inside the bounding box. The result can then be used to reset the original box.

```latex
\box_set_ht:Nn \l__box_internal_box \{ \l__box_top_new_dim \}
\box_set_dp:Nn \l__box_internal_box \{ -\l__box_bottom_new_dim \}
\box_set_wd:Nn \l__box_internal_box
\{ \l__box_right_new_dim - \l__box_left_new_dim \}
\box_use_drop:N \l__box_internal_box
```

These functions take a general point $(#1,#2)$ and rotate its location about the origin, using the previously-set sine and cosine values. Each function gives only one component of the location of the updated point. This is because for rotation of a box each step needs only one value, and so performance is gained by avoiding working out both $x'$ and $y'$ at the same time. Contrast this with the equivalent function in the l3coffins module, where both parts are needed.

```latex
\cs_new_protected:Npm \__box_rotate_xdir:n \#1\#2\#3
\{ \dim_set:Nn \#3
\{ \fp_to_dim:n
\{ \l__box_cos_fp * \dim_to_fp:n \#1
- \l__box_sin_fp * \dim_to_fp:n \#2\}
```

1264
Rotation of the edges is done using a different formula for each quadrant. In every case, the top and bottom edges only need the resulting $y$-values, whereas the left and right edges need the $x$-values. Each case is a question of picking out which corner ends up at with the maximum top, bottom, left and right value. Doing this by hand means a lot less calculating and avoids lots of comparisons.
\begin{verbatim}
{ \_box_rotate_ydir:nnN \l__box_left_dim \l__box_top_dim \\
\l__box_top_new_dim \_box_rotate_ydir:nnN \l__box_right_dim \l__box_bottom_dim \\
\l__box_bottom_new_dim \_box_rotate_ydir:nnN \l__box_left_dim \l__box_bottom_dim \\
\l__box_bottom_new_dim \_box_rotate_ydir:nnN \l__box_right_dim \l__box_top_dim \\
\l__box_top_new_dim }

\end{verbatim}

(End definition for \box_rotate:Nn and others. These functions are documented on page 280.)

Scaling is potentially different in the two axes.

\begin{verbatim}
\fp_new:N \l__box_scale_x_fp \\
\fp_new:N \l__box_scale_y_fp
\end{verbatim}

(End definition for \l__box_scale_x_fp and \l__box_scale_y_fp.)

Resizing a box starts by working out the various dimensions of the existing box.

\begin{verbatim}
\cs_new_protected:Npn \box_resize_to_wd_and_ht_plus_dp:Nnn #1#2#3 
\{ \
\l__box_resize_to_wd_and_ht_plus_dp:NnnN #1 {#2} {#3} \\
\hbox_set:Nn 
\}

\cs_generate_variant:Nn \box_resize_to_wd_and_ht_plus_dp:Nnn { c }
\cs_new_protected:Npn \box_gresize_to_wd_and_ht_plus_dp:Nnn #1#2#3 
\{ \
\l__box_resize_to_wd_and_ht_plus_dp:NnnN #1 {#2} {#3} \\
\hbox_gset:Nn 
\}

\cs_generate_variant:Nn \box_gresize_to_wd_and_ht_plus_dp:Nnn { c }
\cs_new_protected:Npn \__box_resize_to_wd_and_ht_plus_dp:NnnN #1#2#3#4 
\{ 
\l__box_resize_set_corners:N #1 \\
\}

\cs_new_protected:Npn \__box_resize_set_corners:N #1 
\{ 
\}
\cs_new_protected:Npn \__box_resize:N #1 
\{ 
\}
\cs_new_protected:Npn \__box_resize:NNN 
\{
\}
\end{verbatim}

The \textit{x}-scaling and resulting box size is easy enough to work out: the dimension is that given as \texttt{#2}, and the scale is simply the new width divided by the old one.

\begin{verbatim}
\fp_set:Nn \l__box_scale_x_fp \\
\{ \dim_to_fp:n {#2} / \dim_to_fp:n { \l__box_right_dim } \}
\end{verbatim}

The \textit{y}-scaling needs both the height and the depth of the current box.

\begin{verbatim}
\fp_set:Nn \l__box_scale_y_fp \\
\{ \\
\dim_to_fp:n {#3} \\
\dim_to_fp:n { \l__box_top_dim - \l__box_bottom_dim } \\
\}
\end{verbatim}

Hand off to the auxiliary which does the rest of the work.

\begin{verbatim}
\cs_new_protected:Npn \__box_resize:N #1 
\{ 
\}
\cs_new_protected:Npn \__box_resize_set_corners:N #1 
\{
\}
\end{verbatim}

1266
With at least one real scaling to do, the next phase is to find the new edge co-ordinates. In the $x$ direction this is relatively easy: just scale the right edge. In the $y$ direction, both dimensions have to be scaled, and this again needs the absolute scale value. Once that is all done, the common resize/rescale code can be employed.

Scaling to a (total) height or to a width is a simplified version of the main resizing operation, with the scale simply copied between the two parts. The internal auxiliary is called using the scaling value twice, as the sign for both parts is needed (as this allows the same internal code to be used as for the general case).

(End definition for \box_resize_to_wd_and_ht_plus_dp:NNn and others. These functions are documented on page \textbf{279}.)
\cs_new_protected:Npn \box_gresize_to_ht_plus_dp:Nn #1 #2
{ \__box_resize_to_ht_plus_dp:NnN #1 {#2} \hbox_gset:Nn }
\cs_generate_variant:Nn \box_gresize_to_ht_plus_dp:Nn { c }
\cs_new_protected:Npn \__box_resize_to_ht_plus_dp:NnN #1 #2 #3
{ #3 #1
{ \__box_resize_set_corners:N #1
\fp_set:Nn \l__box_scale_y_fp
\dim_to_fp:n {#2} / \dim_to_fp:n { \l__box_top_dim - \l__box_bottom_dim }
\fp_set_eq:NN \l__box_scale_x_fp \l__box_scale_y_fp
\__box_resize:N #1
}
\cs_new_protected:Npn \box_resize_to_wd:Nn #1 #2
{ \__box_resize_to_wd:NnnN #1 {#2} \hbox_set:Nn }
\cs_generate_variant:Nn \box_resize_to_wd:Nn { c }
\cs_new_protected:Npn \box_gresize_to_wd:Nnn #1 #2 #3
{ \__box_resize_to_wd:NnnN #1 {#2} {#3} \hbox_gset:Nn }
\cs_generate_variant:Nn \box_gresize_to_wd:Nnn { c }
\cs_new_protected:Npn \__box_resize_to_wd_and_ht:Nnn #1 #2 #3
{ #3 #1
{ \__box_resize_set_corners:N #1
\fp_set:Nn \l__box_scale_x_fp
\dim_to_fp:n {#2} / \dim_to_fp:n { \l__box_right_dim }
\fp_set:Nn \l__box_scale_y_fp
\dim_to_fp:n {#3} / \dim_to_fp:n { \l__box_top_dim }
\__box_resize:N #1
}
\cs_new_protected:Npn \box_resize_to_wd_and_ht:Nnn #1 #2 #3 #4
{ \__box_resize_set_corners:N #1
\fp_set:Nn \l__box_scale_x_fp
\dim_to_fp:n {#2} / \dim_to_fp:n { \l__box_right_dim }
\fp_set:Nn \l__box_scale_y_fp
\dim_to_fp:n {#3} / \dim_to_fp:n { \l__box_top_dim }
\__box_resize:N #1
}
\cs_new_protected:Npn \__box_resize_to_wd_and_ht:NNn #1 #2 #3 #4
{ \__box_resize_set_corners:N #1
\fp_set:Nn \l__box_scale_x_fp
\dim_to_fp:n {#2} / \dim_to_fp:n { \l__box_right_dim }
\fp_set:Nn \l__box_scale_y_fp
\dim_to_fp:n {#3} / \dim_to_fp:n { \l__box_top_dim }
\__box_resize:N #1
}
When scaling a box, setting the scaling itself is easy enough. The new dimensions are also relatively easy to find, allowing only for the need to keep them positive in all cases. Once that is done then after a check for the trivial scaling a hand-off can be made to the common code. The code here is split into two as this allows sharing with the auto-resizing functions.

\begin{verbatim}
\cs_new_protected:Npn \box_scale:Nnn #1#2#3
{ \__box_scale:NnnN #1 {#2} {#3} \hbox_set:Nn }
\cs_generate_variant:Nn \box_scale:Nnn { c }
\cs_new_protected:Npn \box_gscale:Nnn #1#2#3
{ \__box_scale:NnnN #1 {#2} {#3} \hbox_gset:Nn }
\cs_generate_variant:Nn \box_gscale:Nnn { c }
\cs_new_protected:Npn \__box_scale:NnnN #1#2#3#4
{ #4 #1
\fp_set:Nn \l__box_scale_x_fp {#2}
\fp_set:Nn \l__box_scale_y_fp {#3}
\__box_scale:N #1
}
\cs_new_protected:Npn \__box_scale:N #1
{\dim_set:Nn \l__box_top_dim { \box_ht:N #1 }
\dim_set:Nn \l__box_bottom_dim { -\box_dp:N #1 }
\dim_set:Nn \l__box_right_dim { \box_wd:N #1 }
\dim_zero:N \l__box_left_dim
\dim_set:Nn \l__box_top_new_dim { \fp_abs:n { \l__box_scale_y_fp } \l__box_top_dim }
\dim_set:Nn \l__box_bottom_new_dim { \fp_abs:n { \l__box_scale_y_fp } \l__box_bottom_dim }
\dim_set:Nn \l__box_right_new_dim { \fp_abs:n { \l__box_scale_x_fp } \l__box_right_dim }
\__box_resize_common:N #1
}
\end{verbatim}

Although autosizing a box uses dimensions, it has more in common in implementation with scaling. As such, most of the real work here is done elsewhere.

\begin{verbatim}
\cs_new_protected:Npn \box_autosize_to_wd_and_ht:Nnn #1#2#3
{ \__box_autosize:NnnnN #1 #2 #3 \hbox_set:Nn }
\cs_new_protected:Npn \box_gautosize_to_wd_and_ht:Nnn #1#2#3
{ \__box_autosize:NnnnN #1 #2 #3 \hbox_gset:Nn }
\cs_new_protected:Npn \box_autosize_to_wd_and_ht_plus_dp:Nnn #1#2#3
{ \__box_autosize:NnnnN #1 #2 #3 \hbox_set:Nn #1 + \box_dp:N #1 }
\end{verbatim}
The main resize function places its input into a box which start off with zero width, and includes the handles for engine rescaling.

```latex
\__box_resize_common:N
```

The new height and depth can be applied directly.

```latex
\fp_compare:nNnTF \__box_scale_y_fp > \c_zero_fp
{...
\box_set_ht:Nn \__box_internal_box \__box_top_new_dim
\box_set_dp:Nn \__box_internal_box \__box_bottom_new_dim}
```

Things are not quite as obvious for the width, as the reference point needs to remain unchanged. For positive scaling factors resizing the box is all that is needed. However, for case of a negative scaling the material must be shifted such that the reference point ends up in the right place.

```latex
\fp_compare:nNnTF \__box_scale_x_fp < \c_zero_fp
{...
\hbox_to_wd:nn \__box_right_new_dim}
```
\__kernel_kern:n { \l__box_right_new_dim }
\box_use_drop:N \l__box_internal_box
\tex_hss:D
}
\box_set_wd:Nn \l__box_internal_box \l__box_right_new_dim
\hbox:n
\__kernel_kern:n { Opt }
\box_use_drop:N \l__box_internal_box
\tex_hss:D
}

(End definition for \__box_resize_common:N.)

\/package
Chapter 85

\textbf{l3coffins Implementation}

\section{Coffins: data structures and general variables}

Scratch variables.

\begin{verbatim}
\l__coffin_internal_box \l__coffin_internal_dim \l__coffin_internal_tl
\end{verbatim}

\Box_new:N \l__coffin_internal_box
\dim_new:N \l__coffin_internal_dim
\tl_new:N \l__coffin_internal_tl

(The definition for \l__coffin_internal_box, \l__coffin_internal_dim, and \l__coffin_internal_tl.)

The “corners” of a coffin define the real content, as opposed to the \TeX{} bounding box. They all start off in the same place, of course.

\begin{verbatim}
\prop_const_from_keyval:Nn \c__coffin_corners_prop
\end{verbatim}

\begin{verbatim}
{ tl = { 0pt } { 0pt } ,
  tr = { 0pt } { 0pt } ,
  bl = { 0pt } { 0pt } ,
  br = { 0pt } { 0pt } ,
}

(End definition for \c__coffin_corners_prop.)

Pole positions are given for horizontal, vertical and reference-point based values.

\begin{verbatim}
\prop_const_from_keyval:Nn \c__coffin_poles_prop
\end{verbatim}

\begin{verbatim}
{ l = { 0pt } { 0pt } { 0pt } { 1000pt } ,
  hc = { 0pt } { 0pt } { 0pt } { 1000pt } ,
  r = { 0pt } { 0pt } { 0pt } { 1000pt } ,
  b = { 0pt } { 0pt } { 1000pt } { 0pt } ,
  vc = { 0pt } { 0pt } { 1000pt } { 0pt } ,
  t = { 0pt } { 0pt } { 1000pt } { 0pt } ,
  B = { 0pt } { 0pt } { 1000pt } { 0pt } ,
  H = { 0pt } { 0pt } { 1000pt } { 0pt } ,
  T = { 0pt } { 0pt } { 1000pt } { 0pt } ,
}

\end{verbatim}
There are a number of basic functions needed for creating coffins and placing material in them. This all relies on the following data structures.

\__coffin_to_value:N

Coffins are a two-part structure and we rely on the internal nature of box allocation to make everything work. As such, we need an interface to turn coffin identifiers into numbers. For the purposes here, the signature allowed is N despite the nature of the underlying primitive.

\cs_new_eq:NN \__coffin_to_value:N \tex_number:D
Several of the higher-level coffin functions would give multiple errors if the coffin does not exist. A cleaner way to handle this is provided here: both the box and the coffin structure are checked.

\begin{verbatim}
\prg_new_conditional:Npnn \coffin_if_exist:N #1 { p, T, F, TF }
  { \cs_if_exist:NTF #1 }
    { \cs_if_exist:cTF { coffin \_coffin_to_value:N #1 - poles }
      { \prg_return_true: }
      { \prg_return_false: }
    }
  { \prg_return_true: }
\prg_generate_conditional_variant:Nnn \coffin_if_exist:N
  { c } { p, T, F, TF }
\end{verbatim}

(End definition for \coffin_if_exist:NTF. This function is documented on page 281.)

\begin{verbatim}
\cs_new_protected:Npm \_coffin_if_exist:NT #1#2
  { \cs_if_exist:NTF #1 }
    { \#2 }
    { \msg_error:nxn { coffin } { unknown } }
  { \token_to_str:N #1 }
\end{verbatim}

(End definition for \_coffin_if_exist:NT.)

Clearing coffins means emptying the box and resetting all of the structures.

\begin{verbatim}
\cs_new_protected:Npm \coffin_clear:N #1
  \_coffin_if_exist:NT #1
  { \box_clear:N #1
    \_coffin_reset_structure:N #1 }
\cs_generate_variant:Nn \coffin_clear:N { c }
\cs_new_protected:Npm \coffin_gclear:N #1
  \_coffin_if_exist:NT #1
  { \box_gclear:N #1
    \_coffin_greset_structure:N #1 }
\cs_generate_variant:Nn \coffin_gclear:N { c }
\end{verbatim}

(End definition for \coffin_clear:N and \coffin_gclear:N. These functions are documented on page 281.)
Creating a new coffin means making the underlying box and adding the data structures. The \debug_suspend: and \debug_resume: functions prevent \prop_gclear_new:c from writing useless information to the log file.

\cs_new_protected:Npn \coffin_new:N #1
\begin{verbatim}
{ \box_new:N #1
  \debug_suspend:
  \prop_gclear_new:c \{ coffin - \__coffin_to_value:N #1 - corners \}
  \prop_gclear_new:c \{ coffin - \__coffin_to_value:N #1 - poles \}
  \c__coffin_corners_prop
  \prop_gset_eq:cN \{ coffin - \__coffin_to_value:N #1 - poles \}
  \c__coffin_poles_prop
  \debug_resume:
}
\end{verbatim}

\cs_generate_variant:Nn \coffin_new:N { c }

(End definition for \coffin_new:N. This function is documented on page 281.)

\hcoffin_set:Nn \hcoffin_set:cn \hcoffin_gset:Nn \hcoffin_gset:cn
\begin{verbatim}
Horizontal coffins are relatively easy: set the appropriate box, reset the structures then update the handle positions.
\cs_new_protected:Npm \hcoffin_set:Nn #1#2
\begin{verbatim}
{ \__coffin_if_exist:NT #1
  \hbox_set:Nn #1
  { \color_ensure_current:
    #2
  }
  \__coffin_update:N #1
}
\end{verbatim}
\cs_generate_variant:Nn \hcoffin_set:Nn { c }
\cs_new_protected:Npm \hcoffin_gset:Nn #1#2
\begin{verbatim}
{ \__coffin_if_exist:NT #1
  \hbox_gset:Nn #1
  { \color_ensure_current:
    #2
  }
  \__coffin_gupdate:N #1
}
\end{verbatim}
\cs_generate_variant:Nn \hcoffin_gset:Nn { c }

(End definition for \hcoffin_set:Nn and \hcoffin_gset:Nn. These functions are documented on page 282.)

\vcoffin_set:Nnn \vcoffin_set:cnn \vcoffin_gset:Nnn \vcoffin_gset:cnn
\__coffin_set_vertical:NnnNN \__coffin_set_vertical_aux:
\begin{verbatim}
Setting vertical coffins is more complex. First, the material is typeset with a given width. The default handles and poles are set as for a horizontal coffin, before finding the top baseline using a temporary box. No \color_ensure_current: here as that would add a
\end{verbatim}

1275
whatsit to the start of the vertical box and mess up the location of the T pole (see \TeX \ by \ Topic \ for discussion of the \texttt{vtop} primitive, used to do the measuring).

\begin{verbatim}
\cs_new_protected:Npn \vcoffin_set:Nnn #1#2#3
{ \__coffin_set_vertical:NnnNN #1 {#2} {#3} \vbox_set:Nn \__coffin_update:N }
\cs_generate_variant:Nn \vcoffin_set:Nnn { c }
\cs_new_protected:Npn \vcoffin_gset:Nnn #1#2#3
{ \__coffin_set_vertical:NnnNN #1 {#2} {#3} \vbox_gset:Nn \__coffin_gupdate:N }
\cs_generate_variant:Nn \vcoffin_gset:Nnn { c }
\cs_new_protected:Npx \__coffin_set_vertical_aux:
{ \bool_lazy_and:nnT { \cs_if_exist_p:N \fmtname } { \str_if_eq_p:Vn \fmtname { LaTeX2e } } { \dim_set_eq:NN \exp_not:N \linewidth \tex_hsize:D } { \dim_set_eq:NN \exp_not:N \columnwidth \tex_hsize:D } }
\cs_new_protected:Npx \__coffin_set:pNnnNN #1#2#3#4#5
{ \__coffin_set_vertical:NnnNN #1 {#2} {#3} \vbox_set_top:Nn \l__coffin_internal_box { \vbox_unpack:N #1 } \__coffin_set_pole:Nnx #1 { T } { \dim_eval:n { \box_ht:N #1 - \box_ht:N \l__coffin_internal_box } } { 1000pt } { 0pt } { \box_clear:N \l__coffin_internal_box } }\end{verbatim}

(End definition for \texttt{vcoffin_set:Nnn} and others. These functions are documented on page 282.)

\texttt{\hcoffin_set:Nw} \hcoffin_set:cw \hcoffin_gset:Nw \hcoffin_gset:cw \hcoffin_set_end: \hcoffin_gset_end:

These are the “begin”/“end” versions of the above: watch the grouping!

\begin{verbatim}
\cs_new_protected:Npn \hcoffin_set:Nw #1
{ \bool_lazy_and:nnTT { \cs_if_exist_p:N \fmtname } { \str_if_eq_p:Vn \fmtname { LaTeX2e } } { \dim_set_eq:NN \exp_not:N \linewidth \tex_hsize:D } { \dim_set_eq:NN \exp_not:N \columnwidth \tex_hsize:D } }
\cs_new_protected:Npx \hcoffin_set_end: \hcoffin_gset_end:
\end{verbatim}

1276
The same for vertical coffins.

\vcoffin_set:Nnw \
\vcoffin_set:cnw \
\vcoffin_gset:Nnw \
\vcoffin_gset:cnw 
\vcoffin_set_end: \
\vcoffin_gset_end: 

(End definition for \hcoffin_set:Nw and others. These functions are documented on page 282.)
Setting two coffins equal is just a wrapper around other functions.

\cs_new_protected:Npn \vcoffin_set_end: { }
\cs_new_protected:Npn \vcoffin_gset_end: { }

(End definition for \vcoffin_set:Nnw and others. These functions are documented on page 282.)

\cs_generate_variant:Nn \coffin_set_eq:NN { c , Nc , cc }
\cs_generate_variant:Nn \coffin_gset_eq:NN { c , Nc , cc }

(End definition for \coffin_set_eq:NN and \coffin_gset_eq:NN. These functions are documented on page 281.)

Special coffins: these cannot be set up earlier as they need \coffin_new:N. The empty coffin is set as a box as the full coffin-setting system needs some material which is not yet available. The empty coffin is created entirely by hand: not everything is in place yet.
85.3 Measuring coffins

Coffins are just boxes when it comes to measurement. However, semantically a separate set of functions are required.

85.4 Coffins: handle and pole management

A simple wrapper around the recovery of a coffin pole, with some error checking and recovery built-in.

Resetting the structure is a simple copy job.
Setting the pole of a coffin at the user/designer level requires a bit more care. The idea here is to provide a reasonable interface to the system, then to do the setting with full expansion. The three-argument version is used internally to do a direct setting.
Simple shortcuts.

Updating the corners of a coffin is straightforward as at this stage there can be no rotation. So the corners of the content are just those of the underlying TeX box.
This function is called when a coffin is set, and updates the poles to reflect the nature of size of the box. Thus this function only alters poles where the default position is dependent on the size of the box. It also does not set poles which are relevant only to vertical coffins.

```latex
\cs_new_protected:Npn \__coffin_update_poles:N #1
\cs_new_protected:Npn \__coffin_gupdate_poles:N #1
\cs_new_protected:Npn \__coffin_update_poles:NN #1#2
\cs_new_protected:Npn \__coffin_update_poles:NNN #1#2#3
```

(End definition for \__coffin_update_corners:N and others.)
85.5 Coffins: calculation of pole intersections

The lead off in finding intersections is to recover the two poles and then hand off to the auxiliary for the actual calculation. There may of course not be an intersection, for which an error trap is needed.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_calculate_intersection:Nnn #1#2#3
\__coffin_get_pole:NnN #1 {#2} \l__coffin_pole_a_tl
\__coffin_get_pole:NnN #1 {#3} \l__coffin_pole_b_tl
\bool_set_false:N \l__coffin_error_bool
\exp_last_two_unbraced:Noo
\__coffin_calculate_intersection:nnnnnnnn
\l__coffin_pole_a_tl \l__coffin_pole_b_tl
\bool_if:NT \l__coffin_error_bool
\msg_error:nn { coffin } { no-pole-intersection }
\dim_zero:N \l__coffin_x_dim
\dim_zero:N \l__coffin_y_dim
\end{verbatim}

The two poles passed here each have four values (as dimensions), \((a, b, c, d)\) and \((a', b', c', d')\). These are arguments 1–4 and 5–8, respectively. In both cases \(a\) and \(b\) are the co-ordinates of a point on the pole and \(c\) and \(d\) define the direction of the pole. Finding the intersection depends on the directions of the poles, which are given by \(d/c\) and \(d'/c'\). However, if one of the poles is either horizontal or vertical then one or more of \(c\), \(d\), \(c'\) and \(d'\) are zero and a special case is needed.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_calculate_intersection:nnnnnnnn
#1#2#3#4#5#6#7#8
\dim_compare:nNnTF {#3} = \c_zero_dim
The case where the first pole is vertical. So the \(x\)-component of the interaction is at \(a\). There is then a test on the second pole: if it is also vertical then there is an error.
\dim_set:Nn \l__coffin_x_dim {#1}
\dim_compare:nNnTF {#7} = \c_zero_dim
{ \bool_set_true:N \l__coffin_error_bool }
The second pole may still be horizontal, in which case the \(y\)-component of the intersection is \(b'\). If not,
\[ y = \frac{d'}{c'} (a - a') + b' \]
with the \(x\)-component already known to be \#1.
\dim_set:Nn \l__coffin_y_dim
{ \dim_compare:nNnTF {#8} = \c_zero_dim
{#6}
{ \fp_to_dim:n
{ \dim_to_fp:n {#8} / \dim_to_fp:n {#7} }
\times \dim_to_fp:n {#1} - \dim_to_fp:n {#5} }
\end{verbatim}

1283
If the first pole is not vertical then it may be horizontal. If so, then the procedure is essentially the same as that already done but with the \( x \)- and \( y \)-components interchanged.

\[
\dim_compare:nNnTF {#4} = \c_zero_dim
\]
\[
\dim_set:Nn \l__coffin_y_dim {#2}
\dim_compare:nNnTF {#8} = \c_zero_dim
\bool_set_true:N \l__coffin_error_bool
\]

Now we deal with the case where the second pole may be vertical, or if not we have

\[
x = \frac{c'}{d'} (b - b') + a'
\]

which is again handled by the same auxiliary.

\[
\dim_set:Nn \l__coffin_x_dim
\dim_compare:nNnTF {#7} = \c_zero_dim
\fp_to_dim:n
\]

The first pole is neither horizontal nor vertical. To avoid even more complexity, we now work out both slopes and pass to an auxiliary.

\[
\use:x
\]
\[
\__coffin_calculate_intersection:nnnnn
\]

Assuming the two poles are not parallel, then the intersection point is found in two steps. First we find the \( x \)-value with

\[
x = \frac{sa - s'a' - b + b'}{s - s'}
\]
and then finding the $y$-value with

$$y = s(x - a) + b$$

(End definition for \cs_set_protected:Npn \__coffin_calculate_intersection:nnnnnn #1#2#3#4#5#6, \__coffin_calculate_intersection:nnnnnnnn, and \__coffin_calculate_intersection:nnnnnn.)

85.6 Affine transformations

\l__coffin_sin_fp \l__coffin_cos_fp

Used for rotations to get the sine and cosine values.

(End definition for \l__coffin_sin_fp and \l__coffin_cos_fp.)

\l__coffin_bounding_prop

A property list for the bounding box of a coffin. This is only needed during the rotation, so there is just the one.

(End definition for \l__coffin_bounding_prop.)

\l__coffin_corners_prop \l__coffin_poles_prop

Used to avoid needing to track scope for intermediate steps.

(End definition for \l__coffin_corners_prop and \l__coffin_poles_prop.)
The shift of the bounding box of a coffin from the real content.

These are used to hold maxima for the various corner values: these thus define the minimum size of the bounding box after rotation.

Rotating a coffin requires several steps which can be conveniently run together. The sine and cosine of the angle in degrees are computed. This is then used to set \( \l__\text{coffin}_{-}\text{sin}\_fp \) and \( \l__\text{coffin}_{-}\text{cos}\_fp \), which are carried through unchanged for the rest of the procedure.

Use a local copy of the property lists to avoid needing to pass the name and scope around.

The corners and poles of the coffin can now be rotated around the origin. This is best achieved using mapping functions.

The bounding box of the coffin needs to be rotated, and to do this the corners have to be found first. They are then rotated in the same way as the corners of the coffin material itself.

At this stage, there needs to be a calculation to find where the corners of the content and the box itself will end up.
The correction of the box position itself takes place here. The idea is that the bounding box for a coffin is tight up to the content, and has the reference point at the bottom-left. The x-direction is handled by moving the content by the difference in the positions of the bounding box and the content left edge. The y-direction is dealt with by moving the box down by any depth it has acquired. The internal box is used here to allow for the next step.

```latex
\hbox_set:Nn \l__coffin_internal_box
\{ \_kernel_kern:n
\{ \_coffin_bounding_shift_dim - \_coffin_left_corner_dim \}
\box_move_down:n \{ \_coffin_bottom_corner_dim \}
\{ \box_use:N \#1 \}
\}
```

If there have been any previous rotations then the size of the bounding box will be bigger than the contents. This can be corrected easily by setting the size of the box to the height and width of the content. As this operation requires setting box dimensions and these transcend grouping, the safe way to do this is to use the internal box and to reset the result into the target box.

```latex
\box_set_ht:Nn \l__coffin_internal_box
\{ \_coffin_top_corner_dim - \_coffin_bottom_corner_dim \}
\box_set_dp:Nn \l__coffin_internal_box { 0pt }
\box_set_wd:Nn \l__coffin_internal_box
\{ \_coffin_right_corner_dim - \_coffin_left_corner_dim \}
\#5 \#1 \{ \box_use_drop:N \l__coffin_internal_box \}
```

The final task is to move the poles and corners such that they are back in alignment with the box reference point.

```latex
\prop_map_inline:Nn \l__coffin_corners_prop
\{ \_coffin_shift_corner:Nnnn \#1 \{##1\} ##2 \}
\prop_map_inline:Nn \l__coffin_poles_prop
\{ \_coffin_shift_pole:Nnnnnn \#1 \{##1\} ##2 \}
```

Update the coffin data.

```latex
#4 \{ coffin - \_coffin_to_value:N \#1 - corners \}
\l__coffin_corners_prop
#4 \{ coffin - \_coffin_to_value:N \#1 - poles \}
\l__coffin_poles_prop
```

(End definition for \coffin_rotate:Nn, \coffin_grotate:Nn, and \_coffin_rotate:NNNN. These functions are documented on page 283.)

\_coffin_set_bounding:N

The bounding box corners for a coffin are easy enough to find: this is the same code as for the corners of the material itself, but using a dedicated property list.

```latex
\cs_new_protected:Npn \__coffin_set_bounding:N #1
\{ \prop_put:Nnx \l__coffin_bounding_prop { tl } \{ 0pt \} \{ \dim_eval:n \{ \box_ht:N \#1 \} \}
\prop_put:Nnx \l__coffin_bounding_prop { tr } \{ \dim_eval:n \{ \box_wd:N \#1 \} \}
```

1287
Rotating the position of the corner of the coffin is just a case of treating this as a vector from the reference point. The same treatment is used for the corners of the material itself and the bounding box.

\__coffin_rotate_bounding:nnn
\__coffin_rotate_corner:Nnnn

Rotating a single pole simply means shifting the co-ordinate of the pole and its direction. The rotation here is about the bottom-left corner of the coffin.

\__coffin_rotate_pole:Nnnnnn

A rotation function, which needs only an input vector (as dimensions) and an output space. The values \l__coffin_cos_fp and \l__coffin_sin_fp should previously have been set up correctly. Working this way means that the floating point work is kept to a minimum: for any given rotation the sin and cosine values do no change, after all.
\begin{verbatim}
\dim_to_fp:n \dim_to_fp:n \dim_to_fp:n
\dim_set:Nn \dim_set:Nn \dim_set:Nn
\xprop_map_inline:Nn \xprop_map_inline:Nn \xprop_map_inline:Nn
\end{verbatim}

The idea here is to find the extremities of the content of the coffin. This is done by looking for the smallest values for the bottom and left corners, and the largest values for the top and right corners. The values start at the maximum dimensions so that the case where all are positive or all are negative works out correctly.

\begin{verbatim}
\__coffin_find_bounding_shift: \__coffin_find_bounding_shift_aux:nn
\end{verbatim}

The approach to finding the shift for the bounding box is similar to that for the corners. However, there is only one value needed here and a fixed input property list, so things are a bit clearer.
Shifting the corners and poles of a coffin means subtracting the appropriate values from the $x$- and $y$-components. For the poles, this means that the direction vector is unchanged.

When scaling, the values given have to be turned into absolute values.

Resizing a coffin begins by setting up the user-friendly names for the dimensions of the coffin box. The new sizes are then turned into scale factor. This is the same operation as takes place for the underlying box, but that operation is grouped and so the same calculation is done here.
The poles and corners of the coffin are scaled to the appropriate places before actually resizing the underlying box.

\[ \text{\texttt{\_\_coffin_resize_common:NnnN}} \]

\texttt{\_\_coffin_resize_common:NnnN} for scaling, the opposite calculation is done to find the new dimensions for the coffin. Only the total height is needed, as this is the shift required for corners and poles. The
scaling is done the TeX way as this works properly with floating point values without needing to use the \texttt{fp} module.

\begin{verbatim}
\cs_new_protected:Npn \coffin_scale:Nnn #1#2#3
\exp_args:NNV \box_scale:Nnn \prop_set_eq:cN \__coffin_scale:NnnNN #1 {#2} {#3}
\cs_generate_variant:Nn \coffin_scale:Nnn { c }
\cs_new_protected:Npn \coffin_gscale:Nnn #1#2#3
\exp_args:NNV \box_gscale:Nnn \prop_gset_eq:cN \__coffin_scale:NnnNN #1 {#2} {#3}
\cs_generate_variant:Nn \coffin_gscale:Nnn { c }
\cs_new_protected:Npn \__coffin_scale:NnnNN #1#2#3#4#5
\fp_set:Nn \l__coffin_scale_x_fp {#2}
\fp_set:Nn \l__coffin_scale_y_fp {#3}
#4 \l__coffin_scale_x_fp \l__coffin_scale_y_fp
\dim_set:Nn \l__coffin_internal_dim \coffin_ht:N #1 + \coffin_dp:N #1
\dim_set:Nn \l__coffin_scaled_total_height_dim \fp_abs:n { \l__coffin_scale_y_fp \l__coffin_internal_dim }
\dim_set:Nn \l__coffin_scaled_width_dim \fp_abs:n { \l__coffin_scale_x_fp \coffin_wd:N #1}
\__coffin_resize_common:NnnN #1 \l__coffin_scaled_width_dim \l__coffin_scaled_total_height_dim #5
\end{verbatim}

(End definition for \texttt{coffin_scale:Nnn}, \texttt{coffin_gscale:Nnn}, and \texttt{coffin_scale:NnnNN}. These functions are documented on page 283.)

\subsection{\texttt{\_coffin_scale_vector:nnNN}}

This function scales a vector from the origin using the pre-set scale factors in \textit{x} and \textit{y}. This is a much less complex operation than rotation, and as a result the code is a lot clearer.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_scale_vector:nnNN #1#2#3#4
\dim_set:Nn #3 \fp_to_dim:n { \dim_to_fp:n {#1} * \l__coffin_scale_x_fp }
\dim_set:Nn #4 \fp_to_dim:n { \dim_to_fp:n {#2} * \l__coffin_scale_y_fp }
\end{verbatim}

(End definition for \texttt{\_coffin_scale_vector:nnNN}.)

\subsection{\texttt{\_coffin_scale_corner:Nnnn}}

Scaling both corners and poles is a simple calculation using the preceding vector scaling.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_scale_corner:Nnnn #1#2#3#4
\__coffin_scale_vector:nnNN \l__coffin_x_dim \l__coffin_y_dim
\prop_put:Nnx \l__coffin_corners_prop \l__coffin_x_dim \l__coffin_y_dim
\end{verbatim}

\subsection{\texttt{\_coffin_scale_pole:Nnnnnn}}

(End definition for \texttt{\_coffin_scale_corner:Nnnn}.)
These functions correct for the $x$ displacement that takes place with a negative horizontal scaling.

These commands join two coffins, using a horizontal and vertical pole from each coffin and making an offset between the two. The result is stored as the as a third coffin, which has all of its handles reset to standard values. First, the more basic alignment function is used to get things started.

```
\cs_new_protected:Npn \__coffin_join:NnnNnnnn #1#2#3#4#5#6#7#8
\__coffin_align:NnnNnnnn #1 {#2} {#3} #4 {#5} {#6} {#7} {#8} \l__coffin_aligned_coffin
```

Correct the placement of the reference point. If the $x$-offset is negative then the reference point of the second box is to the left of that of the first, which is corrected using a kern. On the right side the first box might stick out, which would show up if it is wider than the sum of the $x$-offset and the width of the second box. So a second kern may be needed.

```
\hbox_set:Nn \l__coffin_aligned_coffin
```

85.7 Aligning and typesetting of coffins
The coffin structure is reset, and the corners are cleared: only those from the two parent coffins are needed.

\__coffin_reset_structure:N \l__coffin_aligned_coffin
\prop_clear:c
{ coffin ~ \__coffin_to_value:N \l__coffin_aligned_coffin
 \c_space_tl corners }
\__coffin_update_poles:N \l__coffin_aligned_coffin

The structures of the parent coffins are now transferred to the new coffin, which requires that the appropriate offsets are applied. That then depends on whether any shift was needed.

\dim_compare:nNnTF \l__coffin_offset_x_dim < \c_zero_dim
{ \__coffin_offset_poles:Nnn \l__coffin_offset_poles:NNn \l__coffin_offset_corners:NNn }
{ \__coffin_offset_poles:NNn \l__coffin_offset_poles:NNn \l__coffin_offset_corners:NNn }
\__coffin_update_vertical_poles:NNN \l__coffin_aligned_coffin
#9 \l__coffin_aligned_coffin

(End definition for \coffin_join:NnnNnnnn, \coffin_gjoin:NnnNnnnn, and \__coffin_join:NnnNnnnnN. These functions are documented on page 284.)

A more simple version of the above, as it simply uses the size of the first coffin for the new one. This means that the work here is rather simplified compared to the above code. The function used when marking a position is here also as it is similar but without the structure updates.

\cs_new_protected:Npn \coffin_attach:NnnNnnnn #1#2#3#4#5#6#7#8
{ \__coffin_attach:NnnNnnnnN #1 {#2} {#3} #4 {#5} {#6} {#7} {#8} }
\cs_generate_variant:Nn \coffin_attach:NnnNnnnn { c , Nnnc , cnnc }
The internal function aligns the two coffins into a third one, but performs no corrections on the resulting coffin poles. The process begins by finding the points of intersection for the poles for each of the input coffins. Those for the first coffin are worked out after those for the second coffin, as this allows the ‘primed’ storage area to be used for the second coffin. The ‘real’ box offsets are then calculated, before using these to re-box the input coffins. The default poles are then set up, but the final result depends on how the bounding box is being handled.
Transferring structures from one coffin to another requires that the positions are updated by the offset between the two coffins. This is done by mapping over the property list of the source coffins, moving as appropriate and saving to the new coffin data structures. The test for a - means that the structures from the parent coffins are uniquely labelled and do not depend on the order of alignment. The pay off for this is that - should not be used in coffin pole or handle names, and that multiple alignments do not result in a whole set of values.

Saving the offset corners of a coffin is very similar, except that there is no need to worry about naming: every corner can be saved here as order is unimportant.
The T and B poles need to be recalculated after alignment. These functions find the larger absolute value for the poles, but this is of course only logical when the poles are horizontal.

```latex
\cs_new_protected:Npn \___coffin_update_vertical_poles:NNN \#1\#2\#3
\begin{verbatim}
{ %_coffin_get_pole:NnN \#3 \{ \#1 -T \} \l__coffin_pole_a_tl
 %_coffin_get_pole:NnN \#3 \{ \#2 -T \} \l__coffin_pole_b_tl
 \exp_last_two_unbraced:Noo \___coffin_update_T:nnnnnnnnN\l__coffin_pole_a_tl \l__coffin_pole_b_tl \#3

 %_coffin_get_pole:NnN \#3 \{ \#1 -B \} \l__coffin_pole_a_tl
 %_coffin_get_pole:NnN \#3 \{ \#2 -B \} \l__coffin_pole_b_tl
 \exp_last_two_unbraced:Noo \___coffin_update_B:nnnnnnnnN\l__coffin_pole_a_tl \l__coffin_pole_b_tl #3
\end{verbatim}
\end{verbatim}
```

(End definition for \___coffin_offset_corners:NN and \___coffin_offset_corner:Nnnnn.)
An empty-but-horizontal coffin.

(End definition for \c__coffin_empty_coffin.)

Typesetting a coffin means aligning it with the current position, which is done using a coffin with no content at all. As well as aligning to the empty coffin, there is also a need to leave vertical mode, if necessary.

(End definition for \coffin_typeset:Nnnnn. This function is documented on page 284.)

Used for printing coffins with data structures attached.

(End definition for \l__coffin_display_coffin, \l__coffin_display_coord_coffin, and \l__coffin_display_pole_coffin.)

This property list is used to print coffin handles at suitable positions. The offsets are expressed as multiples of the basic offset value, which therefore acts as a scale-factor.

(End definition for \coffin_typeset:cnnnn.)
\prop_put:Nnn \l__coffin_display_handles_prop { Thc } { { t } { hc } { 0 } { -1 } }
\prop_put:Nnn \l__coffin_display_handles_prop { Tr } { { t } { 1 } { 1 } { -1 } }
\prop_put:Nnn \l__coffin_display_handles_prop { Hl } { { vc } { r } { 1 } { -1 } }
\prop_put:Nnn \l__coffin_display_handles_prop { Hhc } { { vc } { hc } { 0 } { 1 } }
\prop_put:Nnn \l__coffin_display_handles_prop { Hr } { { vc } { 1 } { 1 } { 1 } }
\prop_put:Nnn \l__coffin_display_handles_prop { Bl } { { b } { r } { -1 } { -1 } }
\prop_put:Nnn \l__coffin_display_handles_prop { Bhc } { { b } { hc } { 0 } { -1 } }
\prop_put:Nnn \l__coffin_display_handles_prop { Br } { { b } { 1 } { 1 } { -1 } }

(End definition for \l__coffin_display_handles_prop.)

\l__coffin_display_offset_dim
The standard offset for the label from the handle position when displaying handles.
\dim_new:N \l__coffin_display_offset_dim
\dim_set:Nn \l__coffin_display_offset_dim { 2pt }

(End definition for \l__coffin_display_offset_dim.)

\l__coffin_display_x_dim \l__coffin_display_y_dim
As the intersections of poles have to be calculated to find which ones to print, there is a
need to save the intersection into two dedicated values.
\dim_new:N \l__coffin_display_x_dim
\dim_new:N \l__coffin_display_y_dim

(End definition for \l__coffin_display_x_dim and \l__coffin_display_y_dim.)

\l__coffin_display_poles_prop
A property list for printing poles: various things need to be deleted from this to get a
"nice" output.
\prop_new:N \l__coffin_display_poles_prop

(End definition for \l__coffin_display_poles_prop.)

\l__coffin_display_font_tl
Stores the settings used to print coffin data: this keeps things flexible.
\tl_new:N \l__coffin_display_font_tl
\bool_lazy_and:nnT { \cs_if_exist_p:N \fmtname } { \str_if_eq_p:Vn \fmtname { LaTeX2e } }
\tl_set:Nn \l__coffin_display_font_tl { \sffamily \tiny }

(End definition for \l__coffin_display_font_tl.)

\__coffin_rule:nn
Abstract out creation of rules here until there is a higher-level interface.
\cs_new_protected:Npn \__coffin_rule:nn #1#2
{ \mode_leave_vertical:
\hbox:n { \tex_vrule:D width #1 height #2 \scan_stop: } }

1299
Marking a single handle is relatively easy. The standard attachment function is used, meaning that there are two calculations for the location. However, this is likely to be okay given the load expected. Contrast with the more optimised version for showing all handles which comes next.

```latex
\cs_new_protected:Npn \coffin_mark_handle:Nnnn #1#2#3#4
\hcoffin_set:Nn \l__coffin_display_pole_coffin
\color_select:n {#4}
\__coffin_rule:nn { 1pt } { 1pt }
\__coffin_attach_mark:NnnNnnnn #1 {#2} {#3}
\l__coffin_display_pole_coffin { hc } { vc } { Opt } { Opt }
\hcoffin_set:Nn \l__coffin_display_coord_coffin
\color_select:n {#4}
\l__coffin_display_font_tl \tl_to_str:n { #2 , #3 } \prop_get:NnN \l__coffin_display_handles_prop { #2 #3 } \l__coffin_internal_tl\quark_if_no_value:NTF \l__coffin_internal_tl
\__coffin_attach_mark:NnnNnnnn #1 {#2} {#3}
\l__coffin_display_coord_coffin \l__coffin_display_offset_dim
\exp_last_unbraced:No \__coffin_mark_handle_aux:nnnnNnn
\l__coffin_internal_tl #1 {#2} {#3}
\exp_last_unbraced:No \__coffin_mark_handle_aux:nnnnNnn
\l__coffin_internal_tl #1 {#2} {#3}
\cs_new_protected:Npn \__coffin_mark_handle_aux:nnnnNnnn #1#2#3#4#5#6#7
\__coffin_attach_mark:NnnNnnnn #5 {#6} {#7}
\l__coffin_display_coord_coffin {#1} {#2}
\l__coffin_display_offset_dim \l__coffin_display_offset_dim
\cs_generate_variant:Nn \coffin_mark_handle:Nnnn { c }
```

(End definition for \_coffin_rule:nn.)

(End definition for \coffin_mark_handle:Nnnn and \__coffin_mark_handle_aux:nnnnNnnn. This function is documented on page 285.)
Printing the poles starts by removing any duplicates, for which the H poles is used as 
the definitive version for the baseline and bottom. Two loops are then used to find the 
combinations of handles for all of these poles. This is done such that poles are removed 
during the loops to avoid duplication.

For each pole there is a check for an intersection, which here does not give an error if 
one is found. The successful values are stored and used to align the pole coffin with the 
main coffin for output. The positions are recovered from the preset list if available.

This is a dedicated version of \texttt{\coffin_attach:Nnnnn} with a hard-wired first coffin. As the intersection is already known and stored for the display coffin the code simply uses it directly, with no calculation.
For showing the various internal structures attached to a coffin in a way that keeps things relatively readable. If there is no apparent structure then the code complains.

```latex
\cs_new_protected:Npn \coffin_show_structure:N #1 \{ \__coffin_if_exist:NT #1 \{ \coffin \{ show \} \{ \token_to_str:N \#2 \} \{ \iow_newline: >- \dim_eval:n \{ \coffin_ht:N \#2 \} \iow_newline: >- \dim_eval:n \{ \coffin_dp:N \#2 \} \iow_newline: >- \dim_eval:n \{ \coffin_wd:N \#2 \} \} \{ \prop_map_function:cN \{ coffin \{ \__coffin_to_value:N \#2 \} \{ poles \} \msg_show_item_unbraced:nn \} \msg_term:nnxxxx \box_show:Nnn \} \}
```

(End definition for \coffin_show_structure:N, \coffin_log_structure:N, and \__coffin_show_structure:NN. These functions are documented on page 285.)

Essentially a combination of \coffin_show_structure:N and \box_show:Nnn, but we need to avoid having two prompts, so we use \msg_term:nxxxxx in the show case.
85.9 Messages

\msg_new:nnnn { coffin } { no-pole-intersection } { No-intersection-between-coffin-poles. }
\msg_new:nnnn { coffin } { unknown } { Unknown-coffin-#{1}. } { The-coffin-#{1}-was-never-defined. }
\msg_new:nnnn { coffin } { unknown-pole } { Pole-#{1}-unknown-for-coffin-#{2}. } { LaTeX-was-asked-to-find-a-typesetting-pole-for-a-coffin,- but-either-the-coffin-does-not-exist-or-the-pole-name-is-wrong. }
\msg_new:nnn { coffin } { show } { Size-of-coffin-#{1} : #2 \ \ Poles-of-coffin-#{1} : #3 . }

(End definition for \coffin_show:N and others. These functions are documented on page 285.)
Chapter 86

\texttt{l3color} Implementation

\section*{86.1 Basics}

\l__color_current_tl The color currently active for foreground (text, \textit{etc.}) material. This is stored in the form of a color model followed by one or more values. There are four pre-defined models, three of which take numerical values in the range $[0, 1]$:

- \texttt{gray} \langle gray \rangle Grayscale color with the \langle gray \rangle value running from 0 (fully black) to 1 (fully white)
- \texttt{cmyk} \langle cyan \rangle \langle magenta \rangle \langle yellow \rangle \langle black \rangle
- \texttt{rgb} \langle red \rangle \langle green \rangle \langle blue \rangle

Notice that the value are separated by spaces. There is a fourth pre-defined model using a string value and a numerical one:

- \texttt{spot} \langle name \rangle \langle tint \rangle A pre-defined spot color, where the \langle name \rangle should be a pre-defined string color name and the \langle tint \rangle should be in the range $[0, 1]$.

Additional models may be created to allow mixing of spot colors. The number of data entries these require will depend on the number of colors to be mixed.

\textbf{T\textsc{e}Xhacker} note: The content of \l__color_current_tl comprises two brace groups, the first containing the color model and the second containing the value(s) applicable in that model.

(End definition for \l__color_current_tl.)

\color_group_begin: \texttt{color_group_begin:} Grouping for color is the same as using the basic \texttt{group_begin:} and \texttt{group_end:} functions. However, for semantic reasons, they are renamed here.

\color_group_end: \texttt{color_group_end:}

(End definition for \texttt{color_group_begin:} and \texttt{color_group_end:}. These functions are documented on page \pageref{color_group_begin}.)

1305
\color_ensure_current:  A driver-independent wrapper for setting the foreground color to the current color “now”.

\cs_new_protected:Npn \color_ensure_current:
\__color_backend_pickup:N \l__color_current_tl
\__color_select:N \l__color_current_tl
\}

(End definition for \color_ensure_current:. This function is documented on page 287.)

\s__color_stop  Internal scan marks.
\scan_new:N \s__color_stop

(End definition for \s__color_stop.)

\__color_select:N \__color_select:nn  Take an internal color specification and pass it to the driver. This code is needed to ensure the current color but will also be used by the higher-level experimental material.

\cs_new_protected:Npn \__color_select:N #1
\exp_after:wN \__color_select:nn #1
\cs_new_protected:Npn \__color_select:nn #1#2
\use:c { __color_backend_select_ #1 :n } {#2}

(End definition for \__color_select:N and \__color_select:nn.)

\l__color_current_tl  The current color, with the model and
\tl_new:N \l__color_current_tl
\tl_set:Nn \l__color_current_tl { { gray } { 0 } }

(End definition for \l__color_current_tl.)

86.2 Predefined color names

The ability to predefine colors with a name is a key part of this module and means there has to be a method for storing the results. At first sight, it seems natural to follow the usual expl3 model and create a color variable type for the process. That would then allow both local and global colors, constant colors and the like. However, these names need to be accessible in some form at the user level, for selection of colors either simply by name or as part of a more complex expression. This does not require that the full name is exposed but does require that they can be looked up in a predictable way. As such, it is more useful to expose just the color names as part of the interface, with the result that only local color names can be created. (This is also seen for example in key creation in l3keys.) As a result, color names are declarative (no \new functions).

Since there is no need to manipulate colors en masse, each is stored in a two-part structure: a prop for the colors themselves, and a tl for the default model for each color.
86.3 Setup

\l__color_internal_int
\l__color_internal_tl
\l__color_internal_int
\l__color_internal_tl
(End definition for \l__color_internal_int and \l__color_internal_tl.)

\s__color_mark
Internal scan marks. \s__color_stop is already defined in l3color-base.
(End definition for \s__color_mark.)

86.4 Utility functions

\__color_if_defined:nTF
A simple wrapper to avoid needing to have the lookup repeated in too many places.

\__color_model:N
\__color_values:N
Simple abstractions.

\__color_extract:nNN
\__color_extract:VNN
Recover the values for the standard model for a color.

\__color_convert:nnN
\__color_convert:VVN
\__color_convert:nnnN
\__color_convert:nVnN
\__color_convert:nnVN
\__color_convert_gray_rgb:w␣␣␣␣␣\__color_convert_gray_cmyk:w␣␣␣␣␣\__color_convert_cmyk_gray:w␣␣␣␣␣\__color_convert_cmyk_rgb:w␣␣␣␣␣\__color_convert_rgb_gray:w␣␣␣␣␣\__color_convert_rgb_cmyk:w
\__color_convert_rgb_cmyk:nnnn
Model conversion is carried out using standard formulae, as described in the manual for xcolor (see also the PostScript Language Reference Manual).
These rather odd values are based on NTSC television: the set are used for the CMYK conversion.

The conversion from RGB to CMYK is the most complex: a two-step procedure which requires black generation and undercolor removal functions. The PostScript reference describes them as device-dependent, but following xcolor we assume they are linear. Moreover, as the likelihood of anyone using a non-unitary matrix here is tiny, we simplify and treat those two concepts as no-ops.

Working space to store the color data whilst doing calculations: keeping it on the stack is attractive but gets tricky (return is non-trivial).

(End definition for \__color_convert:nnN and others.)

86.6 Color expressions

Working space to store the color data whilst doing calculations: keeping it on the stack is attractive but gets tricky (return is non-trivial).
The main function for parsing color expressions removes actives but otherwise expands, then starts working through the expression itself. At the end, we apply the payload.

Before going to all of the effort of parsing an expression, these two precursor functions look for a pre-defined name, either on its own or with a trailing ! (which is the same thing).

Here, we have to allow for the case where there is a fixed model: that can’t be swept up by generic conversion as we are dealing with a named color.
Once we establish that a full parse is needed, the next job is to get the detail of the first color. That will determine the model we use for the calculation: splitting here makes checking that a bit easier.

This is the loop proper: there can be an open-ended set of colors to parse, separated by ! tokens. There are a few cases to look out for. At the end of the expression and with we find a mix of 100 then we simply skip the next color entirely (we can’t stop the loop as there might be a further valid color to mix in). On the other hand, if we get a mix of 0 then drop everything so far and start again. There is also a trailing white to “read in” if the final explicit data is a mix. Those conditions are separate from actually looping, which is therefore sorted out by checking if we have further data to process: in contrast to xcolor, we don’t allow !! so the test can be simplified.
The “payload” of calculation in the loop first. If the model for the upcoming color is different from that of the existing (partial) color, convert the model. For gray the two are flipped round so that the outcome is something with “real” color. We are then in a position to do the actual calculation itself. The two auxiliaries here give us a way to break the loop should an invalid name be found.

```latex
\cs_new_protected:Npn \__color_parse_loop:nn #1#2
\__color_if_defined:nTF {#2}
\__color_extract:nNN {#2} \l__color_next_model_tl \l__color_next_value_tl
\l_if_eq:NNF \l__color_model_tl \l__color_next_model_tl
\__color_convert:nnn {\str_if_eq:VnTF \l__color_model_tl { gray }}
{ \__color_parse_gray:n {#2} }
{ \__color_parse_mix:NVVn \l__color_model_tl \l__color_value_tl \l__color_next_value_tl {#1} }
\msg_error:nnn { color } { unknown-color } {#2}
\__color_extract:nNN { black } \l__color_model_tl \l__color_value_tl
\__color_parse_break:w
\}
```

The gray model needs special handling: the models need to be swapped: we do that using a dedicated function.

```latex
\cs_new_protected:Npn \__color_parse_gray:n #1
\tl_set_eq:NN \l__color_model_tl \l__color_next_model_tl
\tl_set:Nn \l__color_next_model_tl { gray }
\prop_get:cVNN { l__color_named_ #1_prop } \l__color_model_tl
\l__color_value_tl
\__color_convert:nnn { gray } \l__color_model_tl \l__color_value_tl
\__color_parse_break:w
\}
```

The gray model needs special handling: the models need to be swapped: we do that using a dedicated function.

```latex
\cs_new_protected:Npn \__color_parse_std:n #1
\prop_get:cVNN { l__color_named_ #1_prop } \l__color_model_tl
\l__color_next_value_tl
Do the vector arithmetic: mainly a question of shuffling input, along with one pre-
calculation to keep down the use of division.

Turn the input into internal form, also tidying up the number quickly.
\[
\begin{align*}
\text{\texttt{\textbackslash cs_new:Npn \_\_color_parse_number:n #1}} & - \\
\text{\texttt{\textbackslash cs_new:Npn \_\_color_parse_number:n #2}} & - \\
\text{\texttt{\textbackslash cs_new:Npn \_\_color_parse_number:n #3}} & - \\
\text{\texttt{\textbackslash cs_new:Npn \_\_color_parse_number:n #4}} & - \\
\end{align*}
\]

(End definition for \_\_color_parse_model_gray:w and others.)

The conversion here is non-trivial but is described at length in the \texttt{xcolor} manual. For ease, we calculate the integer and fractional parts of the hue first, then use them to work out the possible values for \(r\), \(g\) and \(b\) before putting them in the correct places.

\[
\begin{align*}
\text{\texttt{\textbackslash cs_new:Npn \_\_color_parse_model_wave:nn \_\_color_parse_model_wave_aux:nn \_\_color_parse_model_wave_auxii:nn \_\_color_parse_model_wave_rho:n}} & - \\
\end{align*}
\]
Following the description in the xcolor manual. As we always use rgb, there is no need to find the sixth, we just pass the information straight to the hsb auxiliary defined earlier.
\cs_new:Npn \_color_parse_model_wave_auxi:nn #1#2
{
  \fp_compare:nNnTF {#1} < { 440 }
  {

    \_color_parse_model_wave_auxii:nn
    { 4 + \_color_parse_model_wave_rho:n { (#1 - 440) / -60 } }
    {#2}
  }
  {
    \fp_compare:nNnTF {#1} < { 490 }
    {

      \_color_parse_model_wave_auxii:nn
      { 4 - \_color_parse_model_wave_rho:n { (#1 - 440) / 50 } }
      {#2}
    }
    {
      \fp_compare:nNnTF {#1} < { 510 }
      {

        \_color_parse_model_wave_auxii:nn
        { 2 + \_color_parse_model_wave_rho:n { (#1 - 510) / -20 } }
        {#2}
      }
      {
        \fp_compare:nNnTF {#1} < { 580 }
        {

          \_color_parse_model_wave_auxii:nn
          { 2 - \_color_parse_model_wave_rho:n { (#1 - 510) / 70 } }
          {#2}
        }
        {
          \fp_compare:nNnTF {#1} < { 645 }
          {

            \_color_parse_model_wave_auxii:nn
            { \_color_parse_model_wave_rho:n { (#1 - 645) / -65 } }
            {#2}
          }
          {
            \_color_parse_model_wave_auxii:nn { 0 } {#2}
          }
        }
      }
    }
  }
}{
  \exp_args:Neee \_color_parse_model_hsb_aux:nnn
  { \fp_eval:n {#1} }
  {#2}
}{
  \_color_parse_model_wave_rho:n {#2}
}
\cs_new:Npn \_color_parse_model_wave:rho:n #1
{
  \fp_eval:n { min(1, max(0,#1) ) }
}(End definition for \_color_parse_model_Gray:w and others.)
86.7 Selecting colors (and color models)

\l_color_fixed_model_tl

For selecting a single fixed model.

\l_color_fixed_model_tl (End definition for \l_color_fixed_model_tl. This variable is documented on page 290.)

\__color_check_model:N
\__color_check_model:nn

Check that the model in use is the one required.

\cs_new_protected:Npn \__color_check_model:N #1
{\tl_if_empty:NF \l_color_fixed_model_tl
{\exp_after:wN \__color_check_model:nn #1
\tl_if_eq:NNF \l__color_model_tl \l_color_fixed_model_tl
{\__color_convert:VVN \l__color_model_tl \l_color_fixed_model_tl
\l__color_value_tl}
\tl_set:Nx #1
{\l_color_fixed_model_tl \l__color_value_tl}}}
\cs_new_protected:Npn \__color_check_model:nn #1#2
{\tl_set:Nn \l__color_model_tl {#1}
\tl_set:Nn \l__color_value_tl {#2}}

(End definition for \__color_check_model:N and \__color_check_model:nn.)

\__color_select:

A backend-neutral location for “last minute” manipulations before handing off to the backend code. We set the special \_ syntax here: this will therefore always be available. The finalisation is separate from the main function so it can also be applied to e.g. page color.

\cs_new_protected:Npx \__color_select:
{\tl_set:Nx \exp_not:c { l__color_named_ . _tl }
{\exp_not:N \__color_model:N \exp_not:N \l__color_current_tl}
\prop_clear:N \exp_not:c { l__color_named_ . _prop }
\prop_put:NVx \exp_not:c { l__color.named_ . _prop }
\exp_not:c { l__color.named_ . _tl }
\exp_not:N \l__color_values:N \exp_not:N \l__color_current_tl
\__color_select:N \exp_not:N \l__color_current_tl}

(End definition for \__color_select:)
34114 } \cs_new_protected:Npn \color_select:nn #1#2
34116 { \_\_color_select_main:Nw \l__color_current_tl
34117 #1 / / \_s__color_mark #2 / / \_s__color_stop \_\_color_select:
34119 \_\_color_select:
34120 }

If the first color model is the fixed one, or if there is no fixed model, we don’t need most
of the data: just set up and apply the backend function.
34121 \cs_new_protected:Npn \_\_color_select_main:Nw
34122 #1 #2 / #3 / #4 \_s__color_mark #5 / #6 / #7 \_s__color_stop
34123 { \_\_color_select:nnN {#2} {#5} #1 \bool_lazy_or:nnF
34124 { \tl_if_empty_p:N \l_color_fixed_model_tl }
34125 { \str_if_eq_p:nV {#2} \l_color_fixed_model_tl }
34126 { \_\_color_select_loop:Nw #1 #3 / #4 \_s__color_mark #6 / #7 \_s__color_stop }
34127 }

If a fixed model applies, we need to check each possible value in order. If there is no hit
at all, fall back on the generic formula-based interchange.
34128 \cs_new_protected:Npn \_\_color_select_loop:Nw
34129 #1 #2 / #3 \_s__color_mark #4 / #5 \_s__color_stop
34130 { \str_if_eq:nVTF {#2} \l_color_fixed_model_tl
34131 { \_\_color_select:nnN {#2} {#4} #1 }
34132 { \tl_if_blank:nTF {#2}
34133 { \exp_after:wN \_\_color_select_swap:Nnn \exp_after:wN #1 #1 }
34134 { \_\_color_select_loop:Nw #1 #3 \_s__color_mark #5 \_s__color_stop }
34135 }
34136 }
34137 \cs_new_protected:Npn \_\_color_select:nnN #1#2#3
34138 { \cs_if_exist:cTF { __color_parse_model_ #1 :w }
34139 { \tl_set:Nx \l_color_fixed_model_tl #3
34140 { \use:c { __color_parse_model_ #1 :w } #2 , 0 , 0 , 0 \_s__color_stop }
34141 } { \msg_error:nnn { color } { unknown-model } {#1} }
34142 }
34143 \cs_new_protected:Npn \_\_color_select_swap:Nnn #1#2#3
34144 { \_\_color_convert:nVnN {#2} \l_color_fixed_model_tl {#3} \_\_color_value_tl
34145 \tl_set:Nx \l_color_fixed_model_tl #1
34146 { \_\_color_fixed_model_tl } { \_\_color_value_tl }
34147 }

(End definition for \color_select:n and others. These functions are documented on page 290.)

86.8 Fill and stroke color

\color_fill:n
\color_stroke:n
\color_fill:nn
\color_stroke:nn
\_\_color_draw:nnn
86.9 Defining named colors

\l_color_named_tl  Space to store the detail of the named color.

Defining named colors means working through the model list and saving both the “main” color and any equivalents in other models. Even if there is only one model, we store a prop as well as a tl, as there could be grouping weirdness, etc. When setting using an expression, we need to avoid any fixed model issues, which is done without a group as in l3keys.

End definition for \color_fill:n and others. These functions are documented on page 290.
When setting an expression-based color, there could be multiple model data available for one or more of the input colors. Where that is true for the first named color in an expression, we re-parse the expression when they are also parameter-based: only cmyk, gray and rgb make any sense here. There is a bit of a performance hit but this should be rare and taking place during set-up.

\cs_new_protected:Npn \__color_set:nnw #1#2#3 ! #4 \s__color_stop
\clist_map_inline:nn { cmyk, gray, rgb }
\prop_get:cnNT { \l__color_named_ #3 _prop } {##1} \l__color_internal_tl
\prop_if_in:cnF { \l__color_named_ #1 _prop } {##1}
\group_begin:
\tl_set:cn { \l__color_named_ #3 _tl } {##1}
\__color_parse:nN {#2} \l__color_internal_tl
\exp_args:NNNV \group_end:
\tl_set:Nn \l__color_internal_tl \l__color_internal_tl
\prop_put:cvx { \l__color_named_ #1 _prop } { \l__color_named_ #1 _tl }
\__color_set:nnw {#1} {#2} ! \s__color_stop
\cs_new_protected:Npx \__color_set_aux:nnn #1#2#3
\exp_not:N \__color_set_colon:nnw {#2} {#3} #1 \c_colon_str \exp_not:N \s__color_stop
\use:x
\{
A small set of colors are always defined.
\color_set:nnn { black } { gray } { 0 }
\color_set:nnn { white } { gray } { 1 }
\color_set:nnn { cyan } { cmyk } { 1, 0, 0, 0 }
\color_set:nnn { magenta } { cmyk } { 0, 1, 0, 0 }
\color_set:nnn { yellow } { cmyk } { 0, 0, 1, 0 }

(End definition for \color_set:nn and others. These functions are documented on page 289.)
\l__color_named_._prop
\l__color_named_._tl
A special named color: this is always defined though not fixed in definition.
\prop_new:c { l__color_named_._prop }
\tl_new:c { l__color_named_._tl }
\tl_set:cx { l__color_named_._tl } { \_color_model:N l__color_current_tl }
(End definition for \l__color_named_._prop and \l__color_named_._tl.)

86.10 Exporting colors
\color_export:nnN
\color_export:nnnN
\__color_export:nN
\__color_export:nnnN
\cs_new_protected:Npn \color_export:nnN #1#2#3
\group_begin:
\tl_if_exist:cT { c__color_export_ #2 _tl }
{ \tl_set_eq:Nc \l_color_fixed_model_tl { c__color_export_ #2 _tl } }
\__color_parse:nN {#1} #3
\__color_export:nN {#2} #3
\exp_args:NNNV \group_end:
\tl_set:Nn #3 #3
\cs_new_protected:Npn \color_export:nnnN #1#2#3#4
\__color_select_main:Nw #4
#1 / / \s__color_mark #2 / / \s__color_stop
\__color_export:nN {#3} #4
\cs_new_protected:Npn \__color_export:nnnNN #1#2#3#4#5
\cs_if_exist_use:cF { __color_export_format_ #3 :nnN }
{ \msg_error:nnn { color } { unknown-export-format } {#3}
  \use_none:nnn }
  {#1} {#2} #4
\end
(End definition for \color_export:nnN and others. These functions are documented on page 292.)
\__color_export_format_backend:nnN
Simple.
\cs_new_protected:Npn \__color_export_format_backend:nnN #1#2#3
{ \tl_set:Nn #3 { {#1} {#2} } }
(End definition for \__color_export_format_backend:nnN)
\__color_export:nnNN
A generic auxiliary for cases where only one model is appropriate.
\cs_new_protected:Npn \__color_export:nnNN #1#2#3#4#5
{}
\l_set:Nx \__color_export_HTML:n \#1
\__color_export_HTML:n \#2\__color_export_HTML:n \#3\__color_export_HTML:n \#4
\cs_new:Npn \__color_export_HTML:n #1
\{\fp_compare:nNnTF {#1} = { 0 } { 00 } { \fp_compare:nNnT { #1 * 255 } < { 16 } { 0 } \int_to_Hex:n \{ \fp_to_int:n \{ #1 * 255 \} \} \}
\cs_new_protected:cpn { \__color_export_space-sep-rgb:Nw } #1#2 \s__color_stop
\(\text{(End definition for } \__color_export_comma-sep-rgb:Nw \text{ and others.)}\)

86.11 Additional color models

\l__color_internal_prop
\prop_new:N \l__color_internal_prop
\(\text{(End definition for } \l__color_internal_prop.\)\)

\g__color_model_int
A tracker for the total number of new models.
\int_new:N \g__color_model_int
\(\text{(End definition for } \g__color_model_int.\)\)

\c__color_fallback_cmyk_tl\c__color_fallback_gray_tl\c__color_fallback_rgb_tl
Conversion from Separation or DeviceN spaces may not be possible; have a fallback to black.
\tl_const:Nn \c__color_fallback_cmyk_tl { 0 - 0 - 0 - 1 }
\tl_const:Nn \c__color_fallback_gray_tl { 1 }
\tl_const:Nn \c__color_fallback_rgb_tl { 1 - 1 - 1 }
\(\text{(End definition for } \c__color_fallback_cmyk_tl, \c__color_fallback_gray_tl, \text{ and } \c__color_fallback_rgb_tl.\)\)

\g__color_colorants_prop
Mapping from names to colorants.
\prop_new:N \g__color_colorants_prop
\prop_gput:Nnn \g__color_colorants_prop { black } { Black }
\prop_gput:Nnn \g__color_colorants_prop { blue } { Blue }
\prop_gput:Nnn \g__color_colorants_prop { cyan } { Cyan }
\prop_gput:Nnn \g__color_colorants_prop { green } { Green }
\prop_gput:Nnn \g__color_colorants_prop { magenta } { Magenta }
\prop_gput:Nnn \g__color_colorants_prop { none } { None }
\prop_gput:Nnn \g__color_colorants_prop { red } { Red }
\prop_gput:Nnn \g__color_colorants_prop { yellow } { Yellow }
\(\text{(End definition for } \g__color_colorants_prop.\)\)
\color_model_new:nnn\_color_model_new:nnn

Set up a new model: in general this has to be handled by a family-dependent function. To avoid some “interesting” questions with casing, we fold the case of the family name. The key-value list should always be present, so we convert it up-front to a \prop, then deal with the detail on a per-family basis.

\cs_new_protected:Npn \color_model_new:nnn #1#2#3
\exp_args:Nee \_color_model_new:nnn
{ \tl_to_str:n {#1} } \str_foldcase:n {#2} \{#3\}
\cs_new_protected:Npn \_color_model_new:nnn #1#2#3
{\cs_if_exist:cTF {__color_parse_model_#1:w} 
{\msg_error:nnn {color} {model-already-defined} {#1}}
\cs_if_exist:cTF {__color_model_#1:w} 
\cs_if_exist:cTF {__color_alternative_model_prop} 
{\clist_map_inline:nnn \clist_map_inline:nnn { cyan , magenta , yellow , black } 
{\prop_gput:Nnn __color_alternative_model_prop {#1} { cmyk } }
\clist_map_inline:nn { red , green , blue } 
{\prop_gput:Nnn __color_alternative_model_prop {#1} { rgb } }
\end_definition_for \g__color_alternative_model_prop.
\end_definition_for \g__color_alternative_values_prop.
\g__color_alternative_values_prop

Same for the values: a bit more involved.

\prop_new:N \g__color_alternative_values_prop
\prop_gput:Nnn __color_alternative_values_prop { cyan } { 1 , 0 , 0 , 0 }
\prop_gput:Nnn __color_alternative_values_prop { magenta } { 0 , 1 , 0 , 0 }
\prop_gput:Nnn __color_alternative_values_prop { yellow } { 0 , 0 , 1 , 0 }
\prop_gput:Nnn __color_alternative_values_prop { black } { 0 , 0 , 0 , 1 }
\prop_gput:Nnn __color_alternative_values_prop { red } { 1 , 0 , 0 }
\prop_gput:Nnn __color_alternative_values_prop { green } { 0 , 1 , 0 }
\prop_gput:Nnn __color_alternative_values_prop { blue } { 0 , 0 , 1 }
\end_definition_for \g__color_alternative_values_prop.

\ color_model_new:nnn

\__color_model_new:nnn

Set up a new model: in general this has to be handled by a family-dependent function. To avoid some “interesting” questions with casing, we fold the case of the family name. The key-value list should always be present, so we convert it up-front to a \prop, then deal with the detail on a per-family basis.

\cs_new_protected:Npm \color_model_new:nnn #1#2#3
\exp_args:Nee \__color_model_new:nnn
{ \tl_to_str:n {#1} } \str_foldcase:n {#2} \{#3\}
\cs_new_protected:Npm \__color_model_new:nnn #1#2#3
\cs_if_exist:cTF \__color_parse_model_#1:w
{\mag_error:nnn {color} {model-already-defined} {#1}}
\}
\}

1324
\cs_if_exist:cTF { __color_model_ #2 :n } 
{ 
  \prop_set_from_keyval:Nn \l__color_internal_prop { #3 } 
  \use:c { __color_model_ #2 :n } { #1 } 
} 
\msg_error:nnn { color } { unknown-model-type } { #2 } 
\end{lstlisting}

(End definition for \color_model_new:nnn and \__color_model_new:nnn. This function is documented on page 292.)

\__color_model_init:nn

A shared auxiliary to do the basics of setting up a new model: reserve a number, create a fallback and white-equivalent, set up links to the backend.

\cs_new_protected:Npn \__color_model_init:nnnnn #1#2#3#4#5
{ 
  \int_gincr:N \g__color_model_int 
  \tl_const:cn { c__color_fallback_ #1 _tl } { #4 } 
  \clist_map_inline:nn { fill , stroke , select } 
  { 
    \cs_new_protected:cpx { __color_backend_ ##1 _ #1 :n } ####1 
    { \exp_not:c { __color_backend_ ##1 _ #3 :nn } 
      { color \int_use:N \g__color_model_int } {####1} 
    } 
  } 
  \cs_new_protected:cpx { __color_model_ #1 _white: } 
  { \prop_put:Nnn \exp_not:N \l__color_named_white_prop { #1 } 
    { \exp_not:n { #5 } } 
    \exp_not:N \int_compare:nNnF { \tex_currentgrouplevel:D } = 0 
    { \group_insert_after:N \exp_not:c { __color_model_ #1 _ white: } } 
  } 
  \use:c { __color_model_ #1 _white: } 
}\cs_generate_variant:Nn \__color_model_init:nnnnn { nnnxx }
\cs_new_protected:Npn \__color_model_init:nnn #1#2#3 
{ \__color_model_init:nnnnn { #1 } { #2 } { #3 } 
  1 \prg_replicate:nn { #2 - 1 } { - 1 } 
  0 \prg_replicate:nn { #2 - 1 } { - 0 } 
}\end{lstlisting}

(End definition for \__color_model_init:nn.)

\__color_model_separation:n
\__color_model_separation:nn
\__color_model_separation:nnn
\__color_model_separation:wm
\__color_model_separation_cmyk:nnnnn
\__color_model_separation_gray:nnnnn
\__color_model_separation_rgb:nnnnn
\__color_model_convert:nnn
\__color_model_separation_CIELAB:nnnnn
\__color_model_separation_CIELAB:nnnnnn

Separations must have a “real” name, which is pretty easy to find.

\cs_new_protected:Npn \__color_model_separation:n #1#2#3#4#5 
{ \prop_put:Nnn \exp_not:N \l__color_named_white_prop { #1 } 
  \exp_not:n { #5 } 
  \exp_not:N \int_compare:nNnF { \tex_currentgrouplevel:D } = 0 
  { \group_insert_after:N \exp_not:c { __color_model_ #1 _ white: } } 
} 
\use:c { __color_model_ #1 _white: } 
\cs_generate_variant:Nn \__color_model_init:nnnnn { nnnxx } 
\cs_new_protected:Npn \__color_model_init:nnn { nnnxx } 
{ \__color_model_init:nnnnn { #1 } { #2 } { #3 } 
  1 \prg_replicate:nn { #2 - 1 } { - 1 } 
  0 \prg_replicate:nn { #2 - 1 } { - 0 } 
}\end{lstlisting}

(End definition for \__color_model_init:nn.)

\__color_model_separation:nn
\__color_model_separation:nnn
\__color_model_separation:wm
\__color_model_separation_cmyk:nnnnn
\__color_model_separation_gray:nnnnn
\__color_model_separation_rgb:nnnnn
\__color_model_convert:nnn
\__color_model_separation_CIELAB:nnnnn
\__color_model_separation_CIELAB:nnnnnn

Separations must have a “real” name, which is pretty easy to find.
We have two keys to find at this stage: the alternative space model and linked values.

\cs_new_protected:Npn \__color_model_separation:nn #1#2
\prop_get:NnNTF \l__color_internal_prop { alternative-model } \l__color_internal_tl
\exp_args:NV \__color_model_separation:nnn \l__color_internal_tl {#2} {#1}
\msg_error:nnn { color } { separation-alternative-model } {#2}
\msg_error:nnn { color } { separation-requires-name } {#1}

\cs_new_protected:Npn \__color_model_separation:nnn #1#2#3
\cs_if_exist:cTF { __color_model_separation_ #1 :nnnnnn }
\prop_get:NnNTF \l__color_internal_prop { alternative-values } \l__color_internal_tl
\exp_after:wN \__color_model_separation:w \l__color_internal_tl , 0 , 0 , 0 , 0 \s__color_stop {#2} {#3} {#1}
\msg_error:nnn { color } { separation-alternative-values } {#2}
\msg_error:nnn { color } { unknown-alternative-model } {#1}

As each alternative space leads to a different requirement for conversion, and as there are only a small number of choices, we manually split the data and then set up. Notice that mixing tints is really just the same as mixing gray. The white color is special, as it allows tints to be adjusted without an additional color space. To make sure the data is set for that at all group levels, we need to work on a per-level basis. Within the output, only the set-up needs the “real” name of the colorspace: we use a simple tracking number for general usage as this is a clear namespace without issues of escaping chars.
\_\_color\_model\_init:nnn \{ 1 \} \{ "separation" \}
\cs_new_eq:N \{ \_color\_parse\_mix\_ #6 \} \_color\_parse\_mix\_gray:nw
\cs_new:cN \{ \_color\_parse\_model\_ #6 \} \#1 \#2 \_s\_color\_stop
\{ \#6 \} \{ \_color\_parse\_number:n \{\#1\} \} \}
\use:c \{ \_color\_model\_init\_separation \#8 \} \{nnnnnn \}
\{\#6\} \{\#7\} \{\#1\} \{\#2\} \{\#3\} \{\#4\}
\prop_gput:Nnn \g\_color\_alternative\_model\_prop \{\#6\} \{\#8\}
\prop_gput:Nnx \g\_color\_colorants\_prop \{\#6\}
\{ \str_convert_pdfname:n \{\#7\} \}
\cs_new_protected:Npn \_\_color\_model\_separation\_cmyk:nnnnnn \#1\#2\#3\#4\#5\#6
\{\#6\} \{\_color\_parse\_number:n \{\#1\}\}
\prop_gput:Nnn \g\_color\_alternative\_model\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\prop_gput:Nnx \g\_color\_colorants\_prop \{\#6\}
\{ \str_convert_pdfname:n \{\#7\} \}
\cs_new_protected:Npn \_\_color\_model\_separation\_gray:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceCMYK \} \{\}
\{ 0 \- 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\cs_new_protected:Npn \_\_color\_model\_separation\_rgb:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceRGB \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\}
\cs_new_protected:Npn \_\_color\_model\_separation\_gray:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceRGB \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\}
\cs_new_protected:Npn \_\_color\_model\_separation\_rgb:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceCMYK \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\cs_new_protected:Npn \_\_color\_model\_separation\_gray:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceRGB \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\cs_new_protected:Npn \_\_color\_model\_separation\_rgb:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceCMYK \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\cs_new_protected:Npn \_\_color\_model\_separation\_gray:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceRGB \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\cs_new_protected:Npn \_\_color\_model\_separation\_rgb:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceCMYK \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\cs_new_protected:Npn \_\_color\_model\_separation\_gray:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceRGB \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\cs_new_protected:Npn \_\_color\_model\_separation\_rgb:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceCMYK \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\cs_new_protected:Npn \_\_color\_model\_separation\_gray:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceRGB \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\cs_new_protected:Npn \_\_color\_model\_separation\_rgb:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceCMYK \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
\cs_new_protected:Npn \_\_color\_model\_separation\_gray:nnnnnn \#1\#2\#3\#4\#5\#6
\prop_gput:Nnn \g\_color\_alternative\_values\_prop \{\#1\} \{\#3\} \{\#4\} \{\#5\}
\prop_gput:Nnx \g\_color\_backend\_separation\_init:nnnnnn \#2 \{/DeviceRGB \} \{\}
\{ 0 \- 0 \- 0 \} \{\#3\} \{\#4\} \{\#5\} \{\#6\}
1327
Generic model conversion via an alternative intermediate.

```latex
\cs_new_protected:Npn \__color_model_convert:nnn #1#2#3
\{
\cs_new:cpx { __color_convert_ #1 _ #3 :w } ##1 \s__color_stop
\exp_not:N \exp_args:NNe \exp_not:N \use:nn
\exp_not:c { __color_convert_ #2 _ #3 :w }
\tl_if_exist:cTF { c__color_model_whitepoint_CIELAB_ #1 _tl }
\{ \msg_error:nnn { color } { unknown-CIELAB-illuminant } {#1} 
\}
\cs_new:cpn { __color_convert_ #2 _rgb:w } ##1 \s__color_stop
\cs_new:cpn { __color_convert_ #2 _gray:w } ##1 \s__color_stop
\}
\}
```

Setting up for CIELAB needs a bit more work: there is the illuminant and the need for an appropriate object.

```latex
\cs_new_protected:Npn \__color_model_separation_CIELAB:nnnnnn #1#2#3#4#5#6#7
\{
\prop_get:NnNF \l__color_internal_prop { illuminant } \l__color_internal_tl
\{ \msg_error:nnn { color } { CIELAB-requires-illuminant } {#1} \}
\tl_set:Nn \l__color_internal_tl { d50 }
\}
\exp_args:NV \__color_model_separation_CIELAB:nnnnnn
\l__color_internal_tl {#1} {#2} {#3} {#4} {#5} {#6} {#7}
\}
```

If a CIELAB space is being set up, we need the illuminant, then create the appropriate setup. At present, this doesn’t include BlackPoint or Range data, but that may be added later. As CIELAB colors cannot be converted to anything else, we fall back to producing black: the user should set up a second model for colors set up this way.

```latex
\cs_new_protected:Npn \__color_model_separation_CIELAB:nnnnnnn #1#2#3#4#5#6#7#8
\{\tl_if_exist:cTF { c__color_model_whitepoint_CIELAB_ #1 _tl } 
\{ \__color_backend_separation_init_CIELAB:nnn {#1} {#3} \} \cs_new:cpn { __color_convert_ #2 _cmykw } ##1 \s__color_stop
\cs_new:cpn { __color_convert_ #2 _rgb:w } ##1 \s__color_stop
\cs_new:cpn { __color_convert_ #2 _gray:w } ##1 \s__color_stop
\}
\}
```

1328
We require a list of component names here: one might call them colorants, but it’s convenient to use \TeX names instead so we slightly adjust the terminology.

All valid models will have an alternative listed, either hard-coded for the core device ones, or dynamically added for Separations, etc.
We now complete the data we require by first finding out how many colorants there are, then moving on to begin constructing the function required to map to the alternative color space.

\begin{verbatim}
\cs_new_protected:Npn \__color_model_devicen:nnn #1 #2 #3
{ \exp_args:Nx \__color_model_devicen:nnnn { \clist_count:n {#2} } {#1} {#2} {#3} }
\end{verbatim}

At this stage, we have checked everything is in place, so we can set up the \TeX{} and backend data structures. As for separations, it’s not really possible in general to have a fallback, so we simply provide “black” for each element.

\begin{verbatim}
\cs_new_protected:Npm \__color_model_devicen:nnn #1 #2 #3 #4
{ \__color_model_init:nnn {#4} {#1} { devicen } \cs_if_exist_use:cF { __color_model_devicen_parse_ #1 :nn } { \__color_model_devicen_parse_generic:nn {#4} {#1} } \__color_model_devicen_init:nnn {#1} {#2} {#3} \__color_model_devicen_convert:nnn {#4} {#2} {#3} }
\end{verbatim}

For short lists of DeviceN colors, we can use hand-tuned parsing. This lines up with other models, where we allow for up to four components. For larger spaces, rather than limit artificially, we use a somewhat slow approach based on open-ended commas-lists.

\begin{verbatim}
\cs_new_protected:cpn { __color_model_devicen_parse_1:nn } #1 #2
{ \cs_new:cpn { __color_parse_model_ #1 :w } ##1 , ##2 \s__color_stop { {#1} { \__color_parse_number:n {##1} } } \cs_new_eq:cN { __color_parse_mix_ #1 :nw } \__color_parse_mix_gray:nw
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { __color_model_devicen_parse_2:nn } #1 #2 #3 #4
{ \cs_new:cpn { __color_parse_model_ #1 :w } ##1 , ##2 , ##3 \s__color_stop { {#1} { \__color_parse_number:n {##1} \__color_parse_number:n {##2} } } \cs_new_eq:cN { __color_parse_mix_ #1 :nw } ##1##2 \__color_mark ##3##4 \s__color_stop { \fp_eval:n { ##2 * ##1 + ##4 * ( 1 - ##1 ) } \c_space_tl \fp_eval:n { ##3 * ##1 + ##5 * ( 1 - ##1 ) } }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:cpn { __color_model_devicen_parse_3:nn } #1 #2 #3 #4 #5 #6 #7 #8 #9 #10 #11 #12
{ \cs_new:cpn { __color_parse_model_ #1 :w } ##1 , ##2 , ##3 , ##4 \s__color_stop { {#1} { \__color_parse_number:n {##1} \__color_parse_number:n {##2} \__color_parse_number:n {##3} } } \cs_new_eq:cN { __color_parse_mix_ #1 :nw } ##1##2 #3##4 \__color_stop { \fp_eval:n { ##2 * ##1 + ##4 * ( 1 - ##1 ) } \c_space_tl \fp_eval:n { ##3 * ##1 + ##5 * ( 1 - ##1 ) } }
\end{verbatim}
To construct the tint transformation, we have to use PostScript. The aim is to have the final tint for each device colorant as

$$1 - \prod_{n}(1 - X_n D X_n)$$

where $X$ is a DeviceN colorant and $D$ is the amount of device colorant that the DeviceN colorant maps to. At the start of the process, the PostScript stack will contain the $X_n$ values, whilst we have the $D$ values on a per-DeviceN colorant basis. The more convenient approach for us is therefore to take each DeviceN colorant in turn and find the value $1 - X_n D X_n$, multiplying as we go, and finalise with the subtraction. That contrasts to colorspace: it splits the process up by process color, which works better when you have a fixed list of colorants. (colorspace only supports up to 4 DeviceN colors, and only cmyk as the alternative space.) To set this up, we first need to know the number of values in the target color space: this is easily handled as there are a very small range of possibles. Once we have that information, it’s relatively easy to build the required PostScript using some generic code.

As we always need to split the alternative values into parts, we use a shared auxiliary and only use a minimal difference between code paths. Construction of the tint transformation is as far as possible done using loops, which means there are some inefficiencies for device colors in the DeviceN space: we roll the stack one-at-a-time even if there is a potential shortcut. However, that way there is nothing to special-case. Once this is sorted, we can write the tint transform object, which will remain as the last object until we sort out the final step: the colorant list.
\tl_put_right:Nx \l__color_internal_tl
{
\prg_replicate:nn {#1}
{ \neg - 1.0 - \add - \#1 - \neg - 1 - \roll - }
\int_eval:n { \#2 + \#1 } - \#1 - \roll
\prg_replicate:nn {#2} { - \pop } - \#1 - 1 - \roll}
\use:x
{
\_color_backend_devicen_init:nnn
{
\clist_map_function:nN {#4}
\_color_model_devicen_colorant:n
}
{ \str_case:nn {#3}
{ { cmyk } { /DeviceCMYK } } { gray } { /DeviceGray } { rgb } { /DeviceRGB } }
}
{ \exp_not:V \l__color_internal_tl }
}
\cs_new_protected:Npn \__color_model_devicen_transform:w
#1 , #2 , #3 , #4 , #5 \s__color_stop #6#7
{
\use:c { \__color_model_devicen_transform_ #6 :nnnnn }
{#1} {#2} {#3} {#4} {#7}
}
\cs_new_protected:cpn { \__color_model_devicen_transform_1:nnnnn } #1#2#3#4#5
{ \clist_mapInline:n { \__color_model_devicen_transform:nnn {#5} { 1 } {#1} }
\cs_new_protected:cpn { \__color_model_devicen_transform_3:nnnnn } #1#2#3#4#5
{ \clist_mapInline:n { \__color_model_devicen_transform:nnn {#1} {#3} {#5} }
{ 3 } {#1} }
\cs_new_protected:cpn { \__color_model_devicen_transform_4:nnnnn } #1#2#3#4#5
{ \clist_mapInline:n { \__color_model_devicen_transform:nnn {#1} {#3} {#4} }
{ 4 } {#1} }
\cs_new_protected:Npn \__color_model_devicen_transform:nnn #1#2#3
{ \tl_put_right:Nx \l__color_internal_tl
{ \fp_compare:nNnF {#3} = \c_zero_fp
{ \int_eval:n { \#1 - \l__color_internal_int + \#2 } - \index - \neg - \#3 - \mul - 1.0 - \add - \mul - } }
Here we need to set up conversion from the DeviceN space to the alternative at the \TeX
level. This also means supplying methods for inter-converting to other parameter-based
spaces. Essentially the approach is exactly the same as the PostScript, just expressed in
\TeX terms.
\cs_new:Npn \__color_convert_devicen_cmyk:nnnnw #1#2#3#4#5 ~ #6 \s__color_mark #7#8 \s__color_stop
\{ \exp_not:N \__color_convert_devicen_cmyk_aux:nnnnw { \fp_eval:n { #2 * (1 - (#1 * #6)) } } { \fp_eval:n { #3 * (1 - (#1 * #7)) } } { \fp_eval:n { #4 * (1 - (#1 * #8)) } } { \fp_eval:n { #5 * (1 - (#1 * #9)) } } \}
\cs_new:Npn \__color_convert_devicen_gray:nw #1#2 ~ #3 \s__color_mark #4#5 \s__color_stop
\{ \exp_not:N \__color_convert_devicen_gray:nw { \fp_eval:n { #2 * (1 - (#1 * #5)) } } { \fp_eval:n { #3 * (1 - (#1 * #5)) } } { \fp_eval:n { #5 * (1 - (#1 * #5)) } } \}
\cs_new:Npn \__color_convert_devicen_gray:nnn #1#2#3 \s__color_stop
\{ \exp_not:N \__color_convert_devicen_gray:nnn { \fp_eval:n { 1 - (#1 ) } } { \fp_eval:n { 1 - (#2 ) } } { \fp_eval:n { 1 - (#3 ) } } \}
\cs_new:Npn \__color_model_devicen_convert_aux:w #1 , , , , \s__color_stop }
\cs_new:Npn \__color_model_devicen_convert_aux:w #1 , #2 , #3 , #4 , #5 \s__color_stop
\{ #1
\tl_if_blank:nF {#2}
\{ #2
\tl_if_blank:nF {#3}
\{ #3
\tl_if_blank:nF {#4} { {#4} }
\}
\}
\cs_new:Npn \__color_convert_devicen_cmyk:nnnnnnnnn #1#2#3#4#5#6#7#8#9
\{ \use:e \begin{array}{l}
\exp_not:N \__color_convert_devicen_cmyk_aux:nnnnw
\{ \fp_eval:n { #2 * (1 - (#1 * #6)) }\}
\{ \fp_eval:n { #3 * (1 - (#1 * #7)) }\}
\{ \fp_eval:n { #4 * (1 - (#1 * #8)) }\}
\{ \fp_eval:n { #5 * (1 - (#1 * #9)) }\}
\end{array}
\}
\cs_new:Npn \__color_convert_devicen_cmyk:nnnnw #1#2#3#4 #5 \s__color_mark #6 \s__color_stop
\{ \tl_if_blank:nTF {#5}
\{ \fp_eval:n { 1 - #1 } - \\
\fp_eval:n { 1 - #2 } - \\
\fp_eval:n { 1 - #3 } - \\
\fp_eval:n { 1 - #4 }
\}
\{ \__color_convert_devicen_cmyk:nnnnw (#1) {#2} {#3} {#4}
\#5 \s__color_mark #6 \s__color_stop 
\}
\cs_new:Npn \__color_convert_devicen_gray:nnn #1#2#3 \s__color_stop
\{ \__color_convert_devicen_gray:nnn (#2) {#1} #4
\#3 \s__color_mark #5 \s__color_stop 
\}
\cs_new:Npn \__color_convert_devicen_gray:nw #1#2 - #3 \s__color_mark #4#5 \s__color_stop
\{ 
\cs_new:Npn \__color_convert_devicen_gray:nw #1#2#3\s__color_stop
\{ 
\fp_eval:n { 1 - (#1 ) } \\
\fp_eval:n { 1 - (#2 ) } \\
\fp_eval:n { 1 - (#3 ) } \\
\fp_eval:n { 1 - (#4 ) }
\}

The signatures in the ICC file header indicating the underlying colorspace. We map it to three values: The number of components, the values corresponding to white, and the range.

\prop_const_from_keyval:Nn \c__color_icc_colorspace_signatures_prop
{ % Gray 47524159 = {1} {1} {0} {}, % RGB}
For an ICC profile, we need a file name and a number of components. The file name is processed here so the backend can treat it as a string.

\begin{verbatim}
\cs_new_protected:Npn \__color_model_iccbased:n #1
\{
  \prop_get:NnNTF \l__color_internal_prop { file }
  \l__color_internal_tl
  \exp_args:NV \__color_model_iccbased:nn
  \l__color_internal_tl {#1}
  \}
  \exp_last_unbraced:NV \__color_model_iccbased_aux:nnnnnn
  \l__color_internal_tl {#2} {#1}
\}
\cs_new_protected:Npn \__color_model_iccbased:nn #1#2
\{
  \exp_args:NNx \prop_get:NnNTF \c__color_icc_colorspace_signatures_prop
  \file_hex_dump:nnn { #1 } { 17 } { 20 }
  \l__color_internal_tl
  \exp_last_unbraced:NV \__color_model_iccbased_aux:nnnnnn
  \l__color_internal_tl { #2 } { #1 }
  \}
  \exp_args:Nx \__color_backend_iccbased_init:nnn
  \file_full_name:n {#6} {#1} {#4}
\}
\end{verbatim}

Here, we can use the same internals as for DeviceN approach as we know the number of components. No conversion is possible, so there is no need to worry about that at all.

86.12 Applying profiles

With a limited range of outcomes, this is largely about getting data to the backend.
(End definition for \color_profile_apply:nn and others. This function is documented on page 299.)

86.13 Diagnostics

Extract the information about a color and format for the user: the approach is similar to the keys module here.
(End definition for \texttt{\_\_color\_show:nn} \#1 and others. These functions are documented on page 290.)

86.14 Messages

\begin{verbatim}
\msg_new:nnnn { color } { CIELAB-requires-illuminant }
{ CIELAB-color-space-'#1'-require-an-illuminant. }
{ LaTeX has been asked to create a separation color space using CIELAB specifications, but no-\iow_indent:n { illuminant=-<basis> } \iow_indent:n { \<basis> } key was given with the correct information. LaTeX will use illuminant \texttt{d50} for recovery. }
\msg_new:nnnn { color } { conversion-not-available }
{ No model conversion available from '#1'-to- '#2'. }
{ LaTeX has been asked to convert a color from model '#1'-to-model '#2', but there is no method available to do that. }
\msg_new:nnnn { color } { DeviceN-inconsistent-alternative }
{ DeviceN-color-spaces-require-a-single-alternative-space. }
{ LaTeX has been asked to create a DeviceN color space '#1', but the constituent colors do not have a common alternative color. }
\msg_new:nnnn { color } { DeviceN-no-alternative }
{ DeviceN-color-spaces-require-an-alternative-space. }
{ LaTeX has been asked to create a DeviceN color space '#1', but the constituent colors do not all have a device-based alternative. }
\msg_new:nnnn { color } { DeviceN-requires-names }
{ DeviceN-color-space-'#1'-require-a-list-of-names. }
{ LaTeX has been asked to create a DeviceN color space, but no-\iow_indent:n { names=<names> } \iow_indent:n { \<names> } key was given with the correct information. }
\msg_new:nnnn { color } { ICC-Device-unknown }
{ Unknown-device-color-space-'#1'. }
{ LaTeX has been asked to apply an ICC profile but the device color space '#1' is unknown. }
\end{verbatim}
LaTeX has been asked to create an ICCBased color space, but the used data color space is not supported. ICC profiles used for defining an ICCBased colorspace should use a Lab, RGB, or CMYK data color space. LaTeX will ignore this request.

LaTeX has been asked to create a separation color space, but no \ioindent:\key\=\<name> key was given with the correct information. LaTeX will ignore this request.

LaTeX was asked to define a new color model called \#1', but this color model already exists.

LaTeX has been asked to create a separation color space, but no \ioindent:\key\=\<model> key was given with the correct information.

LaTeX has been asked to create a separation color space, but no \ioindent:\key\=\<model> key was given with the correct information.

LaTeX has been asked to create a separation color space, but no \ioindent:\key\=\<formal\> name key was given with the correct information.
LaTeX has been asked to use a color named '\#1', but this has never been defined.

\newmsg{color}{unknown-alternative-model}{Separation color space '\#1' require an valid alternative space.}{LaTeX has been asked to create a separation color space, but the model given as alternative-model=-<model>}\newmsg{color}{unknown-export-format}{Unknown export format '\#1'.}{LaTeX has been asked to export a color in format '\#1', but this has never been defined.}\newmsg{color}{unknown-CIELAB-illuminant}{Unknown illuminant model '\#1'.}{LaTeX has been asked to use create a color space using CIELAB illuminant '\#1', but this does not exist.}\newmsg{color}{unknown-model}{Unknown color model '\#1'.}{LaTeX has been asked to use a color model called '\#1', but this model is not set up.}\newmsg{color}{unknown-model-type}{Unknown color model type '\#1'.}{LaTeX has been asked to create a new color model called '\#1', but this type of model was never set up.}\prop_gput:Nnn \g_msg_module_name_prop{color}{LaTeX3}\prop_gput:Nnn \g_msg_module_type_prop{color}{\}
\msg_new:nnn{color}{show}{The color \#1 with the properties: \#2}
Chapter 87

\l3pdf implementation

\s__pdf_stop Internal scan marks.
\scan_new:N \s__pdf_stop

(End definition for \s__pdf_stop.)

\g__pdf_init_bool A flag so we have some chance of avoiding setting things we are not allowed to. As we are potentially early in the format, we have to work a bit harder than ideal.
\bool_new:N \g__pdf_init_bool
\bool_lazy_and:nnT
\{ \str_if_eq_p:Vn \fmtname { LaTeX2e } \}
\{ \tl_if_exist_p:N \@expl@finalise@setup@@ \}
\{ \tl_gput_right:Nn \@expl@finalise@setup@@
\{ \tl_gput_right:Nn \@kernel@after@begindocument
\{ \bool_gset_true:N \g__pdf_init_bool \}
\}
\}

(End definition for \g__pdf_init_bool.)

87.1 Compression

\pdf_uncompress: Simple to do.
\cs_new_protected:Npn \pdf_uncompress:
\{ \bool_if:NF \g__pdf_init_bool
\{ \_pdf_backend_compresslevel:n { 0 }
\_pdf_backend_compress_objects:n { \c_false_bool }
\}
\}

(End definition for \pdf_uncompress:. This function is documented on page 296.)
87.2 Objects

Simple to do.

87.3 Version

To compare version, we need to split the given value then deal with both major and
minor version

(End definition for \pdf_pageobject_ref:n. This function is documented on page 295.)
\bool_lazy_and_p:nn
{ \int_compare_p:nNn \__pdf_backend_version_major: = {#1} }
{ \int_compare_p:nNn \__pdf_backend_version_minor: < {#2} }
{ \prg_return_true: }
{ \prg_return_false: }
\cs_new:cpn { __pdf_version_compare_>:w } #1 . #2 . #3 \s__pdf_stop
{ \bool_lazy_or:nnTF
{ \int_compare_p:nNn \__pdf_backend_version_major: > {#1} }
{ \bool_lazy_and_p:nn
{ \int_compare_p:nNn \__pdf_backend_version_major: = {#1} }
{ \int_compare_p:nNn \__pdf_backend_version_minor: > {#2} }
{ \prg_return_true: }
{ \prg_return_false: }
}
\pdf_version_gset:n
\pdf_version_min_gset:n
\__pdf_version_gset:w
\cs_new_protected:Npn \pdf_version_gset:n #1
{ \__pdf_version_gset:w #1 . . \s__pdf_stop }
\cs_new_protected:Npn \pdf_version_min_gset:n #1
{ \\pdf_version_compare:NnT < {#1}
{ \__pdf_version_gset:w #1 . . \s__pdf_stop }
}
\cs_new_protected:Npn \__pdf_version_gset:w #1 . #2 . #3 \s__pdf_stop
{ \bool_if:NF \g__pdf_init_bool
{ \\__pdf_backend_version_major_gset:n {#1}
\\__pdf_backend_version_minor_gset:n {#2}
}
}
\pdf_version:
\pdf_version_major:
\pdf_version_minor:
\cs_new:Npn \pdf_version:
{ \__pdf_backend_version_major: . \__pdf_backend_version_minor: }
\cs_new:Npn \pdf_version_major: { \__pdf_backend_version_major: }
\cs_new:Npn \pdf_version_minor: { \__pdf_backend_version_minor: }

(End definition for \pdf_version_gset:n and others. This function is documented on page ??.)

(End definition for \pdf_version_gset:n, \pdf_version_min_gset:n, and \__pdf_version_gset:w. These functions are documented on page 295.)

Wrappers,

(End definition for \pdf_version:, \pdf_version_major:, and \pdf_version_minor:. These functions are documented on page 295.)
87.4 Destinations

\pdf_destination:nn
\cs_new_protected:Npn \pdf_destination:nn \#1\#2
\{ \_\_pdf_backend_destination:nn \#1 \#2 \}

(End definition for \pdf_destination:nn. This function is documented on page 296.)

\pdf_destination:nnnn
\cs_new_protected:Npn \pdf_destination:nnnn \#1\#2\#3\#4
\{ \hbox_to_zero:n
\{ \_\_pdf_backend_destination:nnnn \#1 \#2 \#3 \#4 \}
\}

(End definition for \pdf_destination:nnnn. This function is documented on page 297.)

\package
Chapter 88

l3candidates Implementation

88.1 Additions to l3box

88.1.1 Viewing part of a box

A wrapper around the driver-dependent code.

\cs_new_protected:Npn \box_clip:N #1
\{ \hbox_set:Nn #1 { \__box_backend_clip:N #1 } \}
\cs_generate_variant:Nn \box_clip:N { c }
\cs_new_protected:Npn \box_gclip:N #1
\{ \hbox_gset:Nn #1 { \__box_backend_clip:N #1 } \}
\cs_generate_variant:Nn \box_gclip:N { c }

(End definition for \box_clip:N and \box_gclip:N. These functions are documented on page 300.)

Trimming from the left- and right-hand edges of the box is easy: kern the appropriate parts off each side.

\cs_new_protected:Npn \box_set_trim:Nnnnn #1#2#3#4#5
\{ \hbox_set:Nn \l__box_internal_box
\{ \__kernel_kern:n { -#2 } \box_use:N #1 \__kernel_kern:n { -#4 } \}
\}
\cs_generate_variant:Nn \box_set_trim:Nnnnn { c }
\cs_new_protected:Npn \box_gset_trim:Nnnnn #1#2#3#4#5
\{ \__box_set_trim:NnnnnN \box_gset_trim:NnnnnN \box_set_eq:NN \}
\cs_generate_variant:Nn \box_set_trim:Nnnnn { c }
\cs_new_protected:Npn \__box_set_trim:NnnnnN #1#2#3#4#5#6
\{ \hbox_set:Nn \l__box_internal_box
\{ \__kernel_kern:n { -#2 } \box_use:N #1 \__kernel_kern:n { -#4 } \}
\}

For the height and depth, there is a need to watch the baseline is respected. Material always has to stay on the correct side, so trimming has to check that there is enough material to trim. First, the bottom edge. If there is enough depth, simply set the depth, or if not move down so the result is zero depth. \box_move_down:nn is used in both
cases so the resulting box always contains a \lower primitive. The internal box is used here as it allows safe use of \box_set_dp:Nn.

\dim_compare:nTF { \box_dp:N #1 } > (#3)
  \hbox_set:Nn \l__box_internal_box
  \box_move_down:nn \c_zero_dim
  \box_use_drop:N \l__box_internal_box
 \box_set_dp:Nn \l__box_internal_box \{ \box_dp:N #1 - (#3) \}
\box_set_dp:Nn \l__box_internal_box { \box_dp:N #1 - (#3) }

Same thing, this time from the top of the box.

\dim_compare:nTF { \box_ht:N \l__box_internal_box } > (#5)
  \hbox_set:Nn \l__box_internal_box
  \box_move_up:nn \c_zero_dim
  \box_use_drop:N \l__box_internal_box
 \box_set_ht:Nn \l__box_internal_box \{ \box_ht:N \l__box_internal_box - (#5) \}
\box_set_ht:Nn \l__box_internal_box { \box_ht:N \l__box_internal_box - (#5) }

\box_set_viewport:Nnnnn \box_gset_viewport:Nnnnn \__box_set_viewport:Nnnnn

The same general logic as for the trim operation, but with absolute dimensions. As a result, there are some things to watch out for in the vertical direction.

\cs_new_protected:Npn \box_set_viewport:Nnnnn #1#2#3#4#5
  \__box_set_viewport:NnnnnN #1 {#2} {#3} {#4} {#5} \box_set_eq:NN
\cs_generate_variant:Nn \box_set_viewport:Nnnnn { c }
\cs_new_protected:Npn \box_gset_viewport:Nnnnn #1#2#3#4#5
  \__box_set_viewport:NnnnnN #1 {#2} {#3} {#4} {#5} \box_gset_eq:NN
\cs_generate_variant:Nn \box_gset_viewport:Nnnnn { c }
\cs_new_protected:Npn \__box_set_viewport:NnnnnN #1#2#3#4#5#6
(End definition for \box_set_viewport:Nnnn, \box_gset_viewport:Nnnn, and \__box_viewport:NnnnnN. These functions are documented on page 300.)

88.2 Additions to l3flag

⟨@@=flag⟩
It might be faster to just call the “trap” function in all cases but conceptually the function name suggests we should only run it if the flag is zero in case the “trap” made customizable in the future.

```latex
\begin{verbatim}
\flag_raise_if_clear:n
\end{verbatim}
```

(End definition for \flag_raise_if_clear:n. This function is documented on page 301.)

### 88.3 Additions to l3msg

A short-hand used for \texttt{\int_show:n} and similar functions that passes to \texttt{\tl_show:n} the result of applying \texttt{#1} (a function such as \texttt{\int_eval:n}) to the expression \texttt{#2}. The use of \texttt{f}-expansion ensures that \texttt{#1} is expanded in the scope in which the show command is called, rather than in the group created by \texttt{\iow_wrap:nnnN}. This is only important for expressions involving the \texttt{\currentgrouplevel} or \texttt{\currentgrouptype}. On the other hand we want the expression to be converted to a string with the usual escape character, hence within the wrapping code.

```latex
\begin{verbatim}
\msg_show_eval:Nn
\msg_log_eval:Nn
\__msg_show_eval:nnN
\end{verbatim}
```

(End definition for \msg_show_eval:Nn, \msg_log_eval:Nn, and \__msg_show_eval:nnN. These functions are documented on page 302.)

Each item in the variable is formatted using one of the following functions. We cannot use \texttt{\textbackslash{} and so on because these short-hands cannot be used inside the arguments of messages, only when defining the messages.

```latex
\begin{verbatim}
\msg_show_item:n
\msg_show_item_unbraced:n
\msg_show_item:nn
\msg_show_item_unbraced:nn
\end{verbatim}
```

1349
88.4 Additions to \texttt{l3prg}

\begin{verbatim}
\bool_set_inverse:N \bool_set_inverse:c \bool_gset_inverse:N \bool_gset_inverse:c
set to false or true locally or globally.
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \bool_set_inverse:N #1
{ \bool_if:NTF #1 { \bool_set_false:N } { \bool_set_true:N } #1 }

\cs_generate_variant:Nn \bool_set_inverse:N { c }
\cs_new_protected:Npn \bool_gset_inverse:N #1
{ \bool_if:NTF #1 { \bool_gset_false:N } { \bool_gset_true:N } #1 }

\cs_generate_variant:Nn \bool_gset_inverse:N { c }
\end{verbatim}

\begin{verbatim}
\s__bool_mark \s__bool_stop
Internal scan marks.
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \bool_case_true:nTF
{ \exp:w \__bool_case:NnTF \c_true_bool }
\cs_new:Npn \bool_case_true:nT #1#2
{ \exp:w \__bool_case:NnTF \c_true_bool {#1} {#2} { } }
\cs_new:Npn \bool_case_true:nF #1
{ \exp:w \__bool_case:NnTF \c_true_bool {#1} { } }
\cs_new:Npn \bool_case_true:n #1
{ \exp:w \__bool_case:NnTF \c_true_bool {#1} { } { } }
\cs_new:Npn \__bool_case:NnTF #1#2#3#4
{ \bool_if:NTF #1 \__bool_case_true:w \__bool_case_false:w
#1 { } \s__bool_mark {#3} \s__bool_mark {#4} \s__bool_stop }
\cs_new:Npn \__bool_case_true:w #1#2
{ \bool_if:nTF {#1} { \__bool_case_end:nw {#2} } { \__bool_case_true:w } }
\cs_new:Npn \__bool_case_false:w #1#2
{ \bool_if:nTF {#1} { \__bool_case_false:w #1 { } } { \__bool_case_false:w } }
\cs_new:Npn \__bool_case_end:nw #1#2#3#4
{ \\bool_if:NTF \#1 \__bool_case_true:w \__bool_case_false:w
#2 \#1 { } \s__bool_mark \#3 \s__bool_mark \#4 \s__bool_stop }
\cs_new:Npn \__bool_case_end:nw #1#2#3#4
{ \\bool_if:NTF \#1 \__bool_case_true:w \__bool_case_false:w
#2 \#1 { } \s__bool_mark \#3 \s__bool_mark \#4 \s__bool_stop }
\cs_new:Npn \__bool_case_end:nw #1#2#3#4
{ \\bool_if:NTF \#1 \__bool_case_true:w \__bool_case_false:w
#2 \#1 { } \s__bool_mark \#3 \s__bool_mark \#4 \s__bool_stop }
\end{verbatim}

(End definition for \texttt{\bool_set_inverse:N} and \texttt{\bool_gset_inverse:N}. These functions are documented on page \texttt{302}.)
88.5 Additions to \texttt{l3prop}

\texttt{\_prop\_use\_i\_delimit\_by\_s\_stop:nn}

Functions to gobble up to a scan mark.

\texttt{\cs\_new:Npn \_prop\_use\_i\_delimit\_by\_s\_stop:nn #1 #2 #3 \_s\_\_prop\_stop}

(End definition for \texttt{\_prop\_use\_i\_delimit\_by\_s\_stop:nn}.

\texttt{\prop\_rand\_key\_value:N}
\texttt{\prop\_rand\_key\_value:c}
\texttt{\_prop\_rand\_item:w}

Contrarily to \texttt{clist}, \texttt{seq} and \texttt{tl}, there is no function to get an item of a \texttt{prop} given an integer between 1 and the number of items, so we write the appropriate code. There is no bounds checking because \texttt{\int\_rand:nn} is always within bounds. The initial \texttt{\int\_value:w} is stopped by the first \texttt{\_s\_\_prop} in \texttt{#1}.

\texttt{\cs\_new:Npn \prop\_rand\_key\_value:N #1}

\texttt{\prop\_if\_empty:NF #1}

\texttt{\exp\_after:wN \_prop\_rand\_item:w}

\texttt{\int\_value:w \int\_rand:nn \{ 1 \} \{ \prop\_count:N #1 \}}

\texttt{#1 \_s\_\_prop\_stop}

\texttt{\_prop\_rand\_item:w}

\texttt{\cs\_generate\_variant:Nn \prop\_rand\_key\_value:N \{ c \}}

\texttt{\cs\_new:Npn \_prop\_rand\_item:w \#1 \_s\_\_prop \_prop\_pair:wn \#2 \_s\_\_prop \#3}

\texttt{\int\_compare:nNnF \{ #1 \} > 1}

\texttt{\_prop\_use\_i\_delimit\_by\_s\_stop:nn \{ \exp\_not:n \{ #2 \} \{ #3 \} \}}

\texttt{\exp\_after:wN \_prop\_rand\_item:w}

\texttt{\int\_value:w \int\_eval:n \{ #1 - 1 \} \_s\_\_prop}

(End definition for \texttt{\prop\_rand\_key\_value:N} and \texttt{\_prop\_rand\_item:w}. This function is documented on page 303.)

88.6 Additions to \texttt{l3seq}

\texttt{\seq\_map\_thread\_function:NNN}
\texttt{\seq\_map\_thread\_function:NcN}
\texttt{\seq\_map\_thread\_function:ccN}
\texttt{\_seq\_map\_thread\_function:wN}
\texttt{\_seq\_map\_thread\_function:swN}
\texttt{\_seq\_map\_thread\_function:NnNnN}

The idea is to first expand both sequences, adding the usual \{ ? \texttt{\prg\_break:} \} \} to the end of each one. This is most conveniently done in two steps using an auxiliary function. The mapping then throws away the first tokens of \texttt{#2} and \texttt{#5}, which for items in both sequences are \texttt{\_s\_\_seq \_\_seq\_item:n}. The function to be mapped are then be applied to the two entries. When the code hits the end of one of the sequences, the break material stops the entire loop and tidy up. This avoids needing to find the count of the two sequences, or worrying about which is longer.
\cs_new:Npn \seq_mapthread_function:NNN #1\#2\#3
\exp_after:wN \__seq_mapthread_function:wNN #2 \s__seq_stop #1 #3
\cs_new:Npn \__seq_mapthread_function:wNN \s__seq #1 \s__seq_stop #2#3
\exp_after:wN \__seq_mapthread_function:wNw #2 \s__seq_stop #3
\prg_break_point:
\cs_new:Npn \__seq_mapthread_function:wNw \s__seq #1 \s__seq_stop #2
\__seq_mapthread_function:Nnnwnn #2
\prg_break_point:
\cs_new:Npn \__seq_mapthread_function:Nnnwnn #1#2#3#4 \s__seq_stop #5#6
\use_none:n #2
\use_none:n #5
\#1 \{ \#3 \} \{ \#6\}
\__seq_mapthread_function:Nnnwnn \#1 #4 \a__seq_stop
\cs_generate_variant:Nn \seq_mapthread_function:NNN { Nc , c , cc }
\seq_set_filter:NNn
\seq_gset_filter:NNn
\__seq_set_filter:NNNn
\seq_set_from_inline_x:Nnn
\seq_gset_from_inline_x:Nnn
\__seq_set_from_inline_x:NNnn
Similar to \seq_map_inline:Nn, without a \prg_break_point: because the user's code is performed within the evaluation of a boolean expression, and skipping out of that would break horribly. The \__seq_wrap_item:n function inserts the relevant \__seq_item:n without expansion in the input stream, hence in the x-expanding assignment.
\cs_new_protected:Npn \seq_set_filter:NNn
\cs_new_protected:Npn \seq_gset_filter:NNn
\cs_new_protected:Npn \__seq_set_filter:NNNn #1#2#3#4
\__seq_push_item_def:n { \bool_if:nT \#4 \{ \__seq_wrap_item:n \#1 \} }
\#1 \#2 \{ \#3 \}
\__seq_pop_item_def:
\cs_generate_variant:Nn \seq_mapthread_function:NNN { Nc , c , cc }
\seq_set_filter:NNn \seq_gset_filter:NNn \__seq_set_filter:NNNn
\seq_set_from_inline_x:Nnn \seq_gset_from_inline_x:Nnn \__seq_set_from_inline_x:NNnn
Set \__seq_item:n then map it using the loop code.
\cs_new_protected:Npn \seq_set_from_inline_x:Nnn
\cs_new_protected:Npn \seq_gset_from_inline_x:Nnn
\cs_new_protected:Npn \__seq_set_from_inline_x:NNnn
\__seq_push_item_def:n \exp_not:N \__seq_item:n \#4
\#1 \#2 \{ \#3 \} \seq_item:n
\__seq_pop_item_def:
\begin{verbatim}
End definition for \seq_set_from_inline_x:NNn, \seq_gset_from_inline_x:NNn, and \seq_set_from_inline_x:NNn. These functions are documented on page 304.

\seq_set_from_function:NnN
\seq_gset_from_function:NnN

Reuse \seq_set_from_inline_x:NNn.
\cs_new_protected:Npn \seq_set_from_function:NnN #1#2#3
\cs_new_protected:Npn \seq_gset_from_function:NnN #1#2#3

\__seq_set_item:NnnNN
\__seq_set_item:nnNNNN
\__seq_set_item_false:nnNNNN
\__seq_set_item:nNnnNNNN
\__seq_set_item_end:w

\__seq_int_eval:w
\cs_new_eq:NN \__seq_int_eval:w \tex_numexpr:D

The conditionals are distinguished from the Nnn versions by the last argument \use_ii:nn vs \use_i:nn.
\cs_new_protected:Npn \__seq_set_item:NnnNN #1#2#3
\int_compare:nNnTF {#1} > 0
\int_compare:nNnF {#1} > {#2} { \__seq_set_item:nNnnNNNN { #1 - 1 } }
\int_compare:nNnF {#1} < {-#2}
\int_compare:nNnF {#1} = 0
\int_compare:nNnF {#1} = 0
\__seq_set_item_false:nnNNNN {#1} {#2}

Save the item to be stored and evaluate the position and the sequence length only once. Then depending on the sign of the position, check that it is not bigger than the length (in absolute value) nor zero.
\cs_new_protected:Npn \__seq_set_item:Nnn #1#2#3
\exp_args:Nff \__seq_set_item:nnNNNN
\int_compare:nNnTF {#1} > 0
\int_compare:nNnF {#1} > {#2} { \__seq_set_item:nNnnNNNN { #1 - 1 } }
\int_compare:nNnF {#1} < {-#2}
\int_compare:nNnF {#1} < {-#2}
\int_compare:nNnF {#1} = 0
\int_compare:nNnF {#1} = 0
\__seq_set_item_false:nnNNNN {#1} {#2}
\end{verbatim}
If the position is not ok, \_\_seq_set_item_false:nnNNNNN calls an error or returns false (depending on the use_i:nn vs use_ii:nn argument mentioned above).

\begin{verbatim}
cs_new_protected:Npn \_\_seq_set_item_false:nnNNNN #1#2#3#4#5#6
  {
    \msg_error:nxxx { seq } { item-too-large }
    \token_to_str:N #3 \{#2\} \{#1\}
    \prg_return_false:
  }

\msg_new:nnnn { seq } { item-too-large }
  { Sequence-'#1'-does-not-have-an-item-#3 }
  { An-attempt-was-made-to-push-or-pop-the-item-at-position-#3-of-'#1',-but-this-\int_compare:nTF { #3 = 0 } { position-does-not-exist. } { sequence-only-has-#2-item \int_compare:nF { #2 = 1 } {s}. } }

If the position is ok, \_\_seq_set_item:nNnnNNNN makes the assignment and returns true (in the case of conditionals). Here #1 is an integer expression (position minus one), it needs to be evaluated. The sequence #5 starts with \s__seq (even if empty), which stops the integer expression and is absorbed by it. The \if_meaning:w test is slightly faster than an integer test (but only works when testing against zero, hence the offset we chose in the position). When we are done skipping items, insert the saved item \l__seq_internal_a_tl. For put functions the last argument of \_\_seq_set_item_end:w is \use_none:nn and it absorbs the item #2 that we are removing: this is only useful for the pop functions.

\begin{verbatim}
cs_new_protected:Npn \_\_seq_set_item:nNnnNNNN #1#2#3#4#5#6#7#8
  {
    \s__seq
    \exp_after:wN \_\_seq_set_item:wn
    \int_value:w \_\_seq_int_eval:w #1 \s__seq_stop #6
    { } { \prg_return_true: }
  }
cs_new:Npn \_\_seq_set_item:wn #1 \_\_seq_item:n #2
  {
    \s__seq
    \exp_after:wN \_\_seq_set_item:wn
    \int_value:w \_\_seq_int_eval:w #1 - 1 \s__seq
    { } { \prg_return_true: }
  }
cs_new:Npn \_\_seq_set_item_end:w #1 \_\_seq_set_item:wn #1 \_\_seq_item:n #2
  {
    \if_meaning:w 0 #1 \_\_seq_set_item_end:w \fi:
    \exp_not:n \{ \_\_seq_item:n \{#2\} \}
    \exp_after:wN \_\_seq_set_item:wn
    \int_value:w \_\_seq_int_eval:w #1 - 1 \s__seq
  }
cs_new:Npn \_\_seq_set_item_end:v #1 \_\_seq_set_item_end:w #1 \_\_seq_item:n \#2 \#3 \s__seq \#4 \s__seq_stop \#5
  {
    \_\_seq_set_item_end:w #1 \exp_not:n \#2 \#3 \s__seq \#4 \s__seq_stop \#5
    \_\_seq_set_item_end:v #1 \_\_seq_item:n \#2 \#3 \s__seq \#4 \s__seq_stop \#5
    \_\_seq_set_item_end:v #1 \_\_seq_item:n \#2 \#3 \s__seq \#4 \s__seq_stop \#5
  }
cs_new:Npn \_\_seq_set_item_end:v #1 \_\_seq_set_item_end:v #1 \_\_seq_item:o \l__seq_internal_a_tl
  \_\_seq_set_item_end:v #1 \_\_seq_item:o \l__seq_internal_a_tl
\end{verbatim}

1354
The \texttt{NnN} versions simply call the conditionals, for which we will rely on the internals of \texttt{\seq_set_item:Nnn}. The \texttt{NnN} versions simply call the conditionals, for which we will rely on the internals of \texttt{\seq_set_item:Nnn}.

\texttt{\cs_new_protected:Npm \seq_pop_item:NnN #1\#2\#3}
\texttt{\cs_new_protected:Npm \seq_gpop_item:NnN #1\#2\#3}
\texttt{\cs_generate_variant:Nn \seq_pop_item:NnN { c }}
\texttt{\cs_generate_variant:Nn \seq_gpop_item:NnN { c }}
\texttt{\prg_new_protected_conditional:Npnn \seq_pop_item:NnN #1#2#3 { TF , T , F }}
\texttt{\prg_new_protected_conditional:Npnn \seq_gpop_item:NnN #1#2#3 { TF , T , F }}
\texttt{\prg_generate_conditional_variant:Nnn \seq_pop_item:NnN { c } { TF , T , F }}
\texttt{\prg_generate_conditional_variant:Nnn \seq_gpop_item:NnN { c } { TF , T , F }}

Save in \texttt{\l__seq_internal_b_tl} the token list variable \texttt{#3} in which we will store the item. The \texttt{\_\_seq_set_item:nnNNNN} auxiliary eventually inserts \texttt{\_\_seq_internal-a_tl} in place of the item found in the sequence, so we empty that. Instead of the last argument \texttt{\use_i:n} or \texttt{\use_ii:n} used for put functions, we introduce \texttt{\_\_seq_pop_item:nn}, which stores \texttt{\q_no_value} before calling its second argument (\texttt{\prg_return_true:/false:}) to end the conditional. The item found is passed to \texttt{\_\_seq_pop_item-aux:w}, which interrupts the \texttt{x}-expanding sequence assignment and stores the item using the assignment function in \texttt{\_\_seq_internal_b_tl}.

\texttt{\cs_new_protected:Npm \_\_seq_pop_item:nnNNN #1\#2\#3\#4}
\texttt{\_\_seq_set_item:nnNNNN #1 \_\_kernel_tl_set:Nx #3}
\texttt{\exp_args:Nff \_\_seq_set_item:NNNNN #1 \_\_kernel_tl_set:Nx #3}
\texttt{\_\_seq_set_item:nnNNN #1 \_\_kernel_tl_set:Nx #3}
\texttt{\_\_seq_set_item:nnNNN #1 \_\_kernel_tl_set:Nx #3}

(End definition for \texttt{\seq_set_item:NnnTF} and others. These functions are documented on page 305.)
88.7 Additions to \texttt{l3sys}

Various different engines, various different ways to extract the data!

\begin{verbatim}
\cs_after:N \c_sys_engine_version_str
\str_const:Nx \c_sys_engine_version_str
{\str_case:on \c_sys_engine_str
  { \texttt{pdfTeX} }
    \fp_eval:n { round(\int_use:N \tex_pdftexversion:D / 100 , 2) }
    . \tex_pdftexrevision:D
  }
  { \texttt{ptex} }
    \cs_if_exist:NT \tex_ptexversion:D
    p
    \int_use:N \tex_ptexversion:D
    . \int_use:N \tex_ptexminorversion:D
    \tex_ptexrevision:D
    - \int_use:N \tex_uptexversion:D
  }
  { \texttt{luatex} }
    \fp_eval:n { round(\int_use:N \tex_luatexversion:D / 100 , 2) }
    . \tex_luatexrevision:D
  }
  { \texttt{uptex} }
    \cs_if_exist:NT \tex_ptexversion:D
    p
    \int_use:N \tex_ptexversion:D
    . \int_use:N \tex_ptexminorversion:D
    \tex_ptexrevision:D
    u
    \int_use:N \tex_uptexversion:D
    \tex_uptexrevision:D
    - \int_use:N \tex_epTeXversion:D
  }
  { \texttt{xetex} }
    \int_use:N \tex_XeTeXversion:D
\end{verbatim}
88.8 Additions to l3file

\ior_shell_open:Nn
\__ior_shell_open:nN
Actually much easier than either the standard open or input versions! When calling
\__kernel_ior_open:Nn the file the pipe is added to signal a shell command, but the
quotes are not added yet—they are added later by \__kernel_file_name_quote:n.

\cs_new_protected:Npn \ior_shell_open:Nn #1#2
  \SYS_IF_SHELL:TF
  \exp_args:No \__ior_shell_open:nN { \tl_to_str:n {#2} } #1
  \msg_error:nn { ior } { pipe-failed }
\}
\cs_new_protected:Npn \__ior_shell_open:nN #1#2
  \tl_if_in:nnTF {#1} { " }
  \msg_error:nnx
  \{ ior \} \quote-in-shell \{#1\}
\}
\msg_new:nnnn { ior } { pipe-failed }
\{ Cannot-run-piped-system-commands. \}
\{ LaTeX-tried-to-call-a-system-process-but-this-was-not-possible.\}
\msg_new:n { \__kernel_ior_open:Nn #2 \{ \#1 \} }
\msg_new:n { ior \} \quote-in-shell \{#1\}
\}
\msg_new:n { \__kernel_ior_open:Nn #2 \{ \#1 \} }
\msg_new:n { ior \} \quote-in-shell \{#1\}
\}
\msg_new:n { \__kernel_ior_open:Nn #2 \{ \#1 \} }
\msg_new:n { ior \} \quote-in-shell \{#1\}
\}
\msg_new:n { \__kernel_ior_open:Nn #2 \{ \#1 \} }
\msg_new:n { ior \} \quote-in-shell \{#1\}
\}
\msg_new:n { \__kernel_ior_open:Nn #2 \{ \#1 \} }
\msg_new:n { ior \} \quote-in-shell \{#1\}
\}
(End definition for \ior_shell_open:Nn and \__ior_shell_open:nN. This function is documented on
page 301.)

88.9 Additions to l3tl

88.9.1 Building a token list

\tl_build_begin:N {tl var} and \tl_build_end:N {tl var}, the {tl var} has the structure
\exp_end: ... \exp_end: \__tl_build_last:NNn {assignment} {next tl}
{(left)}{(right)}
where (right) is not braced. The “data” it represents is (left) followed by the “data” of
(next tl) followed by (right). The (next tl) is a token list variable whose name is that of
(tl var) followed by ’. There are between 0 and 4 \exp_end: to keep track of when (left)
and \langle right \rangle should be put into the \langle next tl \rangle. The \langle assignment \rangle is \cs_set_nopar:Npx if the variable is local, and \cs_gset_nopar:Npx if it is global.

First construct the \langle next tl \rangle: using a prime here conflicts with the usual expl3 convention but we need a name that can be derived from \#1 without any external data such as a counter. Empty that \langle next tl \rangle and setup the structure. The local and global versions only differ by a single function \cs_(g)set_nopar:Npx used for all assignments: this is important because only that function is stored in the \langle tl var \rangle and \langle next tl \rangle for subsequent assignments. In principle \_\_tl_build_begin:NNN could use \tl_(g)clear_new:N to empty \#1 and make sure it is defined, but logging the definition does not seem useful so we just do \#3 \#1 {} to clear it locally or globally as appropriate.

\cs_new_protected:Npn \tl_build_begin:N \#1
\cs_new_protected:Npn \tl_build_gbegin:N \#1
\cs_new_protected:Npn \_\_tl_build_begin:NN \#1\#2
\cs_new_protected:Npn \_\_tl_build_begin:NNN \#1\#2\#3

\cs_new_eq:NN \tl_build_clear:N \tl_build_begin:N
\cs_new_eq:NN \tl_build_gclear:N \tl_build_gbegin:N

\cs_new_protected:Npn \_\_tl_build_begin:NNN \#1\#2\#3
\exp_not:n { \exp_end: \exp_end: \exp_end: \exp_end: }
\exp_not:n { \__tl_build_last:NNn \#3 \#1 { } }

\end definition for \_\_tl_build_begin:NNN. These functions are documented on page 306.

The \begin \and \gbegin \functions already clear enough to make the token list variable effectively empty. Eventually the begin \and gbegin \functions should check that \#1' is empty or undefined, while the \clear \and gclear \functions ought to empty \#1', \#1'' and so on, similar to \_\_tl_build_end:N. This only affects memory usage.

\_\_tl_build_put_right:Nn
\_\_tl_build_put_right:Nx
\_\_tl_build_gput_right:Nn
\_\_tl_build_gput_right:Nx
\__tl_build_last:NNn
\__tl_build_put:nn
\__tl_build_put:nw

Similar to \_\_tl_put_right:Nn, but apply \exp:w to \#1. Most of the time this just removes one \\exp_end:. When there are none left, \_\_tl_build_last:NNn is expanded instead. It resets the definition of the \langle tl var \rangle by ending the \exp_not:n and the definition early. Then it makes sure the \langle next tl \rangle (its argument \#1) is set-up and starts a new definition. Then \_\_tl_build_put:nn \and \_\_tl_build_put:nnw place the \langle left \rangle part of the original \langle tl var \rangle as appropriate for the definition of the \langle next tl \rangle (the \langle right \rangle part is left in the right place without ever becoming a macro argument). We use \\exp_after:wN rather than some \exp_args:No to avoid reading arguments that are likely very long token lists. We use \cs_(g)set_nopar:Npx rather than \_\_tl_(g)set:Nx partly for the same reason and partly because the assignments are interrupted by brace tricks, which implies that the assignment does not simply set the token list to an x-expansion of the second argument.
See \texttt{\_\_tl_build_last:Nnn} for all the machinery. We could easily provide \texttt{\_\_tl_build_put_left_right:Nnn}, by just add the \texttt{⟨right⟩} material after the \texttt{⟨left⟩} in the x-expanding assignment.
The idea is to expand the ⟨tl var⟩ then the ⟨next tl⟩ and so on, all within an x-expanding assignment, and wrap as appropriate in \exp_not:n. The various ⟨left⟩ parts are left in
the assignment as we go, which enables us to expand the ⟨next tl⟩ at the right place. The various ⟨right⟩ parts are eventually picked up in one last \exp_not:n, with a brace trick to wrap all the ⟨right⟩ parts together.

Get the data then clear the ⟨next tl⟩ recursively until finding an empty one. It is perhaps wasteful to repeatedly use \cs_to_sr:N. The local/global scope is checked by \tl_set:Nx or \tl_gset:Nx.

(End definition for \tl_build_get:NN, \_\_\_tl_build_end:W. These functions are documented on page 307.)

(End definition for \tl_build_end:N, \tl_build_gend:N, and \_\_\_tl_build_end_loop:NN. These functions are documented on page 307.)
88.9.2 Other additions to \texttt{l3tl}

For the braced version \texttt{\_\_tl\_range\_braced:w} sets up \texttt{\_\_tl\_range\_collect\_braced:w} which stores items one by one in an argument after the semicolon. The unbraced version is almost identical. The version preserving braces and spaces starts by deleting spaces before the argument to avoid collecting them, and sets up \texttt{\_\_tl\_range\_collect:nnn} with a first argument of the form \{(\texttt{collected})\} \{\texttt{tokens}\}\}, whose head is the collected tokens and whose tail is what remains of the original token list. This form makes it easier to move tokens to the \{(\texttt{collected})\} tokens.

\begin{Verbatim}
\cs_new:Npn \tl_range_braced:Nnn { \exp_args:No \tl_range_braced:nnn }
\cs_generate_variant:Nn \tl_range_braced:Nnn { c }
\cs_new:Npn \tl_range_braced:nnn { \__tl_range:Nnnn \__tl_range_braced:w }
\cs_new:Npn \tl_range_unbraced:Nnn { \exp_args:No \tl_range_unbraced:nnn }
\cs_generate_variant:Nn \tl_range_unbraced:Nnn { c }
\cs_new:Npn \tl_range_unbraced:nnn { \__tl_range:Nnnn \__tl_range_unbraced:w }
\cs_new:Npn \__tl_range_braced:w #1 ; #2
{ \__tl_range_collect_braced:w #1 ; { } #2 }
\cs_new:Npn \__tl_range_unbraced:w #1 ; #2
{ \__tl_range_collect_unbraced:w #1 ; { } #2 }
\cs_new:Npn \__tl_range_collect_braced:w #1 ; #2#3
{ \if_int_compare:w #1 > \c_one_int
  \exp_after:wN \__tl_range_collect_braced:w
  \int_value:w \int_eval:n { #1 - 1 } \exp_after:wN ;
  \fi:
  { #2 {#3} }
}
\cs_new:Npn \__tl_range_collect_unbraced:w #1 ; #2#3
{ \if_int_compare:w #1 > \c_one_int
  \exp_after:wN \__tl_range_collect_unbraced:w
  \int_value:w \int_eval:n { #1 - 1 } \exp_after:wN ;
  \fi:
  { #2 #3 }
}
\end{Verbatim}

(End definition for \texttt{tl\_range\_braced:Nnn} and others. These functions are documented on page 306.)

88.10 Additions to \texttt{l3token}

\texttt{\_c\_catcode\_active\_space\_tl}

While \texttt{\char\_generate:nn} can produce active characters in some engines it cannot in general. It would be possible to simply change the catcode of space but then the code would need to avoid all spaces, making it quite unreadable. Instead we use the primitive \texttt{\tex\_lowercase:D} trick.

\begin{Verbatim}
\cs_new:Npn \tl_range_braced:Nnn { \exp_args:No \tl_range_braced:nnn }
\cs_generate_variant:Nn \tl_range_braced:Nnn { c }
\cs_new:Npn \tl_range_braced:nnn { \__tl_range:Nnnn \__tl_range_braced:w }
\cs_new:Npn \tl_range_unbraced:Nnn { \exp_args:No \tl_range_unbraced:nnn }
\cs_generate_variant:Nn \tl_range_unbraced:Nnn { c }
\cs_new:Npn \tl_range_unbraced:nnn { \__tl_range:Nnnn \__tl_range_unbraced:w }
\cs_new:Npn \__tl_range_braced:w #1 ; #2
{ \__tl_range_collect_braced:w #1 ; { } #2 }
\cs_new:Npn \__tl_range_unbraced:w #1 ; #2
{ \__tl_range_collect_unbraced:w #1 ; { } #2 }
\cs_new:Npn \__tl_range_collect_braced:w #1 ; #2#3
{ \if_int_compare:w #1 > \c_one_int
  \exp_after:wN \__tl_range_collect_braced:w
  \int_value:w \int_eval:n { #1 - 1 } \exp_after:wN ;
  \fi:
  { #2 {#3} }
}
\cs_new:Npn \__tl_range_collect_unbraced:w #1 ; #2#3
{ \if_int_compare:w #1 > \c_one_int
  \exp_after:wN \__tl_range_collect_unbraced:w
  \int_value:w \int_eval:n { #1 - 1 } \exp_after:wN ;
  \fi:
  { #2 #3 }
}
\end{Verbatim}
Most of the work is done by `\_peek_execute_branches...`, which calls either `\_peek_true:w` or `\_peek_false:w` according to whether the next token `\l_peek_token` matches the search token (stored in `\l_peek_search_token` and `\l_peek_search_tl`). Here, in the `true` case we run `\_peek_collect_true:w`, which generally calls `\_peek_collect:N` to store the peeked token into `\l_peek_collect_tl`, except in special non-N-type cases (begin-group, end-group, or space), where a frozen token is stored. The `true` branch calls `\_peek_execute_branches...:` to fetch more matching tokens. Once there are no more, `\_peek_false_aux:n` closes the safe-align group and runs the user’s inline code.

(End definition for `\c_catcode_active_space_tl`. This variable is documented on page 307.)
\cs_new_protected:Npn \__peek_collect:N #1
\tl_put_right:Nn \l__peek_collect_tl {#1} \__peek_true_aux:w
\cs_new_protected:Npn \__peek_collect_remove:nw #1
\tl_put_right:Nn \l__peek_collect_tl {#1} \exp_after:wN \__peek_true_remove:w

(End definition for \peek_catcode_collect_inline:Nn and others. These functions are documented on page 308.)

{/package}
Chapter 89

\texttt{l3deprecation implementation}

89.1 Helpers and variables

\texttt{\_\_deprecation_grace_period.bool} This is set to \texttt{true} when the deprecated command that is being defined is in its grace period, meaning between the time it becomes an error by default and the time 6 months later where even undo-recent-deprecations stops restoring it.

\[
\texttt{\_\_deprecation_grace_period.bool}
\]

\texttt{\_\_deprecation_mark} Internal scan marks.

\texttt{\_\_deprecation_stop} (End definition for \texttt{\_\_deprecation_mark} and \texttt{\_\_deprecation_stop}.)

\texttt{\_deprecation_date_compare:nNnTF} Expects \#1 and \#3 to be dates in the format YYYY-MM-DD (but accepts YYYY or YYYY-MM too, filling in zeros for the missing data). Compares them using \#2 (one of \texttt{<}, \texttt{=} or \texttt{>}).

\[
\texttt{\_deprecation_date_compare:nNnTF}
\]

(End definition for \texttt{\_deprecation_date_compare:nNnTF} and \texttt{\_deprecation_date_compare_aux:w}.)

\[
\texttt{\_deprecation_date_compare:nNnTF}
\]

1364
\_deprecation_not_yet_deprecated:nTF
\_deprecation_minus_six_months:w

Receives a deprecation \textlangle date\textrangle and runs the \textit{true} (false) branch if the expl3 date is earlier (later) than \textlangle date\textrangle. If \texttt{undo-recent-deprecations} is used we subtract 6 months to the expl3 date (equivalently add 6 months to the \textlangle date\textrangle). In addition, if the expl3 date is between \textlangle date\textrangle and \textlangle date\textrangle plus 6 months, \texttt{l\_deprecation_grace_period_bool} is set to \textit{true}, otherwise \textit{false}.

\cs_new_protected:Npn \__deprecation_not_yet_deprecated:nTF #1
\begin{verbatim}
{ \bool_set_false:N \l__deprecation_grace_period_bool
  \exp_args:No \__deprecation_date_compare:nNnTF { \ExplLoaderFileDate } < {#1}
  { \use_i:nn }
  \exp_args:Nf \__deprecation_date_compare:nNnTF
  { \exp_after:wN \__deprecation_minus_six_months:w
    \ExplLoaderFileDate -0-0- \s__deprecation_stop
  } < {#1}
  { \bool_set_true:N \l__deprecation_grace_period_bool
    \bool_if:NTF \g__kernel_deprecation_undo_recent_bool
    { \use_ii:nn }
  }
}

\cs_new:Npn \__deprecation_minus_six_months:w #1 - #2 - #3 - #4 \s__deprecation_stop
\begin{verbatim}
{ \int_compare:nNnTF {#2} > 6
  { \int_eval:n { \#1 - \int_eval:n { #2 - 6 } - \#3 }
    \int_eval:n { \#1 - 1 } - \int_eval:n { #2 + 6 } - \#3 }
}
\end{verbatim}
\end{verbatim}

(End definition for \_deprecation_not_yet_deprecated:nTF and \_deprecation_minus_six_months:w.)

89.2 Patching definitions to deprecate

\_kernel_patch_deprecation:nnNNpn \{\textlangle date\textrangle\} \{\textlangle replacement\textrangle\} \{\texttt{definition}\}
\{\texttt{function}\} \{\texttt{parameters}\} \{\texttt{(code)}\}

defines the \texttt{\textlangle function\textrangle} to produce a warning and run its \texttt{\textlangle code\textrangle}, or to produce an error and not run any \texttt{\textlangle code\textrangle}, depending on the expl3 date.

- If the expl3 date is less than the \textlangle date\textrangle (plus 6 months in case \texttt{undo-recent-deprecations} is used) then we define the \texttt{\textlangle function\textrangle} to produce a warning and run its code. The warning is actually suppressed in two cases:
  - if neither \texttt{undo-recent-deprecations} nor \texttt{enable-debug} are in effect we may be in an end-user’s document so it is suppressed;
  - if the command is expandable then we cannot produce a warning.
• Otherwise, we define the ⟨function⟩ to produce an error.

In both cases we additionally make \debug_on:n {deprecation} turn the ⟨function⟩ into an \outer error, and \debug_off:n {deprecation} restore whatever the behaviour was without \debug_on:n {deprecation}.

In later sections we use the l3doc key deprecated with a date equal to that ⟨date⟩ plus 6 months, so that l3doc will complain if we forget to remove the stale ⟨parameters⟩ and ⟨code⟩.

In the explanations below, ⟨definition⟩ ⟨function⟩ ⟨parameters⟩ ⟨(code)⟩ or assignments that only differ in the scope of the ⟨definition⟩ will be called “the standard definition”.

\__kernel_patch_deprecation:nnNNpn
\__deprecation_patch_aux:nnNNnn
\__deprecation_warn_once:nnNnn
\__deprecation_just_error:nnNN

(The parameter text is grabbed using #5#.) The arguments of \__kernel_deprecation_code:nn are run upon \debug_on:n {deprecation} and \debug_off:n {deprecation}, respectively. In both scenarios we the ⟨function⟩ may be \outer so we undefine it with \tex_let:D before redefining it, with \__kernel_deprecation_error:Nnn or with some code added shortly.

Then check the date (taking into account undo-recent-deprecations) to see if the command should be deprecated right away (false branch of \__deprecation_not_yet_deprecated:nTF), in which case \__deprecation_just_error:nnNN makes ⟨function⟩ into an error (not \outer), ignoring its ⟨parameters⟩ and ⟨code⟩ completely.

Otherwise distinguish cases where we should give a warning from those where we shouldn’t: warnings can only happen for protected commands, and we only want them if either undo-recent-deprecations or enable-debug is in force, not for standard users.

\cs_new_protected:Npn \__kernel_patch_deprecation:nnNNpn #1#2#3#4#5#
\cs_new_protected:Npn \__deprecation_patch_aux:nnNNnn #1#2#3#4#5#6
\cs_new_protected:Npn \__deprecation_not_yet_deprecated:nTF {#1}
\cs_new_protected:Npn \__deprecation_code:nn
\cs_new_protected:Npn \__deprecation_error:Nnn #4 {#2} {#1}
\cs_new_protected:Npn \__deprecation_warn_once:nnNnn {#1} {#2} #4 {#5} {#6}
\cs_new_protected:Npn \__deprecation_just_error:nnNN {#1} {#2} #3 #4

In case we want a warning, the ⟨function⟩ is defined to produce such a warning without grabbing any argument, then redefine itself to the standard definition that the ⟨function⟩ should have, with arguments, and call that definition. The x-type expansion and \exp_not:n avoid needing to double the #, which we could not do anyways. We then deal with
the code for \debug_off:n \{deprecation\}: presumably someone doing that does not need the warning so we simply do the standard definition.

\cs_new_protected:Npn \__deprecation_warn_once:nnNNN #1#2#3#4#5
\cs_gset_protected:Npx #3
\__kernel_if_debug:TF
{\exp_not:N \msg_warning:nnxxx
 { deprecation } \{ deprecated-command \}
 {#1}
 { \token_to_str:N #3 }
 { \tl_to_str:n {#2} }
}
{ }
\exp_not:n { \cs_gset_protected:Npn #3 #4 {#5} }
\exp_not:N #3
\__kernel_deprecation_code:nn { }
\cs_set_protected:Npn #3 #4 {#5} }

In case we want neither warning nor error, the function is given its standard definition. Here \#1 is \cs_new:Npn or \cs_new_protected:Npn and \#2 is \function \langle \parameters \rangle \{ \langle \code \rangle \}, so \#1\#2 performs the assignment. For \debug_off:n \{deprecation\} we want to use the same assignment but with a different scope, hence the \cs_if_eq:NNTF test.

\cs_new_protected:Npn \__deprecation_patch_aux:Nn #1#2
\exp_args:NNx \__deprecation_patch_aux:Nn #3
{ \exp_not:N #4
{ \cs_if_eq:NNTF #3 \cs_gset_protected:Npn
{ \exp_not:N \msg_error:nnnnnn }
{ \exp_not:N \msg_expandable_error:nnnnnn }
 { deprecation } \{ deprecated-command \}
 {#1}
 { \token_to_str:N #4 }
 { \tl_to_str:n {#2} }
 { \bool_if:NT \l__deprecation_grace_period_bool { grace } }
}

Finally, if we want an error we reuse the same \__deprecation_patch_aux:Nn as the previous case. Indeed, we want \debug_off:n \{deprecation\} into an error, just like it is by default. The error is expandable or not, and the last argument of the error message is empty or is grace to denote the case where we are in the 6 month grace period, in which case the error message is more detailed.

\cs_new_protected:Npn \__deprecation_just_error:nnNN #1#2#3#4
\exp_args:NNn \__deprecation_just_error:nnNN #3
\exp_not:N \__deprecation_just_error:nnNN #4
{ \cs_if_eq:NNTF \cs_gset_protected:Npn
{ \exp_not:N \msg_error:nnnnnn }
{ \exp_not:N \msg_error:nnnnnn }
 { deprecation } \{ deprecated-command \}
 {#1}
 { \token_to_str:N #4 }
 { \tl_to_str:n {#2} }
 { \bool_if:NT \l__deprecation_grace_period_bool { grace } }
}
\_kernel\_deprecation\_error:Nnn

The \texttt{outer} definition here ensures the command cannot appear in an argument. Use this auxiliary on all commands that have been removed since 2015.

\begin{verbatim}
\cs_new_protected:Npn \__kernel_deprecation_error:Nnn #1#2#3
\{
  \tex_protected:D \tex_outer:D \tex_edef:D #1
  \exp_not:N \msg_expandable_error:nnnn
    \{ deprecated-command \}
    \{ \tl_to_str:n {#3} \} \{ \token_to_str:N #1 \} \{ \tl_to_str:n {#2} \}
  \exp_not:N \msg_error:nnxxx
    \{ deprecated-command \}
    \{ \tl_to_str:n {#3} \} \{ \token_to_str:N #1 \} \{ \tl_to_str:n {#2} \}
\}
\end{verbatim}

\begin{verbatim}
(End definition for \__kernel\_deprecation\_error:Nnn.)
\end{verbatim}

\msg_new:nnn { deprecation } { deprecated-command }

\begin{verbatim}
{ \tl_if_blank:nF {#3} { Use- \tl_trim_spaces:n {#3} -not- } \#2-deprecated-on-\#1. \str_if_eq:nnT {#4} { grace } \{ \c_space_tl \For-6-months-after-that-date-one-can-restore-a-deprecated-command-by-loading-the-expl3-package-with-the-option-'undo-recent-deprecations'. \}
\}
\end{verbatim}

\subsection{89.3 Removed functions}

\begin{verbatim}
\__deprecation\_old\_protected:Nnn \__deprecation\_old:Nnn
\begin{verbatim}
\cs_new_protected:Npn \__deprecation\_old\_protected:Nnn #1#2#3
\{
  \__kernel\_patch\_deprecation:nnNNpn {#3} {#2}
  \cs_gset_protected:Npn #1 { }
\}
\cs_new_protected:Npn \__deprecation\_old:Nnn #1#2#3
\{
  \__kernel\_patch\_deprecation:nnNNpn {#3} {#2}
  \cs_gset:Npn #1 { }
\}
\__deprecation\_old\_protected:Nnn \box\_resize:Nnn
\begin{verbatim}
\box\_resize\_to\_wd\_and\_ht\_plus\_dp:Nnn \cs_gset\_protected:Npn \#1 \{ \}
\cs_new_protected:Npn \__deprecation\_old\_protected\_Nnn #1#2#3
\{
  \__kernel\_patch\_deprecation\_Nnn {#3} {#2}
  \cs_gset\_Nnn \#1 \{ \}
\}
\cs_new_protected:Npn \__deprecation\_old\_Nnn \box\_use\_clear\_N
\begin{verbatim}
\box\_use\_clear\_N \cs_gset\_Nnn \#1 \{ \}
\end{verbatim}
\cs_new_protected:Npn \__deprecation\_old\_Nnn \c\_job\_name\_tl
\begin{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}

1368
\{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \} \{ \texttt{\_\_deprecation\_old\_Nnn} \}
\__deprecation_old_protected:Nnn \file_if_exist:nTF \file_input:n \file_list: \seq_put_right:Nn \l_file_search_path_seq \l_file_search_path_seq \seq_remove_all:Nn \l_file_search_path_seq \g_file_current_name_tl \int_case:nnn \int_from_binary:n \int_from_hexadecimal:n \int_from_octal:n \int_to_binary:n \int_to_hexadecimal:n \int_to_octal:n \ior_get_str:NN \ior_list_streams: \ior_log_streams: \ior_show_list: \ior_log_list: \int_case:nnF \int_from_binary:n \int_from_hexadecimal:n \int_from_octal:n \int_to_binary:n \int_to_hexadecimal:n \int_to_octal:n \msginterrupt:nnn \lua_escape_x:n \lua_now:x:n \lua_escape:e \lua_now:e \lua_shipout_x:n \lua_shipout:e:n \lua_tex_if_engine:p \lua_tex_if_engine:F \lua_tex_if_engine:T \msginterrupt:nnn
{ [Defined-error-message] } \msglog:n \msg_term:n
\__deprecation_old:Nnn \pdftex_if_engine_p: { 2017-01-01 }
\__deprecation_old:Nnn \pdftex_if_engine:F { 2017-01-01 }
\__deprecation_old:Nnn \pdftex_if_engine:T { 2017-01-01 }
\__deprecation_old:Nnn \pdftex_if_engine:TF { 2017-01-01 }
\__deprecation_old:Nnn \prop_get:cn { 2016-01-05 }
\__deprecation_old:Nnn \prop_get:Nn { 2016-01-05 }
\__deprecation_old:Nnn \quark_if_recursion_tail_break:N { 2015-07-14 }
\__deprecation_old:Nnn \quark_if_recursion_tail_break:n { 2015-07-14 }
\__deprecation_old:Nnn \sort_ordered: { \sort_return_same: } { 2019-01-01 }
\__deprecation_old:Nnn \sort_reversed: { \sort_return_swapped: } { 2019-01-01 }
\__deprecation_old:Nnn \str_case:nnn { \str_case:nnF } { 2015-07-14 }
\__deprecation_old:Nnn \str_case:onn { \str_case:onF } { 2015-07-14 }
\__deprecation_old:Nnn \str_case_x:nn { \str_case_e:nn } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nnn { \str_case_e:nnF } { 2015-07-14 }
\__deprecation_old:Nnn \str_case_x:xnn { \str_case_e:xnn } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nnF { \str_case_e:nnF } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nnT { \str_case_e:nnT } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nnTF { \str_case_e:nnTF } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nF { \str_case_e:nF } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nT { \str_case_e:nT } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nTF { \str_case_e:nTF } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nnF { \str_case_e:nF } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nnT { \str_case_e:nT } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nFF { \str_case_e:nFF } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nnTF { \str_case_e:nTF } { 2020-01-01 }
\__deprecation_old:Nnn \str_case_x:nTF { \str_case_e:nTF } { 2020-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nn { \str_if_eq:pp } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nnn { \str_if_eq:ppn } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nnT { \str_if_eq:ppT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nFF { \str_if_eq:ppFF } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nTF { \str_if_eq:ppTF } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nTF { \str_if_eq:ppTF } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nT { \str_if_eq:ppT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nF { \str_if_eq:ppF } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nF { \str_if_eq:ppF } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nT { \str_if_eq:ppT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nFF { \str_if_eq:ppFF } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nTT { \str_if_eq:ppTT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nTT { \str_if_eq:ppTT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nFF { \str_if_eq:ppFF } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nFT { \str_if_eq:ppFT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nFT { \str_if_eq:ppFT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nT { \str_if_eq:ppT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nF { \str_if_eq:ppF } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nF { \str_if_eq:ppF } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nT { \str_if_eq:ppT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nFF { \str_if_eq:ppFF } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nTT { \str_if_eq:ppTT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nTT { \str_if_eq:ppTT } { 2001-01-01 }
\__deprecation_old:Nnn \str_if_eq_x_p:nFF { \str_if_eq:ppFF } { 2001-01-01 }
\__deprecation_old_protected:Nnn \tl_analysis_show:N { 2020-01-01 }
\__deprecation_old_protected:Nnn \tl_analysis_show:n { 2020-01-01 }
\__deprecation_old:Nnn \tl_case:cn { 2015-07-14 }
\__deprecation_old:Nnn \tl_case:cnF { 2015-07-14 }
\__deprecation_old:Nnn \tl_case:Nn { 2015-07-14 }
\__deprecation_old:Nnn \tl_case:NnF { 2015-07-14 }
\__deprecation_old_protected:Nnn \tl_to_lowercase:n { 2015-07-14 }
89.4 Loading the patches

When loaded first, the patches are simply read here. Here the deprecation code is loaded with the lower-level `\__kernel_...` macro because we don’t want it to flip the `\g__sys_deprecation_bool` boolean, so that the deprecation code can be re-loaded later (when using `undo-recent-deprecations`).

\begin{verbatim}
\group_begin:
\cs_set_protected:Npn \ProvidesExplFile { \char_set_catcode_space:n { \ } \ProvidesExplFileAux }
\cs_gset_protected:Npn \__kernel_sys_configuration_load:n #1 \file_input:n { #1 .def } \__kernel_sys_configuration_load:n { l3deprecation }
\end{verbatim}

Standard file identification.

\begin{verbatim}
\ProvidesExplFile{l3deprecation.def}{2019-04-06}{L3 Deprecated functions}
\end{verbatim}

89.5 Deprecated l3str functions

\begin{verbatim}
\str_lower_case:n \str_lower_case:f \str_upper_case:n \str_upper_case:f \str_fold_case:n \str_fold_case:V
\end{verbatim}

1372
This command was made internal, with one more argument. There is no easy way to compute a reasonable value for that extra argument so we take a value that is big enough to accommodate all of Unicode.

\texttt{\tl_lower_case:n} \texttt{\tl_lower_case:nn} \texttt{\tl_upper_case:n} \texttt{\tl_upper_case:nn} \texttt{\tl_mixed_case:n} \texttt{\tl_mixed_case:nn}

\texttt{\text_lowercase:n} \texttt{\text_lowercase:nn} \texttt{\text_uppercase:n} \texttt{\text_uppercase:nn} \texttt{\text_titlecase:n} \texttt{\text_titlecase:nn}

\texttt{\seq_map Indexed Inline:Nn} \texttt{\seq_map Function:NN}

\texttt{\tl_lower_case:n} \texttt{\tl_upper_case:n} \texttt{\tl_mixed_case:n}
89.8 Deprecated \l3token functions

(End definition for \textcase and others.)

(End definition for \charcase and others.)

⟨/patches⟩

⟨/package⟩
Index

The italic numbers denote the pages where the corresponding entry is described, numbers underlined point to the definition, all others indicate the places where it is used.

Symbols

! 251
" 18469, 18472, 29163, 31501, 31525, 31531, 31535, 31541, 31545, 31551, 31557, 31564, 31565, 31571, 31575, 31577, 31686
\# 10308, 13590, 18469, 31370
\$ 5358, 13569, 18469, 18472, 31370
% 10310, 13591, 18469, 31370
& 9115, 31582, 18469, 18472
&& 250
\ 29163, 31495, 31522, 31529, 31533, 31538, 31543, 31546, 31548, 31555, 31560, 31561, 31568, 31573, 31576, 31584, 31585, 31632, 31633, 31640, 31641, 31652, 31653, 31658, 31659, 31687, 31688, 31710, 31711, 31714, 31715, 31716, 31717
( 10810, 10996, 11071, 29184
) 29184
* 251
* 13275, 13298, 18846, 18848, 18852, 18860
** 251
+ 250, 251
, 20416, 31382
- 250, 251
- 190
, 29163, 31500, 31588, 31589, 31598, 31599, 31608, 31609, 31626, 31638, 31639, 31689, 31690, 31720, 31721, 31724, 31725
/ 251
/ 189, 4264
/ 13588
/\ 42, 379, 406, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2276, 2280, 2281, 2284, 2287, 2294, 2297, 2300, 2306, 2393, 2469, 2471, 2476, 2483, 2487, 2490, 2545, 2510, 2551, 2552, 2553, 2554, 2565
/\\ 42, 2272, 2393, 2554
/\\ 42, 2300
/\unbraced 42, 2408
/\ 42, 2274
/\unbraced 42, 2408
/\ 42, 2274
/\ 42, 2274
/\unbraced 42, 2274
/\unbraced 42, 2108, 2510
/: error 81
/: f 42, 2287, 2553
/:f_unbraced 42, 2408
/:n 42, 776, 2271, 2551, 2554
/:o 42, 2276, 2552
/:o_unbraced 42, 2408, 2552, 2553, 2554
/:p 42, 379, 2273
/:v 42, 2300
/:v_unbraced 42, 2408
/:x 42, 2294
/:x_unbraced 42, 2408, 2565
< < 250
= 250
= 250
= 20417, 29163, 31498, 31578, 31579, 31594, 31595, 31617, 31618, 31619, 31646, 31647, 31672, 31673, 31726, 31727
> > 250
? ? 250
?: 250
??? 603
:\ 2186, 2685, 3088, 3689, 3713, 3714, 3721, 3722, 4209, 4299, 4710, 4711, 6074, 6081, 6082, 6083, 6207, 7856, 7862, 7867, 7901, 7910, 7914, 7919, 7939, 7941, 7942, 7944, 7947, 7949, 7954, 7956, 7958, 7963, 7967, 7970, 7974, 7976, 7980, 7982, 7988, 7990, 7994, 7996, 8000, 8005, 8007, 8049, 8051, 8056, 8058, 8064, 8069, 8070, 8074, 8078, 8088, 8091, 8095, 8096, 8100, 8108, 8172, 9018, 9036, 9038, 9043, 9044, 9068, 9078, 9085, 9100, 9507, 9515, 9522, 9534, 9535, 9550, 9551, 9558, 9572, 9575, 9576, 9608, 9636, 9669, 9670, 9683, 9731, 9732, 9740, 9747, 9748, 9770, 9786, 9790, 9795, 9802, 10313, 11312, 11316, 11317, 11319, 11325, 11327, 11332, 11333, 11335, 11336, 11338, 11340, 11352, 11362, 11364, 11365, 11366, 11384, 14138, 14139, 14142, 14464, 14467, 14468, 14469, 14470, 14475, 14481, 14486, 14493, 14562, 14655, 14656, 14657, 14659, 1375
<table>
<thead>
<tr>
<th>A</th>
<th>\A</th>
<th>13276, 13299</th>
</tr>
</thead>
<tbody>
<tr>
<td>\AA</td>
<td>29167, 31062, 31419</td>
<td></td>
</tr>
<tr>
<td>\aa</td>
<td>29167, 31062, 31429</td>
<td></td>
</tr>
<tr>
<td>above</td>
<td>191</td>
<td></td>
</tr>
<tr>
<td>abovedisplayshortskip</td>
<td>192</td>
<td></td>
</tr>
<tr>
<td>abovedisplayskip</td>
<td>193</td>
<td></td>
</tr>
<tr>
<td>abovewithdelims</td>
<td>194</td>
<td></td>
</tr>
<tr>
<td>abs</td>
<td>251</td>
<td></td>
</tr>
<tr>
<td>\accent</td>
<td>195</td>
<td></td>
</tr>
<tr>
<td>acos</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>acosd</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>acot</td>
<td>254</td>
<td></td>
</tr>
<tr>
<td>acotd</td>
<td>254</td>
<td></td>
</tr>
<tr>
<td>acsc</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>acscd</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>\adjdemerits</td>
<td>196</td>
<td></td>
</tr>
<tr>
<td>\adjustspacing</td>
<td>909</td>
<td></td>
</tr>
<tr>
<td>\advance</td>
<td>197</td>
<td></td>
</tr>
<tr>
<td>\AE</td>
<td>29168, 31063, 31420, 31714</td>
<td></td>
</tr>
<tr>
<td>\ae</td>
<td>29168, 31063, 31430, 31715</td>
<td></td>
</tr>
<tr>
<td>\afterassignment</td>
<td>198</td>
<td></td>
</tr>
<tr>
<td>\aftergroup</td>
<td>199</td>
<td></td>
</tr>
<tr>
<td>\alignmark</td>
<td>784</td>
<td></td>
</tr>
<tr>
<td>\aligntab</td>
<td>785</td>
<td></td>
</tr>
<tr>
<td>asec</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>asecd</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>asin</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>asind</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>atan</td>
<td>254</td>
<td></td>
</tr>
<tr>
<td>atand</td>
<td>254</td>
<td></td>
</tr>
<tr>
<td>\AtBeginDocument</td>
<td>6/3, 11220, 31080</td>
<td></td>
</tr>
<tr>
<td>\atop</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>\atopwithdelims</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>\attribute</td>
<td>786</td>
<td></td>
</tr>
<tr>
<td>\attributedels</td>
<td>787</td>
<td></td>
</tr>
<tr>
<td>\automaticdiscretionary</td>
<td>788</td>
<td></td>
</tr>
<tr>
<td>\automaticityphenmode</td>
<td>790</td>
<td></td>
</tr>
<tr>
<td>\automaticityphenpenalty</td>
<td>791</td>
<td></td>
</tr>
<tr>
<td>\autoscaling</td>
<td>1108</td>
<td></td>
</tr>
<tr>
<td>\autoexpansion</td>
<td>1109</td>
<td></td>
</tr>
<tr>
<td>\b</td>
<td>29163, 31508</td>
<td></td>
</tr>
<tr>
<td>\badness</td>
<td>202</td>
<td></td>
</tr>
<tr>
<td>\baselineskip</td>
<td>203</td>
<td></td>
</tr>
<tr>
<td>\batchmode</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td>\begin</td>
<td>11306, 11309, 11526, 24600, 29180, 29187, 31365</td>
<td></td>
</tr>
<tr>
<td>\begincsname</td>
<td>793</td>
<td></td>
</tr>
<tr>
<td>\begingroup</td>
<td>3, 20, 24, 29, 48, 104, 182, 205</td>
<td></td>
</tr>
<tr>
<td>\beginL</td>
<td>513</td>
<td></td>
</tr>
<tr>
<td>\beginR</td>
<td>514</td>
<td></td>
</tr>
<tr>
<td>\belowdisplayshortskip</td>
<td>206</td>
<td></td>
</tr>
</tbody>
</table>
Index

\l_box_left_dim ... 32064, 32094, 32149, 32151, 32160, 32164, 32169, 32175, 32180, 32184, 32223, 32238
\l_box_left_new_dim 32068, 32109, 32120, 32152, 32163, 32174, 32185
\l_box_log:nNn 31853
\l_box_resize:NnN 32191
\l_box_resize_common:NnN 32233, 32245, 32277
\l_box_resize_set_corners:NnN 32191
\l_box_resize:NNnN 32191
\l_box_resize:NnN 32191
\l_box_left_new_dim
\l_box_left_dim
\l_box_right_dim ... 32064, 32093, 32147, 32153, 32158, 32162, 32171, 32173, 32182, 32186, 32209, 32222, 32228, 32292, 32309, 32337, 32344
\l_box_right_new_dim ... 32068, 32120, 32154, 32165, 32176, 32187, 32227, 32343, 32397, 32399, 32405
\l_box_rotate:NnN 32073
\l_box_rotate:NNnN 32073
\l_box_rotate_quadrant_four: ... 32073, 32178
\l_box_rotate_quadrant_one: ... 32073, 32145
\l_box_rotate_quadrant_three: ... 32073, 32167
\l_box_rotate_quadrant_two: ... 32073, 32156
\l_box_rotate_xdir:nNn 32073, 32123, 32151, 32153, 32162, 32164, 32173, 32175, 32184, 32186
\l_box_rotate_ydir:nNn 32073, 32134, 32147, 32149, 32158, 32160, 32169, 32171, 32180, 32182
\l_box_scale:NnN 32318, 32348
\l_box_scale:NNnN 32318
\l_box_scale_x_fp ... 32189, 32208, 32228, 32256, 32276, 32291, 32293, 32308, 32328, 32344, 32369, 32371, 32372, 32373, 32483, 32495
\l_box_scale_y_fp ... 32189, 32210, 32230, 32232, 32251, 32256, 32271, 32276, 32293, 32310, 32329, 32340, 32342, 32344, 32370, 32371, 32372, 32373, 32384, 32386
\l_box_set_trim:NnNnN 35350
\l_box_set_viewport:NnNnN 35402, 35405, 35407
\l_box_shown:NnN 31851, 31861, 31865
\l_box_sin_fp
\l_box_top_dim 32064, 32091, 32147, 32151, 32160, 32162, 32171, 32175, 32180, 32186, 32213, 32220, 32232, 32254, 32273, 32313, 32353, 32340
\l_box_top_new_dim ... 32068, 32117, 32148, 32159, 32170, 32181, 32231, 32239, 32388, 32392
\l_box_viewport:NnNnN 35401
\boxdir
\boxdirection
\boxmaxdepth
\breakafterdirmode
\brokenpenalty
\cctab
\cc
code
\cctab_begin:N
\cctab_new:N
\cctab_item:Nn
\cctab_if_exist:NTF
\cctab_gset:Nn
\cctab_end:
\cctab_const:Nn
\cctab_select:N
\cctab_if_exist_p:N
\cctab_start:N

C
\c .. 29163, 31506, 31527, 31553, 31610, 31611, 31630, 31631, 31634, 31635, 31642, 31643, 31654, 31655, 31662, 31663, 31666, 31667, 31722, 31723
\catcode .. 121, 122, 123, 124, 125, 126, 127, 128, 129, 133, 134, 135, 136, 137, 138, 140, 141, 213
\catcodetable .. 799
\cc .. 256
\cctab commands: .. 260, 261, 1175, 1176, 1178–1182, 28622
\cctab_begin:N .. 260, 261, 28574, 28761, 28768, 28812
\cctab_const:Nn .. 260, 261, 28574, 28761, 28768, 28812
\cctab_end: .. 260, 261, 1175, 1176, 1178, 1179, 1181, 1182, 28636
\cctab_gset:Nn .. 260, 261, 28552, 28757
\cctab_if_exist:NTF .. 261, 28709, 28716
\cctab_if_exist_p:N .. 261, 28709
\cctab_item:Nn .. 261, 28693
\cctab_new:N .. 260, 1175, 1176, 28487, 28756, 28760
\cctab_select:N .. 116, 117, 260, 261, 13868, 28557, 28574, 28763, 28770, 28814
\c_code_cctab .. 261, 13868, 28773
\c_document_cctab .. 261, 1178, 28773
\c_intex_cctab .. 261, 28557, 28760
\c_other_cctab .. 261, 28760
Index

1381

cctab internal commands:
\g__cctab_allocate_int ........... 28483, 28615, 28617, 28619
\l_cctab_begin_aux: ............. 1179, 1180, 28603, 28627
\l_cctab_chk_group_begin:n ....... 1181, 28628, 28647
\l_cctab_chk_group_end:n .......... 1181, 28641, 28647
\l_cctab_chk_if_valid:NTF ......... 28554, 28575, 28624, 28713
\g__cctab_endlinechar_prop ......... 1177, 28486, 28533, 28535, 28582
\g__cctab_group_seq ............. 28482, 28649, 28655
\l_cctab_gset:n ............. 28525, 28559, 28631, 28810
\l_cctab_gset_aux:n .......... 28529
\l_cctab_gstore:Nnn ............. 28487
\l_cctab_internal_a_tl .......... 1179, 1180, 28484, 28582, 28583, 28608, 28618, 28626, 28629, 28630, 28631, 28638, 28640, 28642, 28643
\l_cctab_internal_b_tl ........... 28484, 28555, 28569, 28666
\g__cctab_internal_cctab ......... 28564
\l_cctab_item:n ............. 28564, 28585, 28586, 28587, 28588
\l_cctab_nesting_number:N ........... 28629, 28642, 28671
\l_cctab_nesting_number:w ....... 28671
\l_cctab_new:N 1176, 1179, 28487, 28566, 28586, 28607, 28616, 28777
\g__cctab_next_cctab ............ 28603
\l_cctab_select:N .................. 1179, 28574, 28632, 28643
\l_cctab_stack_seq .......... 1175, 28480, 28630, 28638, 28689
\l_cctab_unused_seq .......... 1175, 1180, 28480, 28626, 28640

ceil .................................. 252

char commands:
\l_char_active_seq 83, 185, 18467
\g__cctab_generate:nn ........... 116, 181, 426, 444, 445, 528, 668, 796, 1361, 3739, 3740, 3741, 3742, 3743, 3745, 3746, 3747, 4310, 4326, 4338, 4743, 5766, 6122, 12021, 12037, 13711, 13966, 13982, 18494, 18709, 18713, 18744, 18773, 18828, 18837, 29062, 29100, 29818, 29828, 29829, 29896, 30079, 30080, 30251, 30262, 30309, 30310, 30311, 30322, 30347, 30375, 30402, 30425, 30442, 30454, 30466, 30471, 30496, 30514, 30543, 30545, 30549, 30587, 30588, 30593, 30595, 30603, 30604, 30609, 30611, 30831, 30832, 30837, 30839, 30868, 30869, 30886, 30887, 30888, 30893, 30895, 30897, 30905, 30906, 30907, 30912, 30914, 30916, 31028, 31029, 31030, 31035, 31037, 31392, 31413, 31415, 31474, 31488, 31490
\l_char_active_eq:NN 181, 18473
\l_char_active_eq:NN 181, 18473
\l_char_lower_case:N .......... 36349
\l_char_lowercase:N .......... 36349
\l_char_lowercase:N .......... 36349
\l_char_set_catcode_active_eq:NN 181, 3736, 18473
\l_char_set_catcode_active_eq:NN 181, 4266, 4267, 18473
\l_char_set_catcode:nn 183, 150, 151, 152, 153, 154, 155, 156, 157, 158, 18373, 18380, 18392, 18384, 18386, 18388, 18390, 18392, 18394, 18396, 18398, 18400, 18402, 18404, 18406, 18408, 18410, 18412, 18414, 18416, 18418, 18420, 18422, 18424, 18426, 18428, 18430, 18432, 18434, 18436, 18438, 18440, 18442, 28596
\l_char_set_catcode_active:N 182, 3778, 7168, 9115, 18779, 18474, 18531, 18860, 31374, 35914
\l_char_set_catcode_active:N 183, 18411, 28618, 20146, 20417, 28784, 28791, 28809, 28820, 29030
\l_char_set_catcode_alignment:N .................................. 182, 7220, 18770, 18848
\l_char_set_catcode_alignment:N .................................. 183, 168, 18411, 28566, 28797
\l_char_set_catcode_alignment:N .................................. 183, 18411, 28796
\l_char_set_catcode_end_line:N .................................. 182, 18379
\l_char_set_catcode_end_line:N .................................. 183, 18411, 28792
\l_char_set_catcode_escape:N 182, 18379
Index

\_clist_map_variable:Nmn \ 840, 18111
\_clist_pop:NNN \ ............... 17847
\_clist_pop:w \ 17847
\_clist_pop:wwNNN \ 832, 17847, 17892
\_clist_pop_TF:NNN \ 17873
\_clist_put_left:NNNN \ 17810
\_clist_put_right:NNNN \ 17823
\_clist_rand_item:n \ 18320
\_clist_remove_all: \ 17938
\_clist_remove_all:NNNN \ 17938
\_clist_remove_all:w \ 835, 17938
\_clist_remove_duplicates:NN \ 17916
\_clist_reverse:wwNN .. 836, 17983
\_clist_reverse_end:ww .. 836, 17983
\_clist_sanitize:n \\
............ 17698, 17740, 17805, 17807
\_clist_sanitize:NN \ 828, 17698
\_clist_set_from_seq:n \ 17758
\_clist_set_from_seq:NNNN \ 17758
\_clist_show:NN \ 18333
\_clist_show:NN \ 18355
\_clist_tmp:w ............... 835, 17691, 17951, 17973, 18027, 18036, 18040, 18042, 18176, 18194
\_clist_trim_next:w \ ....... 828, 839, 17692, 17701, 17709, 19077, 18086
\_clist_use:Nw \ 843, 18228
\_clist_use:nn \ 843, 18228
\_clist_use:nnnwnn \ 842, 18195
\_clist_use:wn \ 18195
\_clist_use:end:w \ ....... 843, 18228
\_clist_use_i_delimit_by_s- \ stop:nw \ ....... 17688, 18288
\_clist_use_more:w \ ....... 843, 18228
\_clist_use_none_delimit_by_s- \ mark:w \ ....... 17688, 18243
\_clist_use_none_delimit_by_s- \ stop:w \ 835, 17688, 17706, 17954, 18062, 18069, 18083, 18133, 18141, 18156, 18189, 18230, 18274, 18279
\_clist_use_one:w \ 18289
\_clist_wrap_item:w \ 828, 17707, 17735
\closein \ ................... 217
\closeout \ ................... 218
\clubpenalties \ ................... 516
\clubpenalty \ ................... 219
cm \ ................... 256
code commands:
... code:n \ ................... 224, 20897
coffin commands:
\coffin_attach:NnnNNnnn \ ............... 284, 32677, 33127, 33166, 33623
\coffin_attach:NnnNNNNn \ 283, 33263
\coffin_clear:N \ 281, 32472
\coffin_display_handles:Nn \ 284, 33509
coffin internal commands:
... \_coffin_align:NnnNNnnnN \ ............... 32226, 33277, 33298, 33305, 33401
\_coffin_aligned_coffin \ 32670, 33227,
\delimiterfactor ................. 233
\delimitershortfall .............. 234
deprecation internal commands:
  \_\_deprecation_date_compare:nNnTF
                  ................. 35973, 35990, 35993
  \_\_deprecation_date_compare:_
aux:w ......................... 35973
  \l\_deprecation_grace_period:_
  bool 1365, 35970, 35989, 35999, 36073
  \_\_deprecation_just_error:nNN
                 ..................... 1366, 36011
  \_\_deprecation_minus_six:_
  months:w ...................... 35987
  \_\_deprecation_not_yet_deprecated:nTF
                  ..................... 1366, 35987, 36021
  \_\_deprecation_old:Nnn ...... 36101
  \_\_deprecation_old_protected:Nnn
                  ..................... 36101
  \_\_deprecation_patch_aux:Nn...
                 ..................... 1367, 36011
  \_\_deprecation_patch_aux:nNN
  Nnn ........................................ 36011
  \_\_deprecation_warn_once:nNN
  Nnn ................................ 36011
\detokenize ...................... 43, 131, 522
\DH ............................ 29169, 31064, 31421
\dh .................................. 29169, 31064, 31431
dim commands:
  \dim_abs:n .................. 207, 898, 19828
  \dim_add:N .................. 207, 19810
  \dim_case:nn ................. 210, 19908
  \dim_case:nnn ................ 36165
  \dim_case:nnTF .............. 210, 19908, 19913, 19918, 36166
  \dim_compare:nTF ........... 208–211, 242, 19863, 19932, 19968, 19976, 19985,
                             19991, 20003, 20006, 20017, 32854,
                             32857, 32862, 32876, 32879, 32884,
                             33230, 33235, 33245, 33374, 33386,
                             35364, 35381, 35415, 35429, 35439
  \dim_compare:nTF .......... 208, 209, 211, 19868, 19940, 19948, 19957, 19963
  \dim_compare:p:n ........... 208, 19868
  \dim_compare:p:nNn .......... 208, 19863
  \dim_const:N ................. 206, 891, 901, 19777, 20106, 20107, 21829
  \dim_do_until:nn ............ 211, 19938
  \dim_do_until:nnNn .......... 210, 19966
  \dim_do_while:nn ............ 211, 19938
  \dim_do_while:nnNn .......... 210, 19966
  \dim_eval:n .................. 208, 209, 212, 891, 1252,
                             1355, 19780, 19911, 19916, 19921,
                             19926, 20021, 20049, 20101, 20105,
                             32557, 32634, 32720, 32738, 32775,
                             32779, 32780, 32784, 32788, 32789,
  \dim_gadd:N .................. 207, 19810
  \dim_gset:N .................. 224, 20031
  \dim_gset:NN .................. 207, 891, 19798
  \dim_gset_eq:NN .............. 207, 19804
  \dim_gsub:N .................. 207, 19810
  \dim_gzero:n ................ 206, 19783, 19791, 20127
  \dim_gzero_new:N ................ 206, 19788
  \dim_if_exist:NTF ............ 207, 19789, 19791, 19794
  \dim_if_exist_p:n:NTF ........ 207, 19791
  \dim_log:N ................... 214, 20102
  \dim_log:n ................... 214, 20102
  \dim_max:nn .................. 207, 19828, 33068, 33072
  \dim_min:nn ................... 207, 19828, 33066, 33070, 33083
  \dim_new:N .................. 206, 19771, 19779, 19789,
                             19791, 20108, 20109, 20110, 20111,
                             32064, 32065, 32066, 32067, 32068,
                             32069, 32070, 32071, 32118, 32442,
                             32443, 32446, 32447, 32448, 32449,
                             32942, 32943, 32944, 32945, 32946,
                             33104, 33105, 33446, 33448, 33449
  \dim_ratio:nn ................ 208, 898, 19859, 20065
  \dim_set:N ................... 224, 20931
  \dim_set:NN .................. 207, 19785,
                             32091, 32092, 32093, 32125, 32136,
                             32220, 32221, 32222, 32237, 32335,
                             32336, 32337, 32339, 32431, 32433,
                             32547, 32623, 32856, 32860, 32876,
                             32882, 32913, 32927, 33002, 33037,
                             33045, 33056, 33057, 33058, 33059,
                             33065, 33067, 33069, 33071, 33076,
                             33082, 33165, 33167, 33169, 33177,
                             33179, 33233, 33308, 33309, 33311,
                             33313, 33331, 33332, 33447, 33541,
                             33542, 33588, 33589, 33590, 33592
  \dim_set_eq:NN ............... 207, 19804, 32572, 32573
  \dim_show:N .................. 213, 20068
  \dim_show:n .................. 214, 900, 20100
  \dim_sign:n .................. 212, 20512
  \dim_step_function:nnnN ....... 211, 897, 19994, 20046
  \dim_step_inline:nnn ........ 211, 20024
  \dim_step_variable:nnnN ...... 212, 20024
  \dim_sub:NN .................. 207, 19810
  \dim_to_decimal:n ............ 212, 20071, 20087, 20092
  \dim_to_decimal_in_bp:n ...... 213, 20086
  \dim_to_decimal_in_sp:n ...... 213, 1004, 20088, 23112, 23149, 23743
\dim_to_decimal_in_unit:nn \texttt{213, 20090}
\dim_to_fp:n \texttt{213, 1004, 1023, 20098},
\texttt{27944, 3219, 32130, 32140, 32141,}
\texttt{32209, 32212, 32213, 32238, 32253,}
\texttt{32254, 32273, 32274, 32292, 32309,}
\texttt{32312, 32313, 32867, 32868, 32869,}
\texttt{32889, 32890, 32891, 32901, 32902,}
\texttt{32918, 32919, 32920, 32921, 32931,}
\texttt{32932, 33041, 33042, 33049, 33050,}
\texttt{33123, 33126, 33127, 33178, 33180}
\dim_until_do:nn \texttt{211, 19938}
\dim_until_do:Nnn \texttt{210, 19966}
\dim_use:N \texttt{212, 1252, 19831, 19837,}
\texttt{19838, 19839, 19845, 19846, 19847,}
\texttt{19871, 19890, 20050, 20054, 20069,}
\texttt{20074, 20181, 20182, 20257, 20258,}
\texttt{33004, 33008, 33015, 33021, 33030,}
\texttt{33031, 33032, 33186, 33193, 33339}
\dim_while_do:nn \texttt{211, 19938}
\dim_zero:N \texttt{206, 10783, 19789, 20126,}
\texttt{32094, 32223, 32538, 32847, 32848}
\dim_zero_new:N \texttt{206, 19788}
\c_max_dim \texttt{214, 217, 952,}
\texttt{20106, 20200, 21856, 21898, 21906,}
\texttt{33056, 33057, 33058, 33059, 33076}
\g_tampa_dim \texttt{214, 20108}
\l_tampa_dim \texttt{214, 20108}
\l_tampb_dim \texttt{214, 20108}
\l_tmpb_dim \texttt{214, 20108}
\c_zero_dim \texttt{214, 20003, 20006, 20059,}
\texttt{20106, 20199, 21923, 31950, 31974,}
\texttt{32854, 32857, 32862, 32876, 32879,}
\texttt{32884, 33230, 33235, 33245, 33568,}
\texttt{35379, 35385, 35397, 35415, 35419,}
\texttt{35427, 35429, 35433, 35439, 35449}
dim internal commands:
\__dim_abs:N \texttt{19828}
\__dim_case:nTF \texttt{19908}
\__dim_case:nw \texttt{19908}
\__dim_case_end:nw \texttt{19908}
\__dim_compare:w \texttt{19868}
\__dim_compare:wNN \texttt{893, 19868}
\__dim_compare_once:w \texttt{19868}
\__dim_compare:<w \texttt{19868}
\__dim_compare:@w \texttt{19868}
\__dim_compare:>w \texttt{19868}
\__dim_compare_end:w \texttt{19876, 19900}
\__dim_compare_error: \texttt{893, 19868}
\__dim_eval:w \texttt{899, 19765,}
\texttt{19799, 19801, 19811, 19815, 19820,}
\texttt{19824, 19831, 19837, 19838, 19839,}
\texttt{19845, 19846, 19847, 19862, 19865,}
\texttt{19871, 19890, 19895, 19997, 19998,}
\texttt{19999, 20050, 20054, 20074, 20089}
\__dim_eval_end: \texttt{19765,}
\texttt{19799, 19801, 19811, 19815, 19820,}
\texttt{19824, 19831, 19841, 19849, 19862,}
\texttt{19865, 20050, 20054, 20074, 20089}
\__dim_maxmin:wwN \texttt{19828}
\__dim_ratio:n \texttt{19859}
\__dim_sign:Nw \texttt{20051}
\__dim_step:NnnN \texttt{19994}
\__dim_step:NNnnn \texttt{20024}
\__dim_step:wwN \texttt{19994}
\__dim_to_decimal:w \texttt{20071}
\__dim_use_none_delimit_by_s-_stop:w \texttt{19770, 19886}
\dimen \texttt{235, 19033}
\dimendef \texttt{236}
\dimexpr \texttt{523}
\directlua \texttt{6, 34, 36, 807}
\disablecjktoken \texttt{1168}
\discretionary \texttt{237}
\disinhibitglue \texttt{1113}
\displayindent \texttt{238}
\displaylimits \texttt{239}
\displaystyle \texttt{240}
\displaywidowpenalties \texttt{524}
\displaywidowpenalty \texttt{241}
\divide \texttt{242}
\divide \texttt{243}
\DJ \texttt{29170, 31065, 31422}
\adj \texttt{29170, 31065, 31432}
\do \texttt{1221}
\doublehyphendemerits \texttt{244}
\dp \texttt{245}
\draftmode \texttt{911}
draw commands:
\draw_begin: \texttt{290}
\draw_end: \texttt{290}
\dtou \texttt{1114}
\dump \texttt{246}
\dviextension \texttt{808}
\dvifeedback \texttt{809}
\dvivariable \texttt{810}
\edef \texttt{95, 118, 247}
\efcode \texttt{686}
\c_eight \texttt{36135}
\c_eleven \texttt{36141}
\else 5, 26, 28, 66, 70, 73, 76, 77, 81, 82, 248
else commands:
\else: \texttt{27, 62, 68, 91, 166, 167, 229, 280, 356,
\exp_last_unbraced:Nc 38, 2500, 18996
\exp_last_unbraced:NcV 
\exp_last_unbraced:Nc 38, 2500
\exp_last_unbraced:Ncf 
\exp_last_unbraced:Nco 38, 2500, 18096
\exp_last_unbraced:NcV 38, 2500
\exp_last_unbraced:Nc 38, 2500, 27904
\exp_last_unbraced:Nc 
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500, 27994
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 
\exp_last_unbraced:Nf 38, 2500, 28316
\exp_last_unbraced:Nf 
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 8205
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 8205
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\exp_last_unbraced:Nf 38, 2500
\_\_exp_e_the_toks_register:NTF
\_\_exp_e_the_toks:n
\_\_exp_e_the_toks:N
\_\_exp_e_the_everymath:
\_\_exp_e_the_everydisplay:
\_\_exp_e_the_everycr:
\_\_exp_e_protected:N
\_\_exp_e_put:nn
\_\_exp_last_unbraced:nn
\_\_exp_arg_next:nnn
\_\_exp_eval_register:N
\_\_exp_eval_error_msg:w
ExplSyntaxOn
ExplFileExtension
ExplFileVersion
ExplFileName
ExplFileDate
ExplFileDate
ExplFileName
ExplFileVersion
\_\_explicitdiscretionary
\_\_expanded
\_\_expanded
ExplFileDate
ExplFileExtension
\_\_explicitdiscretionary
\_\_explicitdiscretionary
ExplSyntaxOff
ExplSyntaxOn
\texttt{F}
\texttt{fact}
\texttt{false}
\texttt{fam}
\texttt{fi}
\texttt{fi commands:}
\texttt{fi:}
Index

2253, 2263, 2317, 2320, 2327, 2328,
2601, 2608, 2618, 2631, 2644, 2649,
2652, 2653, 2654, 2655, 2663, 2672,
2688, 2752, 2780, 2786, 2795, 2800,
2806, 2809, 2813, 2821, 2831, 2840,
2844, 2848, 2916, 2932, 2939, 2948,
2962, 2963, 2968, 2969, 2970, 2988,
2989, 2990, 2991, 2992, 2993, 2994,
2995, 2996, 3004, 3023, 3025, 3055,
3056, 3057, 3104, 3206, 3217,
3227, 3347, 3358, 3359, 3403, 3434,
3476, 3487, 3496, 3505, 3560, 3570,
3580, 3692, 3694, 3767, 3771, 3775,
3802, 3813, 3831, 3832, 3833, 3840,
3856, 3862, 3865, 3873, 3883, 3898,
3906, 3914, 3925, 3941, 3961, 3974,
3996, 4014, 4022, 4024, 4027, 4034,
4039, 4123, 4124, 4184, 4215, 4216,
4217, 4224, 4234, 4243, 4265, 4278,
4328, 4329, 4336, 4337, 4342, 4343,
4370, 4371, 4442, 4450, 4493, 4494,
4500, 4504, 4511, 4512, 4517, 4518,
4528, 4529, 4534, 4535, 4549, 4557,
4566, 4567, 4594, 4750, 4786, 4799,
4775, 4787, 4827, 4829, 4836, 4839,
4840, 4844, 4848, 4849, 4850, 4851,
4860, 4861, 4865, 4868, 4869, 4870,
4906, 4909, 4930, 4933, 4941, 4952,
4953, 4964, 4965, 4973, 4985, 4986,
4996, 4997, 5010, 5029, 5030, 5038,
5039, 5090, 5100, 5133, 5147, 5151,
5214, 5262, 5265, 5266, 5281, 5284,
5307, 5365, 5366, 5371, 5398, 5399,
5410, 5414, 5448, 5453, 5461, 5496,
5503, 5508, 5518, 5530, 5556, 5569,
5648, 5687, 5694, 5705, 5793, 5812,
5818, 5823, 5846, 5858, 5859, 5862,
5874, 5908, 5938, 6253, 6292, 6308,
6332, 6352, 6361, 6418, 6425, 6445,
6463, 6474, 6476, 6506, 6509, 6543,
6549, 6555, 6652, 6690, 6721, 6722,
6752, 6778, 6823, 6875, 6920, 6921,
6933, 7011, 7023, 7033, 7043, 7065,
7076, 7102, 7118, 7126, 7176, 7178,
7193, 7195, 7216, 7326, 7347, 7408,
7425, 7426, 7446, 7472, 7474, 7480,
7482, 7487, 7517, 7549, 7560, 7593,
7645, 7647, 7648, 7654, 7834, 7846,
8226, 8269, 8287, 8309, 8329, 8345,
8355, 8371, 8381, 8471, 8473, 8475,
8477, 8481, 8484, 8684, 8692, 10000,
10003, 10006, 10083, 10101, 10356,
10397, 10406, 10427, 10437, 10441,
10448, 10456, 10651, 10655, 10658,
Index

\__file_parse_version:w
\__file_parse_full_name_base:nw
\__file_parse_full_name_area:nw
\__file_name_trim_spaces_aux:w
\__file_name_trim_spaces:nw
\__file_name_str
\__file_name_ext_check:nnnn
\__file_name_ext_check:nnn
\__file_name_expand_error_aux:Nw
\__file_name_expand:Nw
\__file_name_expand_cleanup:w
\__file_name_strip_quotes_-
\__file_internal_ior ............... 10854, 10862, 10864, 10867, 10877, 10878, 10880, 11288, 11299, 11301
\__file_internal_tl ................. 10561, 11134, 11135
\__file_kernel_dependency_-
  compare:nn ....................... 11267
\__file_list:N ..................... 11201
\__file_list_aux:n ................ 11201
\c__file_marker_tl ................
\__file_mdfive_hash:n ............... 10881
\__file_mismatched_dependency_-
  error:nn .......................... 11283, 11286
\__file_name_cleanup:w ............. 10729
\__file_name_end: ................... 10729
\__file_name_expand:n ............... 10605
\__file_name_expand_cleanup:Nw ... 629, 10605
\__file_name_expand_cleanup:w ... 629, 10605
\__file_name_expand:end: 629, 10605
\__file_name_expand_error:Nw ... 629, 10605
\__file_name_expand_error_aux:Nw ... 629, 10605
\__file_name_ext_check:nn ........... 10729
\__file_name_ext_check:nnn ........... 10729
\__file_name_ext_check:nww ........... 10729
\__file_name_quote:nw ............... 10679
\l__file_name_str ................... 10593, 10859, 11126, 11128
\__file_name_strip_quotes:n ........ 10605
\__file_name_strip_quotes:nww ....... 10605
\__file_name_strip_quotes:nwwn ....... 10605
\__file_name_strip_quotes:nnw ........ 10644, 10647, 10653, 10656
\__file_name_strip_quotes:_-
  end:nnnn ......................... 10650, 10655
\__file_name_trimspace:n ........... 10605
\__file_name_trimspace:nu ........... 10605
\__file_name_trimspace:aux:n ......... 10605
\__file_parse_full_name_area:nv .... 641, 11154
\__file_parse_full_name_auxi:nN .. 11151, 11154
\__file_parse_full_name_base:nv ... 642, 11162, 11165
\__file_parse_full_name_base:nu ... 642, 11172, 11173, 11175, 11179
\__file_parse_version:w ............. 11267
\__file_quark_if_nil:nTF ............ 10600, 10669, 10683, 10775, 10784
\__file_quark_if_nil:p:n ............ 10600
\__file_record_seq ................. 640,
\__file_size:n ...................... 10728, 10737, 10757, 10792
\__file_stack_seq ................... 640, 10565, 11119, 11134
\__file_str_cmp:nn ................... 11022, 11052
\__file_timestamp:n .................. 11023
\__file_tmp:w ....................... 10567, 10571, 10575, 10581, 10587
\l__file_tmp_seq ... 10597, 11205,
\l__file_time .. 11208, 11211, 11212, 11224, 11222,
\l__file_times .. 11227, 11298, 11300, 11323, 11327
\filedump ................. 776
\filemoddate ................. 777
\filesize ................. 778
\finalhyphenemerges ............... 272
\firstmark ............... 273
\firstmarks ............... 530
\firstvalidlanguage ........... 819
\c_five ....................... 36129
flag commands:
  \flag_clear:n 168, 169, 5723, 7431,
  7432, 13943, 13971, 14065, 14094,
  14163, 14211, 14212, 14264, 14265,
  14501, 14502, 14503, 14504, 14505,
  14606, 14701, 14702, 14703, 14704,
  14859, 14860, 14861, 17628, 17641
  \flag_clear_new:n 169, 739,
  14446, 14447, 14448, 14449, 14629,
  14630, 14631, 14809, 14810, 17640
  \flag_height:n .......... 169,
  7441, 7443, 13789, 17649, 17665, 17679
  \flag_if_exist:n .......... 169
  \flag_if_exist:nTF . 169, 17641, 17652
  \flag_if_exist:p:n ........ 169, 17652
  \flag_if_raised:n .......... 169
  \flag_if_raised:nTF . 169, 5731, 13782, 13787, 13789,
  14473, 14479, 14484, 14491, 14663,
  14668, 14673, 14824, 14831, 17657
  \flag_if_raised:p:n ....... 169, 17657
  \flag_log:n .......... 169, 17642
  \flag_new:n 168, 169, 739, 824, 5713,
  7309, 7310, 13651, 13652, 17623,
  17641, 22495, 22496, 22497, 22498
  \flag_raise:n .......... 169, 7471,
  7479, 13929, 13979, 14079, 14112,
  14183, 14196, 14234, 14239, 14320,
  14522, 14523, 14546, 14547, 14560,
  14561, 14580, 14581, 14587, 14588,
Index

14618, 14768, 14769, 14878, 14879, 14883, 14884, 14898, 14899, 16767
\flag_raise_if_clear:n ............. 36147
\__flag_clear:wn .................. 17628
\__flag_height:wn ................ 17665
\__flag_height_loop:wn .......... 17665
\__flag_show:N .................... 17642
\floatingpenalty ................ 274
\floor ................................ 252
\fmtname .......................... 107, 8545, 8548,
8549, 8561, 8571, 29441, 29442,
32569, 32570, 33453, 33454, 35240
\font ................................ 275
\fontchardp ........................ 531
\fontcharht ........................ 532
\fontcharic ........................ 533
\fontcharwd ........................ 534
\fontdimen ......................... 944, 276
\fontencoding ...................... 31312
\fontfamily ........................ 31313
\fontid ............................ 820
\fontname .......................... 277
\fontseries ......................... 31314
\fontshape ................................ 31315
\fontsize ................................ 31318
\footnotesize ....................... 31354
\forcecjktoken ...................... 1170
\formatname ........................ 821
\c_four ............................ 36127
\c_fourteen ........................ 36147
\fp .................................
\c_e_fp ............................ 246, 248, 24376
flag_fp_division_by_zero 247, 22495
flag_fp_invalid_operation 247, 22495
flag_fp_overflow ............... 247, 22495
flag_fp_underflow ............. 247, 22495
\fp_abs:n .......................... ...
251, 256, 1158, 27984, 32238,
32340, 32342, 32344, 33168, 33170
\fp_add:N .......................... 239, 1158, 24339
\fp_compare:NnTF ............... 242–244,
24417, 24558, 24564, 24569, 24577,
24638, 24644, 32095, 32097, 32102,
32371, 32386, 32395, 32399, 33142,
34019, 34023, 34031, 34038, 34045,
34052, 34059, 34378, 34381, 34844
\fp_compare:N ................................ 242–244,
250, 24401, 24530, 24536, 24541, 24549
\fp_compare_p:n .................. 243, 24401
\fp_compare_p:nNn ............... 242, 24117
\fp_const:N ....................... 239, 24316, 24376, 24377, 24378, 24379
\fp_do_until:n .................. 244, 24527
\fp_do_until:Nnn ............... 243, 24555
\fp_do_while:n .................. 244, 24527
\fp_do_while:Nnn ............... 243, 24555
\fp_eval:n ...................... 240, 243, 250–256,
1038, 27079, 33725, 33727, 33731,
33735, 33736, 33737, 33741, 33744,
33745, 33746, 33893, 33903, 33907,
33908, 33909, 33914, 33915, 33916,
33917, 33945, 33950, 33958, 33964,
33973, 33974, 33975, 33991, 33992,
33993, 33994, 34001, 34002, 34003, 34010,
34011, 34012, 34074, 34079, 34540,
34541, 34542, 34543, 34554, 34556,
34557, 34569, 34570, 34571, 34572,
34575, 34578, 34579, 34580, 34700,
34701, 34763, 34923, 34924, 34925,
34926, 34934, 34935, 34936, 34937,
34953, 34959, 34976, 34977, 34978,
34986, 34987, 34988, 35709, 35728
\fp_format:n .................... 247
\fp_gadd:N ....................... 239, 24339
\fp_gset:N ....................... 224, 20939
\fp_gset:NN ..................... 239, 24316, 24340, 24342
\fp_gset:eq:NN .................. 239, 24325, 24330
\fp_gsub:N ....................... 240, 24339
\fp_gzero: N ..................... 239, 24339, 24336
\fp_gzero_new:N ............... 239, 24333
\fp_if_exist:NTF ............... 241, 24334, 24336, 24391
\fp_if_exist:NTF ............... 241, 24391
\fp_if_nan:n ..................... 300
\fp_if_nan:NTF .................. 257, 300, 24303
\fp_if_nan_p:n .................. 300, 24303
\fp_log:N ......................... 247, 24349
\fp_log:n ......................... 247, 24372
\fp_max:n ......................... 256, 27086
\fp_min:n ......................... 256, 27086
\fp_new: N ...................... 239, 24313, 24334, 24336,
24380, 24381, 24382, 24383, 32061,
32062, 32063, 32189, 32190, 32439,
32440, 32937, 32938, 33102, 33103
\fp_set:N ......................... 224, 20939
\fp_set: N ....................... 239, 24316, 24339, 24341,
32083, 32084, 32085, 32208, 32210,
32251, 32271, 32291, 32308, 32310,
32328, 32329, 32369, 32370, 2955,
29565, 33122, 33124, 33162, 33163
\fp_set_eq:NN .................. 243, 24329, 214329,
32256, 32276, 32293, 32372, 32373
\fp_show:N ....................... 247, 24349
\fp_show:n ....................... 247, 24372
Index

\_\_fp\_parse\_compare\_auxi:NNNNNNN .................. 24082
\_\_fp\_parse\_compare\_auxi:NNNNN .................. 24082
\_\_fp\_parse\_compare\_end:NNNW .................. 24082
\_\_fp\_parse\_continue:Nw .................. 996, 997, 1024, 23792, 23805,

23981, 24180, 24788, 24798, 24806
\_\_fp\_parse\_continue\_compare:NNw .................. 24173, 24188
\_\_fp\_parse\_digits\_i:N .................. 22974
\_\_fp\_parse\_digits\_ii:N .................. 22974
\_\_fp\_parse\_digits\_iii:N .................. 22974
\_\_fp\_parse\_digits\_v:N .................. 22974
\_\_fp\_parse\_digits\_vi:N .................. 22974
\_\_fp\_parse\_digits\_vii:N .................. 1009, 22974, 23296, 23446
\_\_fp\_parse\_excl\_error: .................. 24082
\_\_fp\_parse\_expand:w .................. 1000, 1001, 22971, 2973,

22983, 23023, 23083, 23127, 23136,

23139, 23143, 23180, 23214, 23252,

23254, 23273, 23275, 23297, 23314,

23327, 23347, 23377, 23405, 23421,

23432, 23455, 23484, 23494, 23501,

23515, 23531, 23551, 23562, 23644,

23667, 23679, 23754, 23763, 23771,

23784, 23902, 23948, 23972, 23998,

24046, 24066, 24135, 24148, 24784
\_\_fp\_parse\_exponent\_N .................. 1014,

23082, 23288, 23437, 23504, 23506
\_\_fp\_parse\_exponent\_Nw .................. 22983, 23023, 23083, 23127, 23136,

23139, 23143, 23180, 23214, 23252,

23254, 23273, 23275, 23297, 23314,

23327, 23347, 23377, 23405, 23421,

23432, 23455, 23484, 23494, 23501,

23515, 23531, 23551, 23562, 23644,

23667, 23679, 23754, 23763, 23771,

23784, 23902, 23948, 23972, 23998,

24046, 24066, 24135, 24148, 24784
\_\_fp\_parse\_exponent\_aux:NNN .................. 23004
\_\_fp\_parse\_exponent\_function:NNNN .................. 23004
\_\_fp\_parse\_exponent\_body:N .................. 23533, 23537
\_\_fp\_parse\_exponent\_digits:N .................. 23541, 23553
\_\_fp\_parse\_exponent\_keep:N .................. 23564
\_\_fp\_parse\_exponent\_keep\_NTF .................. 23544, 23564
\_\_fp\_parse\_exponent\_sign:b .................. 23523, 23527
\_\_fp\_parse\_function:NNN .................. 22608, 22670, 22672,

22675, 23747, 24388, 24390, 26845,

26847, 26849, 26851, 28012, 28014
\_\_fp\_parse\_function\_all_fp_ .................. 22802, 24190, 24686

25441, 25444, 25448, 25760, 25770,

25780, 25790, 25793, 25796, 25803
\_\_fp\_pack\_Bigg:NNNNNNw .................. 22331, 25275,

25278, 25281, 25293, 25296, 25299
\_\_fp\_pack\_eight:NNNNNNNN .................. 970, 1059, 22338,

25088, 25397, 25825, 26074, 26974
\_\_fp\_pack\_twice\_four:wNNNNNNNN .................. 970, 22336, 22925, 22926,

25030, 25031, 25826, 25827, 25828,

25860, 25861, 25862, 26050, 26051,

26386, 26387, 26388, 26976, 26977,

27271, 27272, 27273, 27274, 27967
\_fp\_parse\_n .................. 992, 1004,

1016, 1024, 1037, 1038, 1045, 1158,

1170, 22956, 23107, 23765, 24317,

24319, 24321, 24344, 24395, 24404,

24421, 24431, 24588, 24648, 25573,

27738, 27792, 27870, 27915, 27930,

27983, 27985, 27987, 27989, 28366
\_fp\_parse\_after:w .................. 23765
\_fp\_parse\_apply\_binary:NNwNN .............. 23903, 23834, 23847
\_fp\_parse\_apply\_binary_chk:N .................. 1009, 23953
\_fp\_parse\_apply\_binary\_error:NNN .................. 23803
\_fp\_parse\_apply\_comma:NNwNN .................. 24141, 24150
\_fp\_parse\_apply\_compare:NNwNw .................. 24162, 24165, 24170
\_fp\_parse\_apply\_function:NNNN .................. 1019, 23597, 23758
\_fp\_parse\_apply\_unary:NNNN .................. 23002, 23634, 23749
\_fp\_parse\_apply\_unary_chk:NwNNN .................. 23613, 23614, 23617
\_fp\_parse\_apply\_unary_chk:NwNN .............. 23602
\_fp\_parse\_apply\_unary_chk:NwNw .................. 23602
\_fp\_parse\_apply\_unary\_error:NNN .................. 23602
\_fp\_parse\_apply\_unary\_type:NNN .................. 23602
\_fp\_parse\_caseless\_inf:N .................. 23715
\_fp\_parse\_caseless\_infinity:N .................. 23715
\_fp\_parse\_compare:NNNNNNN .................. 24082
\__fp__parse_function_one_two:nnw
... 1142, 24202, 27439, 27445, 28081
\__fp__parse_function_one_two_aux:nnw ........... 24202
\__fp__parse_function_one_two_auxi:nn ........... 24202
\__fp__parse_function_one_two_error_o:w .......... 24202
\__fp__parse_infix:NN ......................... 1002, 1006, 1022,
1027, 23022, 23192, 23231, 23707,
23722, 23744, 23860, 23863, 23946
\__fp__parse_infix!:N .............. 24082
\__fp__parse_infix&:Nw .............. 24039
\__fp__parse_infix(:N .............. 24022
\__fp__parse_infix:.=:N .............. 23936
\__fp__parse_infix:*:N .............. 24024
\__fp__parse_infix:+:N .............. 24082
\__fp__parse_infix_=:N .............. 24024
\__fp__parse_infix_+:N .............. 24082
\__fp__parse_infix_&:Nw .............. 24066
\__fp__parse_infix:NN .............. 24082
\__fp__parse_function_one_two_minus:NN ........... 1024, 1028, 23722, 23772, 23785, 23934
\__fp__parse_function_one_two_minus:NN ........... 1027, 23871, 23897, 23988
\__fp__parse_pay_and:N ........................................ 23988
\__fp__parse_pay_check:NN .............. 23881, 23891, 23923
\__fp__parse_pay_comma:w .............. 23881, 23891, 23923
\__fp__parse_pay_end:N .............. 1024, 1028, 23772, 23777, 23785, 23934
\__fp__parse_pay_juxt:N .............. 1027, 23871, 23897, 23988
\__fp__parse_pay_mark:NN .............. 23866, 23910, 23933
\__fp__parse_pay_mul:N .............. 1027, 1030, 23896,
23913, 23921, 23988, 24023, 24032
\__fp__parse_pay_or:N ........................................ 23988
\__fp__parse_pay_!:N ........................................ 23988
\__fp__parse_pay_large:i:Nw .............. 24039
\__fp__parse_pay_large=N .............. 1008, 23259, 23342
\__fp__parse_pay_large:wwN .............. 1012, 23344, 23349
\__fp__parse_pay_large_round:NN .............. 1012, 23385, 23457
Index

12716, 19243, 19254, 19260, 20303,
20444, 29217, 29593, 31137, 35936
\group_begin: 13, 667, 1395, 1446, 2179, 2182, 2185,
2594, 3065, 3256, 3419, 3456, 3735,
3779, 3768, 3859, 4019, 4203, 4585,
4677, 5001, 5498, 5832, 6055, 6156,
6532, 6907, 7162, 7351, 7360, 7372,
7382, 7391, 7566, 7597, 7718, 8478,
8523, 8683, 8779, 9086, 9110, 9126,
9192, 10048, 10258, 10304, 10566,
10707, 11237, 11849, 11986, 12028,
12221, 12973, 13000, 13274, 13297,
13704, 13814, 13867, 14162, 14210,
14256, 14263, 14592, 14774, 16329,
16366, 18473, 18479, 18530, 18596,
18841, 18859, 18883, 18971, 18990,
19355, 20281, 20420, 20464, 21470,
24740, 28556, 28778, 28872, 28918,
29002, 29029, 30530, 30537, 30566,
30825, 30862, 31022, 31051, 31082,
31373, 31460, 31609, 33678, 33811,
34214, 34902, 34944, 35913, 36291
\c_group_begin token 165,
193, 308, 44, 686, 860, 3824, 4323,
12579, 12617, 18841, 18865, 29085,
29290, 31912, 31918, 31932, 31938,
32016, 32022, 32037, 32043, 35951
\group_end: 13, 14, 442, 544, 787, 1178,
1181, 1305, 1446, 2179, 2182, 2188,
2603, 3068, 3259, 3423, 3464, 3673,
3749, 3784, 3805, 3866, 4043, 4194,
4597, 4691, 5034, 5042, 5511, 5836,
6115, 6163, 6170, 6178, 6536, 6537,
6948, 7226, 7356, 7367, 7455, 7571,
7637, 7719, 7720, 8482, 8542, 8700,
8807, 9015, 9118, 9137, 9346, 10054,
10262, 10333, 10589, 10725, 11240,
11522, 12008, 12058, 12211, 12225,
12991, 13023, 13279, 13302, 13714,
13827, 13870, 14175, 14222, 14281,
14435, 14773, 14905, 16338, 16376,
16381, 18481, 18488, 18599, 18615,
18858, 18862, 18890, 18989, 19038,
19379, 20419, 20445, 20504, 21484,
21764, 28500, 28811, 28927, 28996,
29016, 29045, 30555, 30565, 30848,
30876, 31046, 31050, 31077, 31108,
31377, 31730, 31875, 33679, 33816,
34217, 34307, 34359, 35917, 36299
\c_group_end token 308, 860, 3827, 18841, 18870,
29086, 29296, 31926, 32031, 35952
\group_insert_after:N 14, 1452, 4593, 34462
\group_log_list: 14, 2191
\group_show_list: 14, 2191

groups commands:
\groups:n 224, 20947

H
\H 29163, 31503, 31650, 31651, 31678, 31679
\halign 282
\hangafter 283
\hangindent 284
\hbadness 285
\hbox 286
\hbox commands:
\hbox:n 269, 273, 31883, 32110, 32406, 33462
\hbox_gaet:Nw 273, 31885, 32077, 32200, 32244, 32264, 32284, 32301, 32322,
32531, 32526, 32520, 32951, 35348
\hbox_gaet:Nw 274, 31909, 32993
\hbox_gaet_to_wd:Nn 274, 31909, 32996
\hbox_gaet_to_wd:Nw 274, 31929
\hbox_overlap_center:n 274, 31953
\hbox_overlap_left:n 274, 31953
\hbox_overlap_right:n 274, 31953
\hbox_set:Nn 269, 273, 274, 287,
31885, 32074, 32106, 32107, 32194,
32241, 32261, 32281, 32298, 32319,
32348, 32356, 32379, 32507, 32948,
32971, 33228, 33315, 33594, 35345,
35358, 35366, 35374, 35383, 35392,
35409, 35417, 35425, 35431, 35444
\hbox_set:Nw 274, 31909, 32580
\hbox_set_end: 274, 31909, 32583
\hbox_set_to_wd:Nn 274, 31987
\hbox_set_to_wd:Nw 274, 31929
\hbox_to_wd:nn 273, 31943, 32397
\hbox_to_zero:n 273, 31943, 31954, 31956, 31958, 35338
\hbox_unpack:N 274, 31950, 33232
\hbox_unpack_drop:N 277, 31959
\hcoffin commands:
\hcoffin_gset:Nn 282, 32503
\hcoffin_gset:Nw 282, 32576
\hcoffin_gset_end: 282, 32576
\hcoffin_set:N 282, 32503, 33466, 33473, 33511, 33546
\hcoffin_set:Nw 282, 32576
\hcoffin_set_end: 282, 32576
\hfi 1118
\hfi 287
Index

16847, 16848, 16862, 16864, 16865, 16882, 16885, 16886, 16887, 16894, 16925, 16959, 16961, 16963, 16983, 16985, 17008, 17042, 17060, 17098, 17106, 17117, 17172, 17173, 17199, 17382, 17409, 17415, 17442

\_\_int_eval_end: \ldots 16799, 16812, 16817, 16831, 16866, 16882, 16888, 16897, 16925, 16959, 16961, 16963, 16965, 16983, 16985, 17060, 17098, 17106, 17382, 17409, 17415, 17442

\_\_int_from_alph:N \ldots 819, 17499
\_\_int_from_alph:nN \ldots 819, 17499
\_\_int_from_base:N \ldots 819, 17516
\_\_int_from_base:nN \ldots 819, 17516
\_\_int_from_roman:NN \ldots 17558
\_\_int_from_roman:C_int \ldots 17544
\_\_int_from_roman:c_int \ldots 17544
\_\_int_from_roman:D_int \ldots 17544
\_\_int_from_roman:d_int \ldots 17544
\_\_int_from_roman:ERROR:w \ldots 17558
\_\_int_from_roman:L_int \ldots 17544
\_\_int_from_roman:l_int \ldots 17544
\_\_int_from_roman:M_int \ldots 17544
\_\_int_from_roman:m_int \ldots 17544
\_\_int_from_roman:V_int \ldots 17544
\_\_int_from_roman:v_int \ldots 17544
\_\_int_from_roman:X_int \ldots 17544
\_\_int_from_roman:x_int \ldots 17544
\_\_int_if_recursion_tail:stop:w \ldots 16806, 17040
\_\_int_if_recursion_tail_stop:N \ldots 16809, 17571
\_\_int_if_recursion_tail_stop_do:N \ldots 16809, 17510, 17527, 17574
\_\_int_internal_a_int \ldots 17618
\_\_int_internal_b_int \ldots 17618
\_\_int_maxconstdef_int \ldots 16905
\_\_int_maximin:WN \ldots 16828
\_\_int_mod:Ww \ldots 16860
\_\_int_pass_signs:wn \ldots 818, 17489, 17503, 17520
\_\_int_pass_signs_end:wn \ldots 17489
\_\_int_show:N \ldots 17594
\_\_int_show:Nw \ldots 16814
\_\_int_step:NNNnN \ldots 17205
\_\_int_step:NNnN \ldots 17168
\_\_int_step:WwwN \ldots 17168
\_\_int_to_Base:nn \ldots 17323
\_\_int_to_base:nn \ldots 17323
\_\_int_to_base:nN \ldots 17323
\_\_int_to_base:nnN \ldots 17323
\_\_int_to_base:nnnN \ldots 17323
\_\_int_to_base:nnnnN \ldots 17323
\_\_int_to_base:nnnnnN \ldots 17323
\_\_int_to_base:nnnnnnN \ldots 17323
\_\_int_to_base:nnnnnnnN \ldots 17323
\_\_int_to_Letter:n \ldots 17323
\_\_int_to_letter:n \ldots 17323
\_\_int_to_roman:N \ldots 17453
\_\_int_to_roman:w \ldots 807, 818, 1450, 16799, 17018, 17456, 17466
\_\_int_to_Roman_aux:N \ldots 17465, 17468, 17471
\_\_int_to_Roman_c:w \ldots 17453
\_\_int_to_Roman_c:w \ldots 17453
\_\_int_to_Roman_d:w \ldots 17453
\_\_int_to_Roman_d:w \ldots 17453
\_\_int_to_Roman_i:w \ldots 17453
\_\_int_to_Roman_i:w \ldots 17453
\_\_int_to_Roman_l:w \ldots 17453
\_\_int_to_Roman_l:w \ldots 17453
\_\_int_to_Roman_m:w \ldots 17453
\_\_int_to_Roman_m:w \ldots 17453
\_\_int_to_Roman_q:w \ldots 17453
\_\_int_to_Roman_q:w \ldots 17453
\_\_int_to_Roman_v:w \ldots 17453
\_\_int_to_Roman_v:w \ldots 17453
\_\_int_to_Roman_x:w \ldots 17453
\_\_int_to_Roman_x:w \ldots 17453
\_\_int_use_none_delimit_by_s:stop:w \ldots 16806, 17040

intarray commands:
\\_\_intarray_const_from_clist:Nn \ldots 236, 21759, 21945, 26410, 27011
\\_\_intarray_count:Nn \ldots 235, 236, 344, 14040, 14049, 14055, 21695, 21875, 28504
\\_\_intarray_gset:NNnN \ldots 236, 21759, 21945, 26410, 27011
\\_\_intarray_gset_rand:Nn \ldots 235, 344, 952, 954, 14024, 14036, 14049, 14055, 21695, 21875, 28504
\\_\_intarray_gset_rand:N \ldots 301, 22028
\\_\_intarray_gset_rand:Nn \ldots 301, 22028
\\_\_intarray_gzero:N \ldots 236, 21708, 21917
\\_\_intarray_item:N \ldots 236, 344, 948, 952, 954, 14034, 14039, 14073, 14102, 14110, 21719, 21757, 21927, 21943, 28704
\\_\_intarray_log:N \ldots 236, 22013
\\_\_intarray_new:Nn \ldots 235, 945, 951, 954, 6589, 6590, 7314, 7315, 7316, 7317, 14022, 14031, 21604, 21831, 28329, 28330, 28331, 28502
\\_\_intarray_rand_item:N \ldots 236, 21756, 21942
\\_\_intarray_show:N \ldots 236, 949, 954, 22013
\\_\_intarray_to_clist:Nn \ldots 301, 21760, 21960
Index

1444

internal commands:
\__intarray: w ..................  21598, 21609
\l__intarray_bad_index_int .........  21595, 21703, 21751
\__intarray_bounds: NNNf .............  21785, 21887, 21938
\__intarray bounded: NNN .............  21857
\__intarray const_from_clist:nN . .... 21945
\__intarray count: w ..................  21827, 21843, 21853, 21951, 21972
\__intarray entry: w .................  21827, 21876, 21923, 21928
\g__intarray_font_int .................  21830, 21834, 21836
\__intarray_get: N ..................  21850
\__intarray_get: Nw ..........  21879, 21885
\__intarray_get: w ...........  21675, 21697
\__intarray_get: wT ...........  21675, 21700
\__intarray_get_all_same:Nn ....  22028
\__intarray_get_count:w ..........  21675, 21879, 21885
\__intarray_get_overflow:Nnn ..  21875
\__intarray_get_overflow: NNN ..........  21899, 21907, 21911
\__intarray_get_overflow_test: w .......  21896, 21904, 21957, 22047, 22054
\__intarray_get_rand: NNN ...........  22028
\__intarray_get_rand auxi: Nnnn ........  22028
\__intarray_get_rand auxiii: Nnnnn ......  22028
\__intarray_get_range: Nw ........  21795
\__intarray_get_range: Nw ........  21998
\__intarray_get_range: w ........  21798
\__intarray_item: N ........  21927
\__intarray_item: Nw ..........  21931, 21936
\__intarray_item: w ........  21719
\g__intarray_table_int .................  21633
\__intarray_table .................  21829, 21836, 21876
\__intarray signed_max_dim: n .........  21855, 21914, 21915
\c__intarray_sp_d .................  21829, 21836, 21876
\__intarray_sp_d .............  21829, 21836, 21876
\__intarray_table .................  21829, 21836, 21876
\g__intarray_table_int .................  21633
\__intarray_to_clist: Nn .............  21769, 21960, 22024
\__intarray_to_clist: w ..........  21769, 21960
\interactionmode ........................  544
\interlinepenalties ..............  545
\interlinepenalty ..................  325
\ior commands:
\ior_close:N .................  84, 301, 9932, 9951, 10854, 10867,
11301, 28928, 28961, 28998, 29015
\ior_get: NN .............  85, 86, 301, 10008, 10099
\ior_get: NNTF ..........  85, 10008, 10099
\ior_get: str: NN .............  36195
\ior_get: term: n .............  301, 10842
\ior_if_eof:N .................  84, 9902, 10014,
10034, 10674, 10864, 10878
\ior_if_eof: p: N ............  84, 9902
\ior_list_streams: .............  36197
\ior_log: N .................  84, 9963
\ior_log: list: ............  84, 9979, 36200
\ior_log: streams: .............  36199
\ior_map: break: .......  84, 10043, 10062
\ior_map: break:n ..........  87, 10057, 10075,
10082, 10094, 10100, 28923, 28994
\ior_map: break:n ..........  87, 10057
\ior_map: inline: Nn .............  86, 10091, 11299
\ior_map: variable: Nn .............  84, 10057, 28920
\ior_new: N .................  84, 9901, 9903, 9904, 10880, 28869
\ior_open: N .................  84, 609, 9905, 28884, 28920, 28962, 29014
\ior_open: NNTF ............  84, 9906, 9909
\ior_shell_open: N .............  301, 11288, 35757
\ior_show: N .................  84, 9963
\ior_show: list: ..........  84, 9979, 36198
\ior_str: get: NN .................  36196
\ior_str: get: NNTF .............  86, 10021, 10022
\ior_str: get: term: N .............  301, 10042
\ior_str: map: inline: Nn .............  86, 87, 10061, 28955, 28985
\ior_str: map: variable: Nn .............  87, 10087
\g__tmpa_ior ..................  91, 9903
\g__tmpb_ior ..................  91, 9903
\ior internal commands:
\l__ior_file_name_tl .............  9908, 9911, 9913
\l__ior_get: N .............  10008, 10043, 10062
\_ior\_get\_term:NNN 10042
\_ior\_internal\_tl ............. 10042
\_ior\_list:N .................. 9979
\_ior\_map\_inline:NNN 10061
\_ior\_map\_inline:NNNN ........ 10061
\_ior\_map\_variable:NNNN .... 10087
\_ior\_map\_variable\_loop:NN .... 10061
\_ior\_new:N .................. 607, 9919, 9936
\_ior\_new\_aux:N ............... 9923, 9927
\_ior\_open\_stream:N .......... 9930
\_ior\_shell\_open:n ........... 35757
\_ior\_show:NN ................ 9963
\_ior\_str\_get:NN 10021, 10045, 10064
\_ior\_stream\_tl ............... 8856, 9933, 9937, 9944
\g\_ior\_streams\_prop .................. 608, 9887, 9945, 9956, 9971, 9986
\g\_ior\_streams\_seq .............. 8855, 9933, 9957, 9958
\c\_ior\_term\_ior .................. 9884, 9901, 9953, 9959, 9995, 10052
\c\_ior\_term\_noprompt\_ior .... 10041, 10051

iow commands:
\iow\_allow\_break: .................. 100, 301, 10273, 10315, 10320
\iow\_allow\_break:n ............... 618
\iow\_char:N ..................... 78, 89, 3685, 3688, 3689, 3713, 3714, 3721, 3722, 4710, 4711, 4718, 4720, 4722, 4724, 4726, 4728, 5357, 5358, 6074, 6078, 6082, 6083, 6207, 7858, 7861, 7862, 7867, 7901, 7910, 7914, 7919, 7939, 7941, 7942, 7944, 7947, 7949, 7954, 7956, 7958, 7963, 7967, 7970, 7971, 7974, 7976, 7980, 7982, 7988, 7990, 7994, 7996, 8000, 8005, 8007, 8049, 8051, 8056, 8058, 8064, 8069, 8074, 8075, 8088, 8091, 8095, 8096, 8100, 8108, 8172, 9572, 9575, 9576, 9608, 9636, 9720, 9724, 9725, 9731, 9734, 10044, 10142, 10145, 10147, 10049
\iow\_close:N .................... 84, 10147, 10165
\iow\_indent:n .................. 90, 619, 620, 9535, 9670, 9732, 9740, 9748, 10284, 10316, 10321, 14140, 14465, 14653, 22631, 22643, 35097, 35126, 35148, 35164, 35173, 35182, 35197
\iow\_line\_count\_int .................. 90, 449, 620, 4163, 4167, 9087, 10245, 10325, 10330, 10368
\iow\_list\_streams: ................ 36201
\iow\_log:N ..................... 84, 10177
\iow\_log:n ................... 88, 1890, 9296, 9303, 10220, 12091, 36222, 36302
\iow\_log\_list: ................. 85, 10193, 36204
\iow\_log\_streams: ................ 36203
\iow\_new:N ................... 84, 10136, 10138, 10139
\iow\_newline: ................... 78, 88–90, 345, 586, 616, 695, 11366, 14142, 26516, 28830, 28832, 4113, 6129, 9109, 10243, 10313, 10322, 10328, 11217, 33622, 33623, 33624, 35468, 35470, 35473, 35480, 9886, 88, 89, 8732, 10223, 10229, 10230, 10231, 10232, 88, 89, 616, 8764, 10208, 88, 89, 616, 10205, 84, 10177, 36202, 88, 301, 1890, 10220, 9121, 9286, 9291, 9309, 9335, 10229, 36224, 88, 90, 301, 620, 695, 1549, 9085, 9088, 9100, 9267, 9301, 9307, 9314, 10276, 10282, 10287, 10299, 10302, 12916, 12931, 91, 618, 10107, 10167, 10229, 10230, 91, 618, 10107, 10136, 10167, 10173, 10231, 10232, 91, 618, 10107, 91, 10138, 91, 10138
\iow\_internal\_commands:
\_\_iow\_allow\_break: .................. 618, 10273, 10315
\_\_iow\_allow\_break\_error: .............. 618, 10273, 10320
\_\_iow\_file\_name\_tl .............. 10142, 10145, 10149, 10153
\_\_iow\_indent:n ................ 619, 10284, 10316
\_\_iow\_indent\_error:n .............. 619, 10284, 10321
\_\_iow\_indent\_int .............. 10252, 10306, 10349, 10496, 10513, 10521
\_\_iow\_indent\_tl .............. 10252, 10306, 10383, 10495, 10514, 10522, 10523
\_\_iow\_internal\_tl .............. 10100, 10185, 10188
\_\_iow\_line\_break\_bool .............. 10256, 10362, 10490, 10504, 10512, 10520, 10528, 10530, 10535, 10537
\_\_iow\_line\_part\_tl .............. 622–624, 10254, 10364,
Index

`:__kernel_backend_postscript:n` 348
`:__kernel_backend_svg:n` 348
`:__kernel_backend_matrix:n` 349
`:__kernel_postscript:n` 349
`:__kernel_backend_scope_begin:` 349
`:__kernel_backend_scope_end:` 349
`:__kernel_chk_cs_exist:N` 342
`:__kernel_chk_defined:NTF`...

`:__kernel_chk_tl_type:NnnTF` 342
`:__kernel_backend_postscript:n`...

`:__kernel_int_arrange:n` 342
`:__kernel_int_free_cs:N`...

`:__kernel_dependency_version:-
check:Nn` 343, 11267
`:__kernel_dependency_version:-
check:n` 343, 11267
`:__kernel_deprecation_code:nnn`...

`:__kernel_deprecation_error:NN`...

`:__kernel_deprecation_undo_:Nn"... 

`:__kernel_file_missing:n`.......

`:__kernel_known_name:nnn` 343, 9006, 11093, 11102
`:__kernel_known_name_quote:n`...

`:__kernel_known_name_sanitize:n`...

`:__kernel_known_if_debug:TF` ....... 

`:__kernel_int_add:nnn` 344, 16892, 28235

`:__kernel_int_arrange:NNn` 344, 947, 949, 952, 9629, 6736, 6739, 7342, 7397, 7399, 7405, 7413, 7415, 7418, 7539, 7543, 7555, 7558, 21695, 21766, 21838, 21850, 21875, 21958, 22008, 22083, 23999, 28400, 28402, 28406, 28407, 28408, 28510, 28511, 28545, 28548

`:__kernel_int_arrange_range:-
clist:Nn` 344, 6799, 21795, 21998

`:__kernel_int_arrange_item:Nn`...

`:__kernel_group_allow_blank:-
keys_bool` 143, 147, 19483, 19493, 19495, 20280, 20453

`:__kernel_group_err:error:n` 9485

`:__kernel_group_err:nnn` 9485

`:__kernel_group_err:nnnn` 9497

`:__kernel_group_err:NNNN` 9497

`:__kernel_group_info:nnn` 9485
Index

\_keys_find_key_module:wNN
\_keys_find_key_module_auxii:Nw
\_keys_find_key_module_auxi:Nw
\_keys_find_key_module_auxiii:Nw
\_keys_find_key_module_auxiv:Nw
\_keys_groups_clist
\_keys_groups_root_str
\_keys_groups_set:n
\_keys_inherit:n
\_keys_inherit_root_str
\_keys_inherit_str
\_keys_initialise:n
\_keys_legacy_if_inverse:nn
\_keys_legacy_if_inverse:nnnn
\_keys_legacy_if_set:nn
\_keys_legacy_if_set:nnnn
\_keys_legacy_if_set_inverse:nn
\_keys_meta_make:n
\_keys_meta_make:nn
\_keys_module_str
\_keys_multichoice_find:n
\_keys_multichoice_make:
\_keys_multichoice_make:nn
\_keys_multichoice_find:nn
\_keys_multichoices_make:nn
\_keys_no_value_bool
\_keys_only_known_bool
\_keys_parent:n
\_keys_parent_auxi:w
\_keys_parent_auxii:w
\_keys_parent_auxiii:n
\_keys_parent_auxiv:w
\_keys_prop_put:Nn
\_keys_property_find:n
\_keys_property_find:auxi:w
\_keys_property_find:auxii:w
\_keys_property_find:auxiii:w
\_keys_property_find:auxiv:w
\_keys_property_find:err:w
\_keys_property_str
\_keys_props_root_str
\_keys_quark_if_no_value:NTF
\_keys_quark_if_no_value_p:N
\_keys_relative_tl
\_keys_selective_bool
\_keys_quark_if_no_value:NTF
\_keys_quark_if_no_value_p:N
\_keys_relative_tl
\_keys_selective_bool

21114, 21149, 21155, 21156, 21168, 21175, 21176, 21257, 21278, 21283
\_keys_find_key_module:wNN
\_keys_find_key_module_auxii:Nw
\_keys_find_key_module_auxi:Nw
\_keys_find_key_module_auxiii:Nw
\_keys_find_key_module_auxiv:Nw
\_keys_groups_clist
\_keys_groups_root_str
\_keys_groups_set:n
\_keys_inherit:n
\_keys_inherit_root_str
\_keys_inherit_str
\_keys_initialise:n
\_keys_legacy_if_inverse:nn
\_keys_legacy_if_inverse:nnnn
\_keys_legacy_if_set:nn
\_keys_legacy_if_set:nnnn
\_keys_legacy_if_set_inverse:nn
\_keys_meta_make:n
\_keys_meta_make:nn
\_keys_module_str
\_keys_multichoice_find:n
\_keys_multichoice_make:
\_keys_multichoices_make:nn
\_keys_no_value_bool
\_keys_only_known_bool
\_keys_parent:n
\_keys_parent_auxi:w
\_keys_parent_auxii:w
\_keys_parent_auxiii:n
\_keys_parent_auxiv:w
\_keys_prop_put:Nn
\_keys_property_find:n
\_keys_property_find:auxi:w
\_keys_property_find:auxii:w
\_keys_property_find:auxiii:w
\_keys_property_find:auxiv:w
\_keys_property_find:err:w
\_keys_property_str
\_keys_props_root_str
\_keys_quark_if_no_value:NTF
\_keys_quark_if_no_value_p:N
\_keys_relative_tl
\_keys_selective_bool

\_keys_find_key_module_auxii:Nw
\_keys_find_key_module_auxi:Nw
\_keys_find_key_module_auxiii:Nw
\_keys_find_key_module_auxiv:Nw
\_keys_groups_clist
\_keys_groups_root_str
\_keys_groups_set:n
\_keys_inherit:n
\_keys_inherit_root_str
\_keys_inherit_str
\_keys_initialise:n
\_keys_legacy_if_inverse:nn
\_keys_legacy_if_inverse:nnnn
\_keys_legacy_if_set:nn
\_keys_legacy_if_set:nnnn
\_keys_legacy_if_set_inverse:nn
\_keys_meta_make:n
\_keys_meta_make:nn
\_keys_module_str
\_keys_multichoice_find:n
\_keys_multichoice_make:
\_keys_multichoices_make:nn
\_keys_no_value_bool
\_keys_only_known_bool
\_keys_parent:n
\_keys_parent_auxi:w
\_keys_parent_auxii:w
\_keys_parent_auxiii:n
\_keys_parent_auxiv:w
\_keys_prop_put:Nn
\_keys_property_find:n
\_keys_property_find:auxi:w
\_keys_property_find:auxii:w
\_keys_property_find:auxiii:w
\_keys_property_find:auxiv:w
\_keys_property_find:err:w
\_keys_property_str
\_keys_props_root_str
\_keys_quark_if_no_value:NTF
\_keys_quark_if_no_value_p:N
\_keys_relative_tl
\_keys_selective_bool

20658, 20661, 20665, 20749, 21301, 21310, 21331, 21339, 21450
\_keys_parent_auxi:w
\_keys_parent_auxii:n
\_keys_parent_auxiv:w
\_keys_prop_put:Nn
\_keys_property_find:n
\_keys_property_find:auxi:w
\_keys_property_find:auxii:w
\_keys_property_find:auxiii:w
\_keys_property_find:auxiv:w
\_keys_property_find:err:w
\_keys_property_str
\_keys_props_root_str
\_keys_quark_if_no_value:NTF
\_keys_quark_if_no_value_p:N
\_keys_relative_tl
\_keys_selective_bool

20794, 20998, 21000, 21002, 21004
\_keys_property_find:n
\_keys_property_find:auxi:w
\_keys_property_find:auxii:w
\_keys_property_find:auxiii:w
\_keys_property_find:auxiv:w
\_keys_property_find:err:w
\_keys_property_str
\_keys_props_root_str
\_keys_quark_if_no_value:NTF
\_keys_quark_if_no_value_p:N
\_keys_relative_tl
\_keys_selective_bool

20512, 20562, 20568, 20574, 20575, 20601, 20609, 20617, 20618, 20621, 20624
\_keys_prop_puts_root_str
\_keys_props_root_str
\_keys_quark_if_no_value:NTF
\_keys_quark_if_no_value_p:N
\_keys_relative_tl
\_keys_selective_bool

20757, 20585, 20606, 20607
\_keys_property_err:w
Index

\l__keys_selective_seq ............
  . . 20529, 21185, 21188, 21190, 21265
\l__keys_set:n:n ............... 21051
\l__keys_set_filter:nnn ........... 21121
\l__keys_set_filter:nnnn ........... 21121
\l__keys_set_keyval:n .............. 21077, 21192
\l__keys_set_keyval:nn ............. 21121, 21192, 21199
\l__keys_set_keyval:nnn ............. 21121
\l__keys_set_keyval:nnnn .......... 21121
\l__keys_set_known:nn ............... 21080
\l__keys_set_known:nnnn ............ 21080
\l__keys_set_selective .......... 21192
\l__keys_set_selective:nn ........... 21121
\l__keys_set_selective:nnn .......... 21121
\l__keys_show:Nn .................. 21121
\l__keys_store_unused:........... 21259, 21279, 21285, 21322
\l__keys_store_unused:w ............ 21406, 21426, 21432
\l__keys_store_unused_aux: .......... 21322
\l__keys_tmp:w ..................... 21471, 21483
\l__keys_tmp:bool .................
  . . 20532, 21264, 21271, 21276
\l__keys_tmp:tl ............... 20532, 20788
\l__keys_tmp:tl .................. 20532, 20788, 20793
\l__keys_trim_spaces:n ............
  . . 918, 20544, 20574, 20603, 20694, 21076, 21208, 21403, 21444, 21445, 21470, 21506, 21514, 21525, 21534
\l__keys_trim_spaces:auxi:w .......... 21470
\l__keys_trim_spaces:auxiii:w ......... 21470
\l__keys_type_root_str .............
  . . 20506, 20658, 20661, 20673
\l__keys.Undefine: .............. 20743, 20798, 21046
\l__keys_unused_clist .............
  . . 931, 20530, 21083, 21089, 21094, 21096, 21097, 21124, 21131, 21136, 21138, 21139, 21380, 21390, 21418
\l__keys_validate_forbidden: ........ 20808
\l__keys_validate_required: .......... 20808
\l__keys_validate_root_str ...........
  . . 20506, 20815, 20821, 20825, 21326, 21345
\l__keys_value_or_default:n ........
  . . 21215, 21288
\l__keys_value:require:nn ...........
  . . 20725, 20808, 20868, 21048, 21050
\l__keys_variable_set:Nnn ...........
  . . 20854, 20900, 20902, 20904, 20906, 21014, 21016, 21018, 21020, 21022, 21024, 21026, 21028, 21030, 21032, 21034, 21036, 21038, 21040, 21042, 21044
\l__keys_variable_set:required:Nnn ...........
  . . 20854, 20932, 20934, 20936, 20938, 20940, 20942, 20944, 20946, 20960, 20962, 20964, 20966, 20990, 20992, 20994, 20996, 21006, 21008, 21010, 21012

keyval commands:
\keyval_parse:Nn ............... ........ 234, 914, 20284, 20545, 21077
\keyval_parse:nn ............... ........ 233, 234, 908, 913, 14949, 20284

keyval internal commands:
\keyval_blank_key_error:w ........
  . . 20425, 20434, 20449
\keyval_blank_true:w ........ 20381, 20449
\keyval_clean_up_active:w ........
  . . 911, 20323, 20336, 20357, 20389, 20409
\keyval_clean_up_other:w ........
  . . 912, 20362, 20367, 20378
\keyval_end_loop_active:w ........
  . . 20310, 20403
\keyval_end_loop_other:w ........
  . . 912, 20320, 20403
\keyval_if_blank:w ..............
  . . 20381, 20425, 20434, 20446
\keyval_if_empty:w ............ 20416
\keyval_if_recursion_tail:w .......
  . . 20399, 20319, 20446
\keyval_key:nn ..................
  . . 912, 20383, 20420, 20449
\keyval_loop_active:nnnw ........
  . . 20291, 20301, 20307, 20410
\keyval_loop_other:nn........... 909, 20311, 20317, 20393, 20401, 20413, 20430, 20439, 20450, 20451, 20456
\keyval_misplaced_eqaul_after:-
  active_error:w ..............
  . . 910, 20330, 20334, 20385
\keyval_misplaced_equal_in:-
  split_error:w ........ 20341, 20346, 20350, 20354, 20370, 20375, 20385
\keyval_pair:nnnn .............
  . . 911, 20356, 20377, 20420
\keyval_split_active:w ........
  . . 909, 910, 20313, 20321, 20340, 20405
\keyval_split_active:auxi:w ........
  . . 20322, 20327, 20358, 20408
\keyval_split_active:auxii:w ....
  . . 910, 20327, 20387
\keyval_split_active:auxiii:w ....
  . . 910, 20327
\keyval_split_active:auxiv:w ....
  . . 910, 20327
\keyval_split_active:other:w ......
  . . 20313, 20329, 20349, 20360, 20369
Index 1453

\_keyval_split_other auxi:w
\_keyval_split_other auxi:w 20364
\_keyval_split_other auxiii:w
\_keyval_tmp:w 913, 20418, 20421, 20443, 20444, 20465, 20503
\_keyval_trim:nN 20337, 20356, 20365, 20377, 20383, 20449, 20464
\_keyval_trim auxi:w 20464
\_keyval_trim auxii:w 20464
\_keyval_trim auxiii:w 20464
\kuten 1139, 1173

L
\L 29172, 31067, 31424
\l 29172, 31067, 31430
l3kernel 97, 11451
l3kernel.charcat 97, 11532
l3kernel.elapsedtime 97, 11535
l3kernel.filedump 97, 11549
l3kernel.filemd5sum 98, 11576
l3kernel.filemoddate 98, 11591
l3kernel.filesize 98, 11650
l3kernel.resettimer 98, 11535
l3kernel.shellescape 98, 11675
l3kernel.strcmp 98, 11666
\labelediag 29180, 29187, 31372
\language 328
\LONG 31352
\Long 31353
\Large 31356
\ln 26621, 26624
\lastallocatedtoks 3382
\lastbox 329
\lastkern 330
\lastlinefit 546
\lastnamedcs 832
\lastnodechar 1140
\lastnodesubtype 1141
\lastnodevalue 547
\logb 252
\longpenalty 331
\longtext 183, 344, 19027, 19031
\lastsavedboxresourceindex 917
\lastsavedimageresourceindex 919
\lastsavedimageresourcepages 921
\lastskip 332
\lastxpos 923
\lastypos 924
\latex 97, 11431
\latexlua 833
\latexluafunction 834
\lccode 333
\leaders 334
\left 335
\left ghost 835
\leftghost 97, 11410, 11412, 36206
\lefthyphenmin 336
\leftmargin kern 688
\leftskip 337
\legacy commands:
\legacy_if:nTF 99, 11761
\legacy_if gset:n 20067
\legacy_if gset:nn 99, 11777
\legacy_if gset_false:n
\legacy_if p:n 99, 11764
\legacy_if set:n 99, 11777
\legacy_if set:nn 99, 11777
\legacy_if set_false:n
\leqno 338
\let 22, 180, 181, 380
\letcharcode 862
\letterspacefont 689
\limits 340
\LineBreak
\linedir 837
\linedirection 838
\linepenalty 341
\lineskip 342
\lineskiplimit 343
\linewidth 32572
\ln 26621, 26624
\logb 252
\longpenalty 331
\longtext 183, 344, 19027, 19031
\lastsavedboxresourceindex 917
\lastsavedimageresourceindex 919
\lastsavedimageresourcepages 921
\lastskip 332
\lastxpos 923
\lastypos 924
\latex 97, 11431
\latexlua 833
\latexluafunction 834
\lccode 333
\leaders 334
\left 335
\left ghost 835
\latexescape:n 97, 11410, 11412, 36206
Index

\luatexifengine: T 36213, 36215, 36217
\luatexifengine:p: 36211
\luatexalignmark 1243
\luatexaligntab 1244
\luatextattribute 1245
\luatextattributedef 1246
\luatexbanner 850
\luatexbodydir 1282
\luatexboxdir 1283
\luatexcatcodetable 1247
\luatexclearmarks 1248
\luatexcrampeddisplaystyle 1249
\luatexcrampedscriptscriptstyle 1251
\luatexcrampedscriptstyle 1252
\luatexcrampedtextstyle 1253
\luatexfontid 1254
\luatexformatname 1255
\luatexleaders 1256
\luatexinitcatcodetable 1257
\luatexitellu 1258
\luatexitleftghost 1284
\luatexitlocalbrokenpenalty 1285
\luatexitlocalinterlinepenalty 1287
\luatexitlocalleftbox 1288
\luatexitlocalrightbox 1289
\luatexluascapestring 1290
\luatexmathdir 1290
\luatexmathstyle 1261
\luatexmathfunction 1260
\luatexmathdir 1290
\luatexmathstyle 1261
\luatexmathfunction 1260
\luatexmathmode 1286
\luatexmathmode 1286
Index

or commands:
\or: .............................. 166, 707,
709, 973, 1417, 1982, 1983, 1984,
2658, 2659, 2660, 2661, 2662, 3837,
3838, 4031, 4032, 4545, 4546, 4547,
4548, 4819, 4820, 4821, 4822, 4823,
6306, 6359, 6997, 6998, 6999, 7000,
7001, 7002, 7003, 7004, 7005, 7006,
7007, 7008, 7009, 7345, 7346, 10420,
10421, 10422, 10423, 10424, 10425,
10426, 13372, 13448, 13692, 13693,
13694, 13695, 13696, 14721, 14722,
16790, 17384, 17385, 17386, 17387,
17388, 17389, 17390, 17391, 17392,
17393, 17394, 17395, 17396, 17397,
17398, 17399, 17400, 17401, 17402,
17403, 17404, 17405, 17406, 17407,
17408, 17417, 17418, 17419, 17420,
17421, 17422, 17423, 17424, 17425,
17426, 17427, 17428, 17429, 17430,
17431, 17432, 17433, 17434, 17435,
17436, 17437, 17438, 17439, 17440,
17441, 18558, 18562, 18565, 18567,
18568, 18570, 18572, 18574, 18575,
18577, 18579, 18581, 18583, 18611,
22173, 22174, 22175, 22424, 22439,
22440, 22811, 22812, 22837, 24120,
24121, 24122, 24158, 24843, 24844,
24845, 24968, 25053, 25139, 25140,
25141, 25142, 25143, 25144, 25145,
25146, 25147, 25226, 25229, 25565,
25566, 25580, 25581, 25595, 25597,
26102, 26127, 26133, 26134, 26135,
26136, 26137, 26286, 26321, 26323,
26331, 26524, 26574, 26577, 26586,
26701, 26724, 26725, 26757, 26758,
26762, 26815, 26816, 26856, 26861,
26871, 26876, 26886, 26891, 26901,
26906, 26916, 26921, 26931, 26936,
27463, 27464, 27509, 27594, 27597,
27609, 27615, 27662, 27664, 27665,
27675, 27681, 27758, 27759, 27766,
27812, 27813, 27820, 27886, 27887,
28107, 28390, 28391, 28392, 28469,
28470, 28471, 35951, 35952, 35953
\over .................................. 395
\overfullrule .......................... 395
\overline ................................ 395
\overwithdelims ........................ 395

P
\PackageError .................................. 89, 97
\pagebottomoffset ................................ 874
\pagedepth .................................. 399
\pagedir .................................. 875
\pagedirection ................................ 876
\pagediscard ................................ 553
\pagefillstretch ................................ 400
\pagefillstretch ................................ 401
\pagefillstretch ................................ 402
\pagefillstretch ................................ 1144
\pagegoal .................................. 403
\pageheight ................................ 927
\pageloftoffset ................................ 877
\pagerightoffset ................................ 878
\pageshrink .................................. 404
\pagestretch .................................. 405
\pagetopoffset ................................ 879
\pagetotal .................................. 406
\pagewidth .................................. 928
\par ................................ 14, 19, 85, 369, 1259, 407, 31964,
31966, 31970, 31975, 31980, 31985,
31992, 31997, 32004, 32009, 32029
\pardir .................................. 880
\pardirection ................................ 881
\parfillskip ................................ 408
\parindent .................................. 409
\parshape .................................. 410
\parshapedimen ................................ 554
\parshapindent ................................ 555
\parshapelength ................................ 556
\parskip .................................. 411
\partokencontext ................................ 1184
\partokenname ................................ 1185
\patterns .................................. 412
\pausing .................................. 413
\pc .................................... 256
\pdf commands:
\pdf_destination:nn ...... 296, 3534
\pdf_destination:nnnn .. 296, 297, 3536
\pdf_object_if_exist:nTF . 294, 4527
\pdf_object_if_exist_p:n . 294, 4527
\pdf_object_new:nn ... 294, 4527
\pdf_object_if_exist:nTF . 294, 4527
\pdf_object_ref:n .. 295, 4527
\pdf_object_ref_last: 295, 4527
\pdf_object_unnamed_write:nn ...
\pdf_object_write:nn ... 294, 4527
\pdf_pageobject_ref:n .. 295, 4527
\pdf_pageobject_ref:nn 295, 4527
pdf internal commands:
\_\_pdf_backend_compress_objects:n 35254
\_\_pdf_backend_compresslevel:n 35253
\_\_pdf_backend_destination:nn 35335
\_\_pdf_backend_destination:n mun
\_\_pdf_backend_object_last: 35339
\_\_pdf_backend_object_new:nn 35268
\_\_pdf_backend_pageobject_ref:n
\_\_pdf_backend_version_major:.n
\_\_pdf_backend_version_minor:-
gset:n 35326
\_\_pdf_backend_version_minor:
... 35287, 35298, 35310, 35331, 35333
\_\_pdf_backend_version_minor-
gset:n 35327
\_\_pdf_backend_version_major-
gset:n 35260
\_\_pdf_backend_version_major-
gset:n 35285
\_\_pdf_backend_version_major-
gset:n 35280
\_\_pdf_backend_version_major-
gset:n 35267
\_\_pdf_backend_version_major-
gset:n 35266
"pdf_adjustspacing" 651
"pdfannot" 579
"pdfcatalog" 580
"pdfcolorstack" 582
"pdfcolorstackinit" 583
"pdfcompresslevel" 581
"pdfcopyfont" 652
"pdfcreationdate" 584
"pdfdecimaldigits" 585
"pdfdest" 586
"pdfdestmargin" 587
"pdfdraftmode" 653
"pdfachinedepth" 654
"pdfachlineheight" 655
"pdfelapsedtime" 656
"pdfendlink" 588
"pdfendthread" 589
"pdfextension" 882
"pdf_feedback" 883
"pdffiledump" 713
"pdffilemoddate" 712
"pdffilesize" 657
"pdffontstr" 590
"pdffontexpand" 658
"pdffontname" 591
"pdffontobjnum" 592
"pdffontsize" 659
"pdfgamma" 593
"pdfgencounter" 596
"pdfglyphtounicode" 597
"pdfhori" 598
"pdfignoreddimen" 600
"pdfimagemappic" 594
"pdfimagegamma" 595
"pdfimagecolor" 599
"pdfimagegamma" 600
"pdfimagegamma" 601
"pdfincludechars" 602
"pdfincludecopyfonts" 603
"pdfincludeerrorlevel" 604
"pdfinfo" 605
"pdfinsertline" 606
"pdflastannot" 607
"pdflastlink" 608
"pdflastobj" 609
"pdflastxi" 610
"pdflastximagecol" 611
"pdflastximagecol" 612
"pdflastximag" 613
"pdflastximag" 614
"pdflinkmargin" 615
"pdfmajorversion" 616
"pdfmapfile" 665
"pdfmapfile" 666
"pdfmapline" 666
"pdfmd5sum" 711
"pdfminorversion" 617
"pdfnames" 618
"pdfnoligatures" 667
"pdfnormaldeviate" 668
"pdfobj" 619
"pdfobjcompresslevel" 620
"pdfoutline" 621
"pdfoutput" 622
"pdfpageattr" 623
"pdfpagebox" 625
"pdfpageheight" 669
\_\_peek\_collect\_remove: mw ... 35920
\_\_peek\_collect\_tl ... 1362,
35919, 35931, 35933, 35958, 35963
\_\_peek\_collect\_true: w ... 1362, 35920
\_\_peek\_execute\_branches: catcode: ... 874, 19285, 35921
\_\_peek\_execute\_branches: catcode\_aux: ... 19285
\_\_peek\_execute\_branches: catcode\_auxii: N ... 19285
\_\_peek\_execute\_branches: charcode: ... 874, 19285, 35923
\_\_peek\_execute\_branches: meaning: ... 874, 19277, 35925
\_\_peek\_execute\_branches\_N\_type: ... 19355
\_\_peek\_false: w ... 875, 1362,
19215, 19223, 19244, 19258, 19282,
19305, 19315, 19372, 19385, 35932
\_\_peek\_false\_aux: N ... 1362, 35933, 35934
\_\_peek\_N\_type: w ... 19355
\_\_peek\_N\_type\_aux: mw ... 19355
\_\_peek\_remove\_spaces: ... 19231
\l\_\_peek\_search\_tl ... 870, 873, 1362,
19214, 19251, 19302, 19312, 35930
\l\_\_peek\_search\_token ... 870, 1362, 19213, 19250, 19279, 35929
\_\_peek\_tmp: w ... 19215, 19220, 19356, 19378
\_\_peek\_token\_generic: NNTF ... 874, 875, 19265,
19267, 19269, 19389, 19393, 19395
\_\_peek\_token\_generic\_aux: NNTF ... 19247, 19266, 19272
\_\_peek\_token\_remove\_generic: NNTF ... 874, 19265,
19273, 19275
\_\_peek\_true: w ... 875, 1362,
19215, 19257, 19280, 19303,
19313, 19370, 19384, 19385, 35939
\_\_peek\_true\_aux: w ... 871,
872, 19215, 19228, 19235, 19236,
19252, 19266, 35940, 35941, 35959
\_\_peek\_true\_remove: w ... 871, 872, 19226,
19241, 19272, 35964
\_\_peek\_use\_none\_delimit\_by\_s: stop: w ... 875, 19221, 19368
\penalty ... 414
\pi ... 23122, 23123
pi ... 255
\pm ... 24593, 24594
\postbreakpenalty ... 1145
\postdisplaypenalty ... 415
\postexhyphenchar ... 885
\posthyphenchar ... 886
\prebinoppenalty ... 887
\prebreakpenalty ... 1146
\predisplaydirection ... 557
\predisplaypenalty ... 416
\predisplaysize ... 417
\preexhyphenchar ... 885
\prehyphenchar ... 890
\preinhyphenchar ... 891
\prelinepenalty ... 418
\prewidth ... 419
\prevgraf ... 420
prg commands:
\prg\_break: ... 69, 518, 741,
791, 792, 2256, 3598, 3673, 4026,
4107, 4137, 4138, 4139, 4140, 4141,
4142, 4710, 4714, 5958, 5968, 5973,
5982, 6006, 6034, 6846, 6874, 7553,
8486, 10878, 12783, 13754, 13770,
13888, 13920, 14026, 14029, 14170,
14217, 14270, 14276, 14509, 14590,
14762, 14903, 16507, 16540, 16591,
16644, 17184, 17702, 18184, 22235,
22244, 24244, 24264, 24265, 24489,
24490, 24503, 24603, 24604, 24605,
27996, 28048, 28272, 35554, 35560
\prg\_break\_n: ... 69, 413, 420, 634,
1352, 2256, 3426, 3467, 3591, 3598,
3949, 4108, 4144, 4688, 5954, 6003,
6383, 6716, 6868, 7547, 7553, 8486,
10851, 12773, 13657, 13665, 13755,
13771, 13889, 13921, 14027, 14030,
14171, 14218, 14271, 14277, 14510,
14710, 14867, 16378, 16508, 16542,
16593, 16644, 16689, 16696, 17189,
17702, 18184, 21967, 21986, 22001,
22236, 22425, 24245, 24266, 24491,
24607, 27997, 28048, 28280, 35555
\prg\_break\_point: N ... 69, 69, 139, 377, 441, 451,
792, 812, 897, 2247, 4078, 4197,
6607, 6621, 6667, 7732, 8486, 10075,
10094, 12329, 12358, 12369, 13197,
13223, 13243, 13265, 16543, 16584,
16594, 16617, 16624, 16633, 17236,
18057, 18079, 18101, 18128, 18150,
19680, 19702, 19718, 20047, 24676
Index

\_prg_replicate_first_5:n ... 8422
\_prg_replicate_first_6:n ... 8422
\_prg_replicate_first_7:n ... 8422
\_prg_replicate_first_8:n ... 8422
\_prg_replicate_first_9:n ... 8422
\_prg_set_eq_conditional:NNN 1716
\_prg_set_eq_conditional:nnNnNw
                               1724, 1732
\_prg_set_eq_conditional:F_form:nn ... 1732
\_prg_set_eq_conditional:F_form:wNn ... 1769
\_prg_set_eq_conditional:_loop:nnnnNw ... 1732
\_prg_set_eq_conditional:p_form:nn ... 1732
\_prg_set_eq_conditional:p_form:wNn ... 1732
\_prg_set_eq_conditional:T_form:nn ... 1732
\_prg_set_eq_conditional:T_form:wNn ... 1767
\_prg_set_eq_conditional_TF_form:nn ... 1732
\_prg_set_eq_conditional_TF_form:wNn ... 1767
\_prg_T_true:w ... 1669
\_prg_use_none_delimit_by_q_recursion_stop:w
                      359, 1669
\primitive .................. 781
prop commands:
\c_empty_prop ............ 205, 879, 19404, 19414, 19418, 19421, 19641
\prop_clear:N ......... 198, 19417, 19424, 19456, 19470, 33239, 34104
\prop_clear_new:N ........ 198, 19423, 34200, 34231, 34272
\prop_concat:N ....... 200, 879, 19442
\prop_const_from_keyval:Nn
                               199, 19454, 32420, 32427, 34995
\prop_count:N .......... 201, 19565, 35536
\prop_gclear:N ......... 198, 19417, 19427, 19462
\prop_gclear_new:N .......... 198, 1275, 19423, 32404, 32495
\prop_gconcat:N ........ 200, 19442
\prop_get:N .............. 69, 36233, 36235
\prop_get:Nn .......... 136, 137, 200, 201, 19524, 33479, 33438, 33552, 33556, 33706, 33785, 33873, 34791
\prop_get:NnNF ........ 200–202, 9382, 9402, 9457, 9971, 10185, 13861, 19659, 28582, 32685, 33779, 33878, 34210, 34479, 34492, 34507, 34597, 34626, 34642, 35008, 35021
\prop_gprop:Nn ............ 200, 19532
\prop_gprop:NnNF ....... 200, 202, 19675
\prop_gput:N .............. 226, 20909
\prop_gput:Nn
                    ........ 199, 3725, 3726, 8136, 8137, 9172, 9174, 9865, 9866, 9873, 9874, 9876, 9877, 9878, 9879, 9899, 9945, 10130, 10160, 13624, 13625, 13626, 13627, 13628, 13629, 13630, 13631, 13632, 13633, 13634, 13635, 13636, 13637, 13638, 13645, 13648, 14497, 14498, 19486, 19597, 20462, 20463, 21584, 21585, 22490, 22491, 28535, 28864, 28865, 32711, 32729, 32764, 32795, 34393, 34394, 34395, 34396, 34397, 34398, 34399, 34400, 34411, 34413, 34415, 34416, 34417, 34418, 34419, 34420, 34421, 34532, 34533, 34547, 34561, 34582, 35225, 35226
\prop_gput_from_keyval:Nn 200, 19454
\prop_gput_if_new:Nn .... 199, 19618
\prop_gremove:N
                          .... 201, 19556, 10170, 19508, 28533
\prop_gaet_eq:NN ........ 199, 19421, 19429, 19446, 32496, 32498, 32603, 32665, 32702, 32704, 32951, 33117, 33158
\prop_gaet_from_keyval:Nn 199, 19454
\prop_hput:Nn .......... 19597
\prop_if_empty:NTF ...... 201, 19639, 35533
\prop_if_empty:p:N ...... 201, 19639
\prop_if_exist:NTF ...... 201, 19424, 19427, 19633, 20786, 33697
\prop_if_exist:p:N ...... 201, 19635
\prop_if_in:NNTF ...... 201, 9177, 9183, 19646, 34212
\prop_if_in:p:N ...... 201, 19646
\prop_item:Nn ......... 201, 203, 9178, 9184, 19554, 34354, 34394, 36234, 36236
\prop_log:N .......... 204, 19737
\prop_map_break:
                      .... 203, 886, 888, 19676, 19677, 19678, 19679, 19680, 19702, 19714, 19715, 19716, 19717, 19718, 19733
\prop_map_break:n .... 204, 19662, 19655, 19733
\prop_map_function:NNN .... 202, 203, 302, 888, 9986, 10200, 19570, 19670, 19752, 33627, 35080
\prop_map_inline:NNn .... 203, 19451, 19656, 32961, 32963, 32966, 32984, 32986,
Index

1466

\__prop_internal_tl .............. 884, 19400, 19403,
19601, 19607, 19608, 19624, 19631
\__prop_item:nn .......... 883, 19554
\__prop_keyval_parse:NNn ....
... 19476, 19477, 19484, 19485, 19491
\__prop_map_function:NNw .... 886, 19670
\__prop_map_tokens:nn .... 19710
\__prop_missing_eq:n .... 19454
\__prop_pair:wn ...... 877, 881, 887,
888, 19400, 19401, 19502, 19505,
19603, 19626, 19676, 19677, 19678,
19679, 19683, 19684, 19685, 19686,
19698, 19700, 19705, 19714, 19715,
19716, 19717, 19721, 19722, 19723,
19724, 19747, 19756, 19759, 35541
\__prop_put:NNn .......... 19597
\__prop_put_if_new:NNnn ...... 19618
\__prop_rand_item:w .... 35531
\__prop_show:NN .... 19737
\__prop_show_validate:w .... 19737
\__prop_split:NTF ............ 878, 884–886, 19497,
19510, 19516, 19526, 19534, 19543,
19577, 19587, 19606, 19629, 19661
\__prop_split_aux:NTF .... 19497
\__prop_split_aux:w .... 881, 19497
\__prop_use_i_delimit_by_s_-
stop:nn .... 35530, 35544
\protect ............ 1209, 10318,
23057, 29437, 29460, 29462, 31278
\protected 118, 120, 144, 558, 19029, 19031
\protrudechars ........... 929
\ProvidesExplClass ........ 9
\ProvidesExplFile ........ 9, 36292, 36309
\ProvidesExplFileAux .... 36295, 36297
\ProvidesExplPackage .... 9
\ProvidesFile ........ 36300, 36301
pt .................. 256
\ptexminversion ........ 1147
\ptexversion ........ 1148
\ptexversion ........ 1149
\pdimen ................ 930

Q
quark commands:
\q_mark .......... 137, 426, 15790
\q_nil ............ 25,
117, 137, 354, 769, 771, 773, 1536,
1539, 15796, 15850, 15869, 15875,
15890, 15891, 15897, 15921, 15925,
21452, 21453, 21455, 21457, 21459,
21461, 21467, 21473, 21473, 34743, 34744

\prop_map_tokens:Nn ...... 202,
203, 793, 883, 886, 19556, 19648, 19710
\prop_new:N ...... 198, 9171, 9173,
9195, 9354, 9887, 10118, 13623,
19411, 19424, 19427, 19437, 19438,
19439, 19440, 19441, 20786, 28486,
32939, 32940, 32941, 33409, 33450,
34297, 34387, 34392, 34409, 34414
\prop_pop:NNn .......... 200, 19532
\prop_pop:NNnNTF ...... 200, 202, 19575
\prop_put:N .... 226, 20997
\prop_put:NNn ...........
...... 199, 200, 399, 877, 878,
9430, 9446, 9463, 19451, 19478,
19507, 32708, 32726, 32745, 32762,
32793, 32995, 32997, 33003, 33005,
33014, 33202, 33208, 33087, 33095,
33185, 33191, 33199, 33206, 33350,
33410, 33412, 33414, 33416, 33418,
33420, 33422, 33424, 33426, 33428,
33430, 33432, 33434, 33436, 33438,
33440, 33442, 33444, 33575, 34105,
34201, 34219, 34262, 34277, 34459
\prop_put_from_keyval:NNn . 200, 19454
\prop_put_if_new:NNnn ...... 199, 19618
\prop_rand_key_value:N .... 303, 35531
\prop_remove:Nn ........ 201,
9427, 9442, 19508, 33521, 33524, 33528
\prop_set_eq:NN ............
...... 199, 19418, 19429, 19443,
19450, 32651, 32653, 32695, 32697,
32948, 32957, 32959, 33110, 33134,
33136, 33155, 33283, 33516, 34282
\prop_set_from_keyval:NNn ..
...... 199, 200, 879, 19451, 34437
\prop_show:N ........ 204, 19437
\_g_tapn_prop ........ 204, 19437
\_l_tapn_prop ........ 204, 19437
\_l_tapb_prop ........ 204, 19437
prop internal commands:
\_prop_concat:NNNN ........ 19442
\_prop_count:nn ........ 19565
\_prop_from_keyval_key:w .... 880
\_prop_from_keyval_value:w .... 880
\_prop_if_in:nnn ........ 19646
\_prop_if_reursion_tail_stop:n .... 19409
\_l_prop_internal_prop ....
.............. 19441, 19450,
19451, 19452, 19468, 19469, 19470
\ProvidesSupportedClass
........ 144, 558, 19029, 19031
\Providesfile ........ 256
\ptexversion ........ 1147
\ptexversion ........ 1148
\ptexversion ........ 1149
\pdimen ................ 930
Index

\regex_replace_once:nnN 55, 56, 195, 7290
\regex_replace_once:NaNNTF 56, 7290
\regex_replace_once:nnNTF 56, 548, 7290
\regex_set:Nn 46, 54, 55, 7238
\regex_show:N 54, 491, 7253
\regex_show:n 47, 52, 54, 7253
\regex_split:NnN 56, 7290
\regex_split:NNNTF 56, 7290
\regex_split:nNTF 56, 7290
\regex_split:nnNTF 56, 7290
\g_tapa_regex 37, 7234
\l_tapa_regex 37, 7234
\g_tapb_regex 37, 7234
\l_tapb_regex 37, 7234

regex internal commands:
\__regex_A_test: 463, 5335, 5337, 5973, 5976, 5982, 6082, 6512

\__regex_action_cost:n 499, 501, 6301, 6302, 6310, 6760, 6786
\__regex_action_free:n 499, 510, 6524, 6303, 6331, 6342, 6400, 6404, 6429, 6454, 6458, 6461, 6489, 6497, 6507, 6521, 6564, 6758, 6762
\__regex_action_free_aux:nnn 6762
\__regex_action_free_group:n 499, 510, 6350, 6469, 6472, 6762

\__regex_action_startWildcard:N 498, 6236, 6255
\__regex_action_submatch:nN 499, 6423, 6424, 6562, 6811, 6813
\__regex_action_submatch:auxw 6813
\__regex_action_submatch_auxi:w 6813
\__regex_action_submatch_auxii:w 6813
\__regex_action_submatch_auxiii:w 6813
\__regex_action_success: 498, 6239, 6252, 6834

\__regex_actionWildcard: 515
\l_regex_added_begin_int 7318, 7440, 7448, 7452, 7493, 7626, 7631, 7634, 7645, 7660
\l_regex_added_end_int 7318, 7442, 7448, 7453, 7494, 7628, 7631, 7635, 7647, 7661
\c_regex_allCatcodes_int 4883, 5005, 5095, 5933
\c_regex_ascii_lower_int 4406, 4527, 4533
\c_regex_ascii_maxControl_int 4463, 4644
\c_regex_ascii_max_int 4463, 4637, 4645, 4835
\c_regex_ascii_min_int 4463, 4636, 4643
\__regex_assertion:Nn 463, 477, 508, 5331, 5353, 5962, 6075, 6512
\__regex_b_test: 463, 508, 5343, 5345, 5979, 6080, 6512
\l_regex_balance_int 455, 521, 542, 4462
6909, 6942, 7175, 7192, 7330, 7343, 7345, 7346, 7576, 7603, 7627, 7629
\g_regex_balance_intarray 449, 534, 6890, 6897, 7317, 7342
\l_regex_balance_tl 521, 524, 6852, 6910, 6943, 7040

\__regex_begin 7309
\__regex_branch:n 463, 481, 504, 4459, 5010, 5503, 5556, 5741, 5851, 5859, 5943, 5945, 5948, 6057, 6395
\__regex_break_point:TF 453, 476, 501, 4747, 4480, 6301, 6302, 6318, 6353
\__regex_break_true:w 453, 4747, 4480, 4485, 4492, 4499, 5306, 4516, 4566, 4577, 4593, 5306, 6542, 6548, 6554
\__regex_build:N 532, 6219, 7283, 7289, 7290, 7297
\__regex_build:n 532, 6219, 7278, 7287, 7292, 7295
\__regex_build_aux:NN 543, 6219, 7682, 7701, 7771
\__regex_build_aux:Nn 543, 6219, 7673, 7691, 7763
\__regex_build_for_cs:n 4588, 6241
\__regex_build_new_state: 6233, 6244, 6245, 6274, 6283, 6315, 6349, 6435, 6397, 6412, 6417, 6456, 6475, 6510, 6514, 6559
\l_regex_build_t1 481, 548, 4456, 5002, 5009, 5027, 5032, 5035, 5036, 5039, 5040, 5043, 5089, 5092, 5132, 5146, 5150, 5275, 5289, 5330, 5352, 5365, 5397, 5410, 5414, 5496, 5499, 5502, 5508, 5509, 5512, 5555, 5845, 5849, 5856, 5862, 5883, 5899, 5917, 6056, 6113, 6116, 6127, 6157, 6172, 6176, 6179, 6185, 6908, 6931, 6946, 6949, 6968, 7037, 7097, 7100, 7114, 7182, 7830, 7833, 7841, 7844
\__regex_build_transition_left:NNN 6270, 6458, 6472, 6489
\__regex_build_transition_right:NN

\regex_split:nnN 4466, 4527, 4533
\regex_split:NNNTF 4463, 4644
Index

6284, 6287, 6288, 6290, 6291, 6317, 6325, 6328, 6351, 6399, 6401, 6411, 6431, 6451, 6453, 6481, 6484, 6487, 6490, 6502, 6515, 6524, 6560, 6567
\_\_regex_left_state_seq

\_\_regex_maplike_break

\_\_regex_match_once_init:

\_\_regex_match_init:
\_\_regex_match_cs:n

\_\_regex_match_once_init:

\_\_regex_match_once_init_aux:

\_\_regex_match_once_active:n

\_\_regex_match_token:n:n

\_\_regex_match_successbool

\_\_regex_matched_analysis_tl

\_\_regex_max_pos_int

\_\_regex_max_thread_int

\_\_regex_min_pos_int

\_\_regex_min_state_int

\_\_regex_min_submatch_int

\_\_regex_min_thread_int

\_\_regex_mode_int

\_\_regex_mode_quit_c:

\_\_regex_msg_repeated:n:n

\_\_regex_multi_match:n:n

\c\_\_regex_no_match_regex

\_\_regex_peek:nnTF

\_\_regex_peek_aux:nnTF

\_\_regex_peek_end:

\_\_regex_peek_false_tl

\_\_regex_peek_false_tl

\_\_regex_peek_reinsert:N

\_\_regex_peek_remove_end:n

\_\_regex_peek_replace:nnTF

\_\_regex_peek_replace_end:

\_\_regex_peek_replace_put:n

\_\_regex_peek_replace_put:n

\_\_regex_peek_replace_put:n

\_\_regex_peek_replace_token:n

\_\_regex_peek_replace_token:n

\_\_regex_peek_replace_token:n

\_\_regex_peek_replace_token:n

\_\_regex_peek_replace_token:n

\_\_regex_peek_remove_end:n

\_\_regex_peek_remove_end:n

\_\_regex_peek_remove_end:n

\_\_regex_peek_remove_end:n

\_\_regex_posix_xdigit:

\_\_regex_posix_upper:

\_\_regex_posix_digit:

\_\_regex_posix_ascii:

\_\_regex_posix_alpha:

\_\_regex_posix_alnum:

\_\_regex_pop_lr_states:

\_\_regex_peek_replacement_var:N

\_\_regex_peek_replacement_-

\_\_regex_peek_replacement_put_-

\_\_regex_peek_replacement_remove:

\_\_regex_peek_replacement_variable:n

\_\_regex_peek_replacement_variable:n

\_\_regex_pop_lr_states:

\_\_regex_pop_lr_states:

\_\_regexposix_alnum:

\_\_regexposix_alpha:

\_\_regexposix_asciicntrl:

\_\_regexposix_artist:

\_\_regexposix_basename:

\_\_regexposix_bytes:

\_\_regexposix_digits:

\_\_regexposix_graph:

\_\_regexposix_lower:

\_\_regexposix_print:

\_\_regexposix_punctuation:

\_\_regexposix_quotes:

\_\_regexposix_space:

\_\_regexposix_upper:

\_\_regexposix_word:

\_\_regexposix_xdigit:

\_\_regexprop:

\_\_regexprop_d:

\_\_regexprop_h:

\_\_regexprop_n:
Index

\__regex_use_state_and_submatches:w
. . . . . . . . . . . . . . . . 514, 6727, 6743
\__regex_Z_test: . . . . . . . . . . . . . .
463, 5339, 5341, 5358, 5978, 6081, 6512
\l__regex_zeroth_submatch_int . . .
. . . . . . . . . 533, 539, 7311, 7398,
7400, 7403, 7406, 7536, 7544, 7556,
7559, 7579, 7584, 7588, 7800, 7804
register commands:
register_luadata . . . . . . . . . . . . 11724
\relax . . . . . . . . . . . . . . . . . 4, 21, 25,
30, 65, 67, 68, 69, 80, 87, 108, 121,
122, 123, 124, 125, 126, 127, 128,
129, 130, 131, 133, 134, 135, 136,
137, 138, 139, 140, 141, 142, 143, 424
\relpenalty . . . . . . . . . . . . . . . . . . . . 425
\RequirePackage . . . . . . . . . . . . . . . . 111
\resettimer . . . . . . . . . . . . . . . . . . . . 782
reverse commands:
\reverse_if:N . . . . . . . . . . . . . 27,
712, 807, 808, 1002, 1417, 4228,
4490, 4491, 4508, 4509, 4514, 4515,
8376, 13530, 16895, 17045, 17047,
17049, 17051, 17106, 19888, 19893,
19897, 19899, 22998, 26578, 27400,
27423, 28909, 28911, 28933, 28957
\right . . . . . . . . . . . . . . . . . . . . . . . . 426
\rightghost . . . . . . . . . . . . . . . . . . . . 892
\righthyphenmin . . . . . . . . . . . . . . . . 427
\rightmarginkern . . . . . . . . . . . . . . . 692
\rightskip . . . . . . . . . . . . . . . . . . . . 428
\rmfamily . . . . . . . . . . . . . . . . . . . . 31341
\romannumeral . . . . . . . . . . . . . . . . . . 429
round . . . . . . . . . . . . . . . . . . . . . . . . . 252
\rpcode . . . . . . . . . . . . . . . . . . . . . . . 693
S
\saveboxresource . . . . . . . . . . . . . . . 935
\savecatcodetable . . . . . . . . . . . . . . 893
\saveimageresource . . . . . . . . . . . . . . 936
\savepos . . . . . . . . . . . . . . . . . . . . . . 934
\savinghyphcodes . . . . . . . . . . . . . . . 560
\savingvdiscards . . . . . . . . . . . . . . . 561
scan commands:
\scan_align_safe_stop: . . . . . . . 36241
\scan_new:N . . . . . . . . . . 140, 696,
778, 3321, 3322, 3730, 8972, 8973,
10132, 10133, 10598, 12679, 12954,
12955, 12956, 12965, 12966, 13621,
16067, 16092, 16093, 16094, 16804,
16805, 17686, 17687, 18370, 18839,
18840, 19219, 19220, 19400, 19405,
19406, 19768, 19769, 20113, 20276,
20277, 20278, 20279, 20535, 20536,

1477
20537, 22114, 22117, 22118, 22119,
22120, 22122, 22123, 22124, 22125,
22126, 22225, 29021, 33685, 33694,
35237, 35492, 35493, 35971, 35972
\scan_stop: . . . . . . . . . . . . . . . 13,
22, 23, 140, 155, 344, 347, 364,
365, 367, 376, 382, 393, 442, 455,
464, 486, 557, 662, 666, 673, 674,
676, 681, 687, 712, 777, 807, 812,
863, 873, 875, 877, 887, 891, 892,
897, 998, 1002–1004, 1007, 1204,
160, 174, 1446, 1801, 1819, 1829,
1847, 1873, 2202, 2225, 2234, 2243,
2315, 2757, 2881, 2882, 2897, 2937,
2963, 2987, 3004, 3208, 3214, 3355,
3736, 3869, 3909, 3913, 3919, 3921,
3968, 3970, 4230, 4236, 4238, 4254,
4256, 4266, 4267, 4268, 4274, 4283,
4574, 4575, 4688, 4748, 4773, 4785,
4921, 5743, 6043, 6047, 6050, 6205,
6753, 7165, 7217, 7433, 8687, 8691,
8943, 9944, 9949, 10049, 10159,
10162, 10711, 10718, 11113, 11992,
12162, 12172, 12239, 12268, 12270,
12634, 12654, 12668, 13531, 13731,
14720, 15802, 16076, 16079, 16452,
17236, 18884, 18956, 19291, 19390,
19393, 19395, 19780, 19799, 19801,
19805, 19808, 19811, 19815, 19820,
19824, 20047, 20123, 20141, 20143,
20151, 20153, 20157, 20159, 20180,
20185, 20188, 20214, 20234, 20236,
20244, 20246, 20250, 20252, 20256,
21614, 21697, 21798, 21836, 21843,
22005, 22100, 22289, 22996, 23000,
23201, 23218, 23519, 23566, 23567,
23822, 23865, 23893, 23907, 24676,
26489, 26497, 27242, 27245, 27248,
27251, 27254, 27257, 27260, 27263,
27266, 28531, 28558, 29034, 29035,
31736, 31884, 33462, 36017, 36020
\s_stop . . . . . . 140, 778, 16079, 16088
scan internal commands:
\s__bool_mark . . . 35492, 35513, 35527
\s__bool_stop . . . 35492, 35513, 35527
\s__char_stop . . . . . . . . . . . . . . . .
. . 18370, 18701, 18706, 18747, 18749
\s__clist_mark . . . . 832, 835–837,
842, 17686, 17688, 17716, 17717,
17734, 17856, 17866, 17870, 17892,
17954, 17960, 17974, 17986, 17987,
17988, 17991, 17992, 17993, 18002,
18003, 18012, 18208, 18209, 18221,
18222, 18235, 18243, 18249, 18252


\seq_set_from_function:NNN
\seq_set_from_inline_x:NNN
\seq_item:n
\seq_if_in:
\seq_get_left:wnw
\seq_count_end:w
\l_tmpb_seq
\g_tmpb_seq
\l_tmpa_seq
\g_tmpa_seq
\seq_use:Nn
\seq_shuffle:N
\seq_show:N
\seq_set_split_keep_spaces:Nnn
\seq_set_map:NNn
\seq_set_item:Nnn
\seq_set_from_inline_x:Nnn
\l__seq_internal_b_int
\__seq_count:w
\seq_show:NN
\seq_shuffle:NN
\seq_count:NN
\seq_show:NN
\__seq_set_from_inline_x:NNnn
\__seq_reverse_item:nwn
\__seq_reverse:NN
\l__seq_remove_seq
\__seq_pop_left:wnwNNN
\__seq_pop_left:NNN
\__seq_pop_item_def:
\__seq_pop_item:NnNN
\__seq_pop:NNNN
\__seq_mapthread_function:wNN
\__seq_mapthread_function:Nnnwnn
\__seq_map_tokens:nw
\__seq_map_indexed:nNN
\seq_internal_commands:
\__seq_count:w
\__seq_count_end:w
\__seq_get_left:ww
\__seq_get_right_end:NNN
\__seq_get_right_loop:nn
\__seq_if_in:
\__seq_int_eval:w
\l__seq_internal_a_int
\l__seq_internal_a_tl
\l__seq_internal_b_int
\l__seq_internal_b_tl
\g__seq_internal_seq
\seq_item:n
\seq_reverse:NN
\seq_reverse_item:nn
\seq_set_filter:NNN
\seq_set_from_inline_x:NNn
\seq_set_item:NNN
\seq_set_item:NNNN
Index

\special ................................ 452
\splitbotmark .......................... 453
\splitbotmarks .......................... 566
\splitdiscards .......................... 567
\splitfirstmark ...................... 454
\splitfirstmarks .......................... 568
\splitmaxdepth .......................... 455
\splittopskip .......................... 456
sqrt .................................. 254
\SS ................................ 29176, 31071, 31444
\ss ................................ 29176, 31071, 31440

str commands:
\c_\_ampersand_str ............. 131, 13582
\c_at__sign_str ............. 131, 13582
\c_backslash_str .......... 131, 4700,
5288, 13582, 14284, 14286, 14309,
1438, 14340, 14372, 1338, 14385
\c_circumflex_str ........ 131, 13582
\c_colon_str ............ 131, 13582, 18947,
19124, 19130, 20627, 34239, 34244
\c_dollar_str ............. 131, 13582
\c_hash_str ............ 131, 13582,
14252, 14355, 14930, 14931, 14934,
14937, 28957, 28988, 28989, 28993
\c_left_brace_str ............
..................... 131, 465, 4763, 5161,
5165, 5185, 5198, 5222, 5696, 5707,
5711, 5790, 5814, 6916, 13652, 29274
\c_percent_str ........ 131, 13582, 14254, 14408
\c_right_brace_str ..... 131, 4700, 5171, 5191, 5204,
5714, 5718, 5811, 6913, 13652, 29292
str_byte .................. 13651
\str_case:nn ..... 122, 5428, 9080,
10997, 13130, 34774, 34812, 35705
\str_case:nnn .............................. 36247, 36249
\str_case:nnTF ........ 122,
894, 8598, 8633, 8947, 9044, 13130,
13135, 13140, 20810, 36248, 36250
\str_case:en ...... 123, 13133, 36252
\str_case:enTF .... 123,
2692, 5169, 13130, 13166, 13171,
14307, 36254, 36256, 36258, 36260
\str_case:x:nn ........ 36251
\str_case:x:nnn .......... 36253
\str_case:x:nnTF .......... 36255,
36259, 36259
\cнструct:nnnTF .... 261, 1178, 28760
\cstrclear:N ........ 120, 12973, 20609, 20752, 21210, 21211
\cstrclearnew:N ........ 120, 12973
\cstrcompare:nnTF ........ 123, 13085
\cstrcompare:p:nn ........ 123, 13085
\cstrconcat:NNN ........ 120, 12973
\str_count:NNN ..............
..................... 120, 8506, 8525, 8543, 8589,
8909, 11382, 11389, 11393, 11397,
13000, 13582, 13583, 13584, 13585,
13586, 13587, 13588, 13589, 13590,
13591, 13592, 13593, 13594, 14348,
14349, 14371, 20506, 20507, 20508,
20509, 20510, 20511, 20512, 35703
\str_convert_pdfname:n ....
.............................. 134, 14906, 34534
\str_count:N ............ 125, 4168, 9093,
9094, 9276, 9277, 10247, 10325, 13472
\str_count:n ............ 125, 4162, 13472
\str_countignore_spaces:n ..
............................ 125, 710, 3776, 13472
\str_count_spaces:N ...... 125, 13452
\str_count_spaces:n ......
............................ 125, 710, 13452, 13478
\strdeclareeight_bit_encoding:mm
.............................. 36323
\str_end .................. 14809
\str_error .................. 13651
\strfold_case:n .......... 36311
\strfoldcase:n ..........
...................... 129, 130, 182, 265, 13540, 23187,
34426, 36319, 36320, 36321, 36322
\strgclear:N .......... 120, 12973
\strgclearnew:N .......... 12973
\strgconcat:NNN .......... 120, 12973
\strgput_left:N ........ 121, 13000
\strgput_right:N ........ 121, 13000
\strgremove_all:N ........ 128, 13070
\strgremove_once:N .... 128, 13064
\strgreplace_all:N ......
............................. 128, 13024, 13073
\strgreplace_once:N ......
............................. 128, 13024, 13067
\str_gset:N ........ 226, 21013
\str_gset:N ........ 128, 13024, 13067
\str_gset:NNn .......... 226, 21013
\str_gset:NNn .......... 121, 11140, 11141, 11142, 13000
\str_gset_convert:NNnn ..... 134, 13790
\str_gset_convert:NNnnTF .. 134, 13790
\str_gset_eq:N ........
............................. 120, 11127, 11128, 11129, 12973
\str_gset_x:N ........ 226, 21013
\strhead:N ........ 126, 711, 13510
\strhead:n ..........
............................. 126, 685, 711, 12563, 12608, 13510
\strheadignore_spaces:n .... 126, 13510
\str_if_empty:NTF ....
............................. 121, 10860, 13076, 20565, 20588, 21435
\str_if_empty:p:N .... 121, 13076
\str_if_eq:n ........ 198, 878, 886
\str_if_eq:NNTF ....
............................. 121, 701, 13109
Index
\str_if_eq:nnTF . . . . . . . . . . 103,
104, 122, 123, 201, 203, 784, 864,
2713, 8003, 8140, 8561, 8582, 8613,
8782, 9416, 9459, 9735, 9738, 11183,
11246, 11261, 13097, 13157, 13185,
14691, 14694, 14849, 14852, 16270,
18950, 19007, 19561, 19654, 20164,
20660, 21269, 23057, 23130, 24357,
28968, 28987, 28991, 29383, 29437,
29750, 29783, 31278, 33851, 34133,
34194, 34228, 34273, 34331, 34648,
34658, 36093, 36264, 36266, 36268
\str_if_eq_p:NN . . . . . . . . 121, 13109
\str_if_eq_p:nn . . . . . . . . . . . . . .
122, 8521, 8548, 8549, 8621, 8622,
8917, 8919, 11401, 13097, 29442,
32570, 33454, 34127, 35240, 36262
\str_if_eq_x:nnTF 36263, 36265, 36267
\str_if_eq_x_p:nn . . . . . . . . . . . 36261
\str_if_exist:NTF . . . 120, 8580, 13076
\str_if_exist_p:N . . . . . . . 120, 13076
\str_if_in:NnTF . . . . . . . . 122, 13116
\str_if_in:nnTF 122, 3244, 13116, 28749
\str_item:Nn . . . . . . . . . . . 126, 13314
\str_item:nn . . . . 126, 706, 710, 13314
\str_item_ignore_spaces:nn . . . . .
. . . . . . . . . . . . . . . . 126, 706, 13314
\str_log:N . . . . . . . . . . . . . 130, 13599
\str_log:n . . . . . . . . . . . . . 130, 13599
\str_lower_case:n . . . . . . . . . . . 36311
\str_lowercase:n . . . . . . . . . . 129,
265, 13540, 36311, 36312, 36313, 36314
\str_map_break: 124, 5937, 13191, 13265
\str_map_break:n 124, 125, 3248, 13191
\str_map_function:NN
123, 705, 13191
\str_map_function:nN . . . . . . . . . .
. . . 123, 124, 703, 5930, 13191, 14909
\str_map_inline:Nn . . . . . . 124, 13191
\str_map_inline:nn . . . . . . . . . . . .
. . . . . . . . . . . 124, 3242, 6614, 13191
\str_map_tokens:Nn . . . . . . 124, 13259
\str_map_tokens:nn . . . . . . 124, 13259
\str_map_variable:NNn . . . 124, 13191
\str_map_variable:nNn . . . 124, 13191
\str_new:N . 120, 8970, 8971, 10562,
10563, 10564, 10593, 10594, 10595,
12973, 13595, 13596, 13597, 13598,
20516, 20518, 20521, 20523, 20526
str_overflow . . . . . . . . . . . . . . . 14809
\str_put_left:Nn . . . . . . . . 121, 13000
\str_put_right:Nn . . . . . . . 121, 13000
\str_range:Nnn . . . . . . . . . 127, 13375
\str_range:nnn 93, 127, 710, 4165, 13375

1486
\str_range_ignore_spaces:nnn . . .
. . . . . . . . . . . . . . . . . . . . 127, 13375
\str_remove_all:Nn . . . . . . 128, 13070
\str_remove_once:Nn . . . . . 128, 13064
\str_replace_all:Nnn 128, 13024, 13071
\str_replace_once:Nnn . . . . . . . . .
. . . . . . . . . . . . . . 128, 13024, 13065
.str_set:N . . . . . . . . . . . . . 226, 21013
\str_set:Nn . . . . . . . . . . 121, 128,
226, 9056, 9057, 9264, 9265, 11196,
11197, 11198, 13000, 13250, 20544,
20546, 21076, 21078, 21219, 21344
\str_set_convert:Nnnn . . . . . . . . .
. . . . . . . . . 134, 135, 720, 731, 13790
\str_set_convert:NnnnTF . . . . . . . .
. . . . . . . . . . . . . . . . 134, 720, 13790
\str_set_eq:NN . . . . . . . . . 120, 12973
.str_set_x:N . . . . . . . . . . . 226, 21013
\str_show:N . . . . . . . . . . . . 130, 13599
\str_show:n . . . . . . . . . . . . 130, 13599
\str_tail:N . . . . . . . . . . . . 126, 13525
\str_tail:n . . . 126, 430, 13525, 29070
\str_tail_ignore_spaces:n 126, 13525
\str_upper_case:n . . . . . . . . . . . 36311
\str_uppercase:n . . . . . . . . . . 129,
265, 13540, 36315, 36316, 36317, 36318
\str_use:N . . . . . . . . . . . . . 125, 12973
\c_tilde_str . . . . . . . . . . . 131, 13582
\g_tmpa_str . . . . . . . . . . . . 131, 13595
\l_tmpa_str . . . . . . . . 128, 131, 13595
\g_tmpb_str . . . . . . . . . . . . 131, 13595
\l_tmpb_str . . . . . . . . . . . . 131, 13595
\c_underscore_str . . . . . . . 131, 13582
\c_zero_str . . . . . . . . . . . . 131, 13582
str internal commands:
\g__str_alias_prop . 723, 13623, 13861
\c__str_byte_-1_tl . . . . . . . . . . 13704
\c__str_byte_0_tl . . . . . . . . . . . 13704
\c__str_byte_1_tl . . . . . . . . . . . 13704
\c__str_byte_255_tl . . . . . . . . . 13704
\c__str_byte_⟨number ⟩_tl . . . . . . 718
\__str_case:nnTF . . . . . . . . . . . . 13130
\__str_case:nw . . . . . . . . . . . . . 13130
\__str_case_e:nnTF . . . . . . . . . . 13130
\__str_case_e:nw . . . . . . . . . . . . 13130
\__str_case_end:nw . . . . . . . . . . 13130
\__str_change_case:nn . . . . . . . 13540
\__str_change_case_aux:nn . . . . 13540
\__str_change_case_char:nN . . . 13540
\__str_change_case_end:nw . . . . 13540
\__str_change_case_end:wn . . . . . .
. . . . . . . . . . . . . . . . . . 13559, 13577
\__str_change_case_loop:nw . . . 13540
\__str_change_case_output:nw . 13540


Index

30268, 30275, 30353, 30431, 30509, 30532, 30568, 31384, 31469, 36212
\sys_if_engine_pdftex:TF

8551, 8931, 29933, 36228, 36230, 36232
\sys_if_engine_pdftex:p

71, 8506, 8565, 36226
\sys_if_engine_pdftex:TF

71, 8506, 8529, 8554
\sys_if_engine_putex:p

71, 3794, 8506
\sys_if_engine_putex:TF

71, 8506, 8530, 8555
\sys_if_engine_putex:p

71, 3795, 8506
\sys_if_engine_xetex:TF

6, 71, 2698, 8506, 8528, 8553, 8596, 8926, 8936, 36226, 36228, 36290
\sys_if_engine_xetex:p

71, 8506, 8788, 13642, 13914, 13938, 13960, 14130, 14913, 28782, 28808, 28870, 29804, 29878, 29918, 29979, 30015, 30221, 30269, 30276, 30354, 30432, 30510, 30533, 30569, 31385, 31470, 36284
\sys_if_output_dvi:TF

71, 8907
\sys_if_output_dvi:p

71, 8907
\sys_if_output_pdf:TF

71, 8611, 8907, 8929
\sys_if_output_pdf:p

71, 8907
\sys_if_platform_unix:TF

72, 8577, 11398
\sys_if_platform_unix:p

72, 8577, 11398
\sys_if_platform_windows:TF

72, 8577, 11398
\sys_if_platform_windows:p

72, 8577, 11398
\sys_if_rand_exist:TF

305, 571, 8575, 8811, 8825, 22031, 28015, 28039
\sys_if_rand_exist:p

305, 8575
\sys_if_shell:

73
\sys_if_shell:TF

73, 8670, 8887, 35759
\sys_if_shell:p

73, 8887
\sys_if_shell Restricted:TF

73, 8887
\sys_if_shell Restricted:p

73, 8887
\sys_if_shell unrestricted:TF

73, 8887
\sys_if_shell unrestricted:p

73, 8887
\sys_if_timer_exist:TF

8838
\sys_if_timer_exist:p

8838
\c_sys_jobname_str

70, 91, 579, 8773, 36116
\sys_load_backend:n

73, 8577
\sys_load_debug:

74, 8649
\sys_load_deprecation:

74, 8649
\c_sys_minute_int

70, 8777
\c_sys_month_int

70, 8777
\c_sys_output_str

71, 8907
\c_sys_platform_str

72, 8577, 11380, 11401
\c_sys_rand.seed:

72, 146, 255, 8809
\c_sys_shell_escape_int

73, 8875, 8890, 8892, 8894
\c_sys_shell_now:n

73, 8705
\c_sys_shell_shipout:n

73, 8736
\c_sys_timer:

71, 8838
\c_sys_year_int

70, 8777
\sys internal commands:

\g__sys_backend_tl

8587, 8588, 8589, 8921
\_sys_const:n

8490, 8520, 8575, 8873, 8889, 8891, 8916, 9198
\g__sys_debug bool

8647, 8651, 8653
\g__sys_deprecation bool

1772, 8647, 8657, 8659
\_sys_elapsedtime:

8838
\_sys_everyjob:n

8767, 8775, 8777, 8809, 8823, 8875, 8887, 8896
\g__sys_everyjob_tl

8767
\_sys_finalize:n

8988, 8907, 8922, 8939
\g__sys_finalize_tl

8988
\_sys_get:mm

8663
\_sys_get_do:N

8663
\_sys_internal_tl

8601
\_sys_load_backend_check:N

8577
\c__sys_marker_tl

8602, 8686, 8698
\c__sys_shell_now:n

8705
\c__sys_shell_shipout:n

8736
\c__sys_shell_stream_int

8703, 8732, 8764
\_sys_tmp:w

8780, 8801, 8803, 8804, 8805, 8806
syst commands:

\c__sys_catcodes n

28738, 28742
\c__sys_last_allocated_toks

3382
\t

29163, 31509
\tabskip

458
\tagcode

695
tan

253
tand

253
tate

1157
\tbaselineshift

1158
\c_ten

36139
\TeX

1125
<table>
<thead>
<tr>
<th>Command</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>\input</td>
<td>336</td>
</tr>
<tr>
<td>\input@path</td>
<td>92, 632, 10745, 10747, 10845, 10847</td>
</tr>
<tr>
<td>\italicccr</td>
<td>866</td>
</tr>
<tr>
<td>\jobname</td>
<td>70, 576</td>
</tr>
<tr>
<td>\lastnamedcs</td>
<td>366</td>
</tr>
<tr>
<td>\lccode</td>
<td>429, 433, 822, 1188</td>
</tr>
<tr>
<td>\leavevmode</td>
<td>28</td>
</tr>
<tr>
<td>\let</td>
<td>316</td>
</tr>
<tr>
<td>\letclrcode</td>
<td>851</td>
</tr>
<tr>
<td>\LL@shipout</td>
<td>1214</td>
</tr>
<tr>
<td>\loctoks</td>
<td>411</td>
</tr>
<tr>
<td>\long</td>
<td>5, 197, 391, 695</td>
</tr>
<tr>
<td>\lower</td>
<td>1347</td>
</tr>
<tr>
<td>\lowercase</td>
<td>529, 529</td>
</tr>
<tr>
<td>\luaescapestring</td>
<td>97</td>
</tr>
<tr>
<td>\makeatletter</td>
<td>9</td>
</tr>
<tr>
<td>\MakeUppercase</td>
<td>1230</td>
</tr>
<tr>
<td>\mathchar</td>
<td>197</td>
</tr>
<tr>
<td>\mathchardef</td>
<td>189, 804, 1193</td>
</tr>
<tr>
<td>\meaning</td>
<td>20</td>
</tr>
<tr>
<td>\message</td>
<td>33</td>
</tr>
<tr>
<td>\newcatcodetable</td>
<td>1176</td>
</tr>
<tr>
<td>\newif</td>
<td>63, 99</td>
</tr>
<tr>
<td>\newlinechar</td>
<td>116, 312, 366, 586, 616, 667, 668, 695</td>
</tr>
<tr>
<td>\newread</td>
<td>607</td>
</tr>
<tr>
<td>\newtoks</td>
<td>43, 422, 449</td>
</tr>
<tr>
<td>\newwrite</td>
<td>613</td>
</tr>
<tr>
<td>\nexpand</td>
<td>39, 195, 390–393</td>
</tr>
<tr>
<td>\nullfont</td>
<td>866</td>
</tr>
<tr>
<td>\number</td>
<td>166, 801, 1059</td>
</tr>
<tr>
<td>\numexpr</td>
<td>344, 394</td>
</tr>
<tr>
<td>\openafihipout</td>
<td>1216</td>
</tr>
<tr>
<td>\or</td>
<td>166</td>
</tr>
<tr>
<td>\outer</td>
<td>197, 313, 441, 427, 607, 619, 875, 1366, 1368</td>
</tr>
<tr>
<td>\parindent</td>
<td>28</td>
</tr>
<tr>
<td>\pdfescapehex</td>
<td>731</td>
</tr>
<tr>
<td>\pdfescapename</td>
<td>133, 731</td>
</tr>
<tr>
<td>\pdfescapestring</td>
<td>133, 731</td>
</tr>
<tr>
<td>\pdffilesize</td>
<td>631, 632</td>
</tr>
<tr>
<td>\pdfmapfile</td>
<td>338</td>
</tr>
<tr>
<td>\pdfmapline</td>
<td>338</td>
</tr>
<tr>
<td>\pdfstrcmp</td>
<td>310, 311, 313, 327, 655</td>
</tr>
<tr>
<td>\pdfunidencodetable</td>
<td>255</td>
</tr>
<tr>
<td>\pdfuniformdeviate</td>
<td>277</td>
</tr>
<tr>
<td>\pgfpages@originalshipout</td>
<td>1217</td>
</tr>
<tr>
<td>\pi</td>
<td>248</td>
</tr>
<tr>
<td>\pre@shipout</td>
<td>1218</td>
</tr>
<tr>
<td>\protect</td>
<td>619, 1002, 1003, 1241</td>
</tr>
<tr>
<td>\protected</td>
<td>197, 391, 695</td>
</tr>
<tr>
<td>\ProvidesClass</td>
<td>9</td>
</tr>
<tr>
<td>\ProvidesFile</td>
<td>9</td>
</tr>
<tr>
<td>\ProvidesPackage</td>
<td>9</td>
</tr>
<tr>
<td>\quitvmode</td>
<td>378</td>
</tr>
<tr>
<td>\read</td>
<td>85, 611</td>
</tr>
<tr>
<td>\relax</td>
<td>86, 611</td>
</tr>
<tr>
<td>\RequirePackage</td>
<td>10, 313, 626</td>
</tr>
<tr>
<td>\reserveinserts</td>
<td>313</td>
</tr>
<tr>
<td>\romanumeral</td>
<td>41, 959</td>
</tr>
<tr>
<td>\savecatcodetable</td>
<td>1178</td>
</tr>
<tr>
<td>\strangepar</td>
<td>117, 135, 631, 666</td>
</tr>
<tr>
<td>\shipout</td>
<td>336</td>
</tr>
<tr>
<td>\show</td>
<td>20, 108, 375</td>
</tr>
<tr>
<td>\showbox</td>
<td>1256</td>
</tr>
<tr>
<td>\showgroups</td>
<td>14, 376</td>
</tr>
<tr>
<td>\showthe</td>
<td>375, 821, 900, 904, 906</td>
</tr>
<tr>
<td>\showtokens</td>
<td>108, 591, 695</td>
</tr>
<tr>
<td>\sin</td>
<td>248</td>
</tr>
<tr>
<td>\space</td>
<td>866</td>
</tr>
<tr>
<td>\splitbotmark</td>
<td>866</td>
</tr>
<tr>
<td>\splitfirstmark</td>
<td>866</td>
</tr>
<tr>
<td>\SS</td>
<td>1244</td>
</tr>
<tr>
<td>\strcmp</td>
<td>310, 327</td>
</tr>
<tr>
<td>\string</td>
<td>186, 499–431</td>
</tr>
<tr>
<td>\tenrm</td>
<td>195</td>
</tr>
<tr>
<td>\the</td>
<td>157, 195, 212, 216, 219, 382, 390–392, 394, 395, 801, 1253</td>
</tr>
<tr>
<td>\toks@</td>
<td>396</td>
</tr>
<tr>
<td>\toks</td>
<td>43, 146, 166, 394–396, 409–417, 422, 428, 430, 432, 433, 445, 453, 454, 456, 457, 555, 786</td>
</tr>
<tr>
<td>\topmark</td>
<td>196, 866</td>
</tr>
<tr>
<td>\tracingfonts</td>
<td>337</td>
</tr>
<tr>
<td>\tracingnesting</td>
<td>630, 666</td>
</tr>
<tr>
<td>\tracingonline</td>
<td>1256</td>
</tr>
<tr>
<td>\typeout</td>
<td>619</td>
</tr>
<tr>
<td>\ucode</td>
<td>1188</td>
</tr>
<tr>
<td>\Ucharcat</td>
<td>853</td>
</tr>
<tr>
<td>\unbox</td>
<td>277</td>
</tr>
<tr>
<td>\unboxtext</td>
<td>274</td>
</tr>
<tr>
<td>\unhcopy</td>
<td>255</td>
</tr>
<tr>
<td>\unidencodetable</td>
<td>255</td>
</tr>
<tr>
<td>\unless</td>
<td>27</td>
</tr>
<tr>
<td>\unvbox</td>
<td>277</td>
</tr>
<tr>
<td>\unvcopy</td>
<td>276</td>
</tr>
<tr>
<td>\uppercase</td>
<td>528</td>
</tr>
<tr>
<td>\usepackage</td>
<td>626</td>
</tr>
</tbody>
</table>
Index

\valign \value \verb \verso\orig\shipout \vskip \vtop \wd \write

tex commands:
\tex_above:D \tex_afterassignment:D \tex_aftergroup:D \tex_alignmark:D
\tex_aligntab:D \tex_above:D \tex_atop:D \tex_atopwithdelims:D
\tex_attenuation:D \tex_attribute:D \tex_automatichyphenmode:D
\tex_automaticdiscretionary:D \tex_autoxspacing:D \tex硊
\tex_above:D \tex_adjustspacing:D \tex_adjustedmarks:D \tex_adjustedmarks:
\tex_align:D \tex_align:D \tex_align:D \tex_align:D
\tex_assist:D \tex_advancing:D \tex_adjacent:D \tex_adjacent:D
\tex_advancing:D \tex_advancing:D \tex_advancing:D \tex_advancing:D
\tex_advancing:D \tex_advancing:D \tex_advancing:D \tex_advancing:D
\tex_advancing:D \tex_advancing:D \tex_advancing:D \tex_advancing:D
<table>
<thead>
<tr>
<th>\text{Index}</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{tex_Umathcharnum:D}</td>
</tr>
<tr>
<td>\text{tex_Umathcharnumdef:D}</td>
</tr>
<tr>
<td>\text{tex_Umathcharslot:D}</td>
</tr>
<tr>
<td>\text{tex_Umathclosebinspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathcloseclosespacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathcloseinnernerspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathcloseopenspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathcloseordclosespacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathcloseopenpunctspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathclosepunctpunctspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathconnectoroverlamin:D}</td>
</tr>
<tr>
<td>\text{tex_Umathfractiondelsize:D}</td>
</tr>
<tr>
<td>\text{tex_Umathfractiondenomdown:D}</td>
</tr>
<tr>
<td>\text{tex_Umathfractiondenomvgap:D}</td>
</tr>
<tr>
<td>\text{tex_Umathfractionnumvgap:D}</td>
</tr>
<tr>
<td>\text{tex_Umathfractionrule:D}</td>
</tr>
<tr>
<td>\text{tex_Umathinnerbinspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathinnerclosespacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathinnerinnernerspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathinneropenspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathinnerordpacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathinnerordspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathinnerpunctpunctspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathinnerpunctspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathlimitabovexgap:D}</td>
</tr>
<tr>
<td>\text{tex_Umathlimitabovexv:gap:D}</td>
</tr>
<tr>
<td>\text{tex_Umathlimitabovexv:gap:D}</td>
</tr>
<tr>
<td>\text{tex_Umathlimitbelowxgap:D}</td>
</tr>
<tr>
<td>\text{tex_Umathlimitbelowxv:gap:D}</td>
</tr>
<tr>
<td>\text{tex_Umathlimitbelowxv:gap:D}</td>
</tr>
<tr>
<td>\text{tex_Umathlimitsubfactor:D}</td>
</tr>
<tr>
<td>\text{tex_Umathlimitsubfactor:D}</td>
</tr>
<tr>
<td>\text{tex_Umathopbinspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathopclosespacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathopenbinspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathopenpenclosespacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathopeninnernerspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathopenpunctpunctspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathopenpunctspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathopenrelspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathoperatorsize:D}</td>
</tr>
<tr>
<td>\text{tex_Umathpinnernerspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathpinnerspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathpinnerspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathpinnerspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathpinnerspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathpinnerspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umathpinnerspacing:D}</td>
</tr>
<tr>
<td>\text{tex_Umatheodvalues:D}</td>
</tr>
<tr>
<td>\text{tex_Umatheodvalues:D}</td>
</tr>
<tr>
<td>\text{tex_Umatheodvalues:D}</td>
</tr>
<tr>
<td>\text{tex_Umatheodvalues:D}</td>
</tr>
</tbody>
</table>
\_._text_token_to_explicit_char:N \ 29029
\_._text_token_to_explicit_cs:N \ 29029
\_._text_token_to_explicit_cs_- aux:N \ 29029
\_text_use_i_delimit_by_q- recursion_stop:nw ...........
\ 29026, 29317, 29385, 29407, 29427, 29683, 29752, 31176, 31245
\textbaselineshiftfactor \ 1159
\textbf \ 31329
\textit \ 459
\texttt \ 460
\textsf \ 463
\textsl \ 462
\textup \ 31327
\textulc \ 31328
\textsc \ 31334
\textnormal \ 31326
\th \ 1161
\th \ 569
\TH \ 29177, 31072, 31428
\th \ 29177, 31072, 31441
\the \ 42, 121, 122, 123,
\ 124, 125, 126, 127, 128, 129, 130, 461
\th \ 462
\thinmuskip \ 31335
\th \ 463
\_c_thirteen \ 36145
\_c_thirty_two \ 36153
\thousand commands:
\_c_one_thousand \ 36161
\_c_ten_thousand \ 36163
\_c_three \ 36125
\_time \ 464, 1309, 18801, 8803
\_tiny \ 31360, 33457
tl commands:
\_c_catcode_active_space_tl \ 307, 35913
\_c_catcode_active_tl \ 185, 862, 18850, 18919
\_c_catcode_other_space_tl \ 182, 621, 10270,
\ 10314, 10394, 10483, 10559, 18837
\_c_empty_tl \ 117, 815, 828,
\ 2678, 9260, 11795, 11811, 11813,
\ 11848, 12152, 17331, 17337, 17684,
\ 17700, 28909, 35823, 35860, 35879
\_c_job_name_tl \ 36115
\_c_novaluen_tl \ 104, 117, 11849, 12261
internal commands:

\tl_to_uppercase:n ............ 36279
\tl_trims_spaces:N ........... 108, 12454
\tl_trims_spaces:n ........... 108, 683, 940, 10674, 12454, 16163, 16165, 36091
\tl_trims_spaces_apply:nN ........ 108, 914, 10671, 12454, 17696, 18157, 18244, 18317
\tl_upper_case:n ............ 36331
\tl_upper_case:nn ............ 36331
\tl_upper_case:N .......... 167, 212, 216, 219, 8769, 8901, 12402, 20804
\bg temporarily ............... 118, 12957
\l_tmpa_tl .................. 6, 116, 117, 1206, 1208, 1225, 1308, 1310, 1314, 1316, 1320, 1322, 1326, 1328, 12959
\g_tmpb_tl .................. 118, 12957
\l_tmpb_tl .................. 117, 1207, 1208, 1223, 1225, 1309, 1310, 1315, 1316, 1321, 1322, 1327, 1328, 12959
\tl internal commands:
\_\_\_tl_act:nNN ................. 689, 690, 12428, 12680, 12747
\_\_\_tl_act_count_group:n 12430, 12437
\_\_\_tl_act_count_group:nn ... 12424
\_\_\_tl_act_count_normal:n 12429, 12435
\_\_\_tl_act_count_normal:nn ... 12424
\_\_\_tl_act_count_space: ... 12431, 12436
\_\_\_tl_act_count_space:n 12424
\_\_\_tl_act_end:wn ........... 681, 12680
\_\_\_tl_act_group:nWWNNN .... 12680
\_\_\_tl_act_if_head_is_space:n TF ... ............... 689, 12680
\_\_\_tl_act_if_head_is_space:w 12680
\_\_\_tl_act_if_head_is_space-_true:w .................. 12680
\_\_\_tl_act_loop:w ........... 689, 12680
\_\_\_tl_act_normal:WWNNN .... 12680
\_\_\_tl_act_output:n ........ 690, 12680

\_\_\_tl_result:n \ldots 690, 12715, 12736, 12738, 12739, 12740, 12741
\_\_\_tl_act_reverse ............. 690
\_\_\_tl_act_reverse_output:n ....... 12680, 12756, 12758, 12760
\_\_\_tl_act_space:WWNNN .... 689, 12680
\_\_\_tl_analysis:n ............ 429, 438, 3777, 4056, 4089, 4101
\_\_\_tl_analysis:a:n ........ 3781, 3806
\_\_\_tl_analysis_a_bgroup:w 3837, 3859
\_\_\_tl_analysis_a_cs:w ........ 3916
\_\_\_tl_analysis_a_egroup:w ... 3839, 3859
\_\_\_tl_analysis_a_group:nw .... 3859
\_\_\_tl_analysis_a_group:aux:w ... 3859
\_\_\_tl_analysis_a_group_auxii:w ... 3859
\_\_\_tl_analysis_a_group_test:w ... 3859
\_\_\_tl_analysis_a_loop:w ........ 3813, 3816, 3857, 3899, 3913, 3931
\_\_\_tl_analysis_a_safe:N ......... 3838, 3880, 3916
\_\_\_tl_analysis_a_space:w ........ 3836, 3842
\_\_\_tl_analysis_a_space_test:w .... 431, 3842
\_\_\_tl_analysis_a_store: ...... 432, 3853, 3895, 3901
\_\_\_tl_analysis_a_type:w ........ 3817, 3818
\_\_\_tl_analysis_b:n ........... 3782, 3944
\_\_\_tl_analysis_b_char:Nww .......... 3971, 3977
\_\_\_tl_analysis_b_cs:Nww .... 3973, 4001
\_\_\_tl_analysis_b_cs_test:ww ... 4001
\_\_\_tl_analysis_b_loop:w .......... 437, 3944, 4047, 4052
\_\_\_tl_analysis_b_normal:WWNN .... 3957, 4022
\_\_\_tl_analysis_b_normals:ww ........ 436, 3954, 3957, 3998, 4008
\_\_\_tl_analysis_b_special:w 3960, 4019
\_\_\_tl_analysis_b_special_char:WN ........... 4019
\_\_\_tl_analysis_b_special_space:w ...... 4180, 4295, 4354
\_\_\_tl_analysis_char_arg:Nw .... 4180
\_\_\_tl_analysis_char_arg_aux:Nw 4180
\_\_\_tl_analysis_char_token .... 426, 431, 3751, 3846, 3851, 3889, 3894
\_\_\_tl_analysis_cs_count:n .................. 3761, 3939, 4004
\_\_\_tl_analysis_cs_count:w .... 3761
\_\_\_tl_analysis_cs_count_end:w ........ 3761
\_\_\_tl_analysis_disable:n ......... 3768, 3808, 3874, 3927
Index

\_token_tmp:w ........ 864, 18972, 18981, 18982, 18983, 18984, 18985, 18986, 18987, 18988, 18991, 19025, 19026, 19027, 19028, 19030, 19032, 19033, 19034, 19035, 19036, 19037
\toks .................. 465, 19037
\toksapp ................. 905
\toksdef ................ 466, 3676
\tokspre ................ 19033, 19034, 19035, 19036, 19037
\tolerance .............. 906
\topmark .................. 467
\topmarks ................ 570
\topskip .................. 469
\tpack .................. 907
\tracingassigns .......... 571
\tracingcommands ........ 470
\tracingfonts ............ 938
\tracinggroups .......... 572
\tracingifs ............. 573
\tracinglastchars ....... 471
\tracingmacros .......... 472
\tracingnesting .......... 574
\tracingonline .......... 473
\tracingoutput .......... 474
\tracingpages ........... 475
\tracingparagraphs ..... 476
\tracingrestores ........ 477
\tracinngscantokens .... 575
\tracingstacklevels ..... 1186
\tracingstats ........... 478
\true ..................... 256
\trunc .................. 252
\ttfamily ............... 31343
\_c_twelve .............. 36143
\_c_two ................ 36123
\_c_two_hundred_fifty_five .......... 36157
\_c_two_hundred_fifty_six ....... 36159
\u .................. 29163, 31499, 31580, 31581, 31596, 31597, 31606, 31607, 31620, 31621, 31622, 31648, 31649, 31674, 31675
\uccode .................. 479
\Uchar .................. 940
\Ucharcat ............... 941
\uchyph ................ 480
\ucs ..................... 1174
\Udelcode ............... 942
\Udelcodenum ........... 943
\Udelimiter ............. 944
\Udelimiterover ........ 945
\Udelimiterunder ...... 946
\Uhextensible .......... 947
\Umathaccent ........... 948
\Umathaxis ............. 949
\Umathbinbinspacing .... 950
\Umathbinclosebinspacing .... 951
\Umathbininnerspacing .... 952
\Umathbinopenbinspacing .... 953
\Umathbinoppspacing .... 954
\Umathbinorbinspacing .... 955
\Umathbinpunctspacing .... 956
\Umathbinrelspacing .... 957
\Umathchar ............. 958
\Umathcharclass ......... 959
\Umathchardef .......... 960
\Umathcharfam .......... 961
\Umathcharnum .......... 962
\Umathcharnumdef ...... 963
\Umathcharslot ........ 964
\Umathclosebinspacing .... 965
\Umathcloseclosebinspacing .... 966
\Umathcloseinnerspacing .... 968
\Umathcloseopenbinspacing .... 970
\Umathcloseopspacing .... 971
\Umathcloseordspacing .... 972
\Umathclosepunctspacing .... 973
\Umathcloserelspacing .... 975
\Umathcode ............. 976
\Umathcodenum .......... 977
\Umathconnectoroverlapmin .... 978
\Umathfractiondelsize .... 980
\Umathfractiondenomdown .... 981
\Umathfractiondenomvgap .... 983
\Umathfractionnumup .... 985
\Umathfractionnumvgap .... 986
\Umathfractionrule .... 987
\Umathinnerbinspacing .... 988
\Umathinnernerspacing .... 989
\Umathinnernumspacing .... 991
\Umathinneropspacing .... 993
\Umathinnerordspacing .... 994
\Umathinnerrelspacing .... 998
\Umathlimitabovebgap ...... 999
\Umathlimitabovexkern ..... 1000
\Umathlimitbelowbgap ...... 1001
\Umathlimithtbelowxkern ...... 1002
\Umathlimithtbelowvgap ...... 1004
\Umathlimithtleftkern ...... 1003
\Umathlimithtleftvgap ...... 1004
\Umathlimithtopenbgap ...... 1005
\Umathlimithtopenxkern ...... 1006
\Umathlimithtopenvgap ...... 1007
\Umathopenspacing ...... 1008
\Umathopenbinspacing ...... 1009
\Umathopeninnerspacing ...... 1010
\Umathopenpunctspacing ...... 1011
<table>
<thead>
<tr>
<th>Command</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>\mathopenoperatorspacing</td>
<td>1027</td>
</tr>
<tr>
<td>\mathopclosespacing</td>
<td>1024</td>
</tr>
<tr>
<td>\mathunderdelimitervgap</td>
<td>1035</td>
</tr>
<tr>
<td>\mathunderdelimiterbgap</td>
<td>1037</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1039</td>
</tr>
<tr>
<td>\mathunderbarkern</td>
<td>1042</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1044</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1045</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1046</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1047</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1049</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1050</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1051</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1053</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1055</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1058</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1059</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1060</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1061</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1062</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1063</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1064</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1065</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1066</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1067</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1068</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1070</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1072</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1073</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1074</td>
</tr>
<tr>
<td>\mathunderbarrule</td>
<td>1075</td>
</tr>
</tbody>
</table>

For more commands, see the documentation.
Index

W
\wd ........................................ 506
\widowpenalties ............................ 578
\widowpenalty ............................... 507
\write ...................................... 508
\xdef .................................. 509
\xetex commands:
  \xetex_if_engine:TF ...................... 36285, 36287, 36289
  \xetex_if_engine_p: ....................... 36283
\XeTeXcharclass .............................. 716
\XeTeXcharglyph .............................. 717
\XeTeXcountfeatures ......................... 718
\XeTeXcountglyphs ............................ 719
\XeTeXcountselectors ......................... 720
\XeTeXcountvariations ........................ 721
\XeTeXdashbreakstate ........................ 722
\XeTeXdefaultencoding ....................... 723
\XeTeXfeaturecode ............................ 724
\XeTeXfeaturename ............................. 725
\XeTeXfindfeaturebyname ...................... 726
\XeTeXfindselectorbyname ..................... 727
\XeTeXfindvariationbyname .................... 728
\XeTeXfirstfontchar ......................... 729
\XeTeXglyph ................................ 730
\XeTeXglyphbounds ............................ 731
\XeTeXglyphindex .............................. 732
\XeTeXglyphname .............................. 733
\XeTeXgenerateactualtext ..................... 734
\XeTeXglyph .................. 36119
\XeTeXglyphbounds .................... 735
\XeTeXglyphindex ....................... 736
\XeTeXglyphname ....................... 737
\XeTeXinputencoding ......................... 738
\XeTeXinputnormalization ..................... 739
\XeTeXinterchartokenstate .................... 740
\XeTeXinterchartoks .......................... 741
\XeTeXisdefaultselector ..................... 742
\XeTeX getXexclusivefeature ................. 743
\XeTeXlastfontchar ......................... 744
\XeTeXlastlinebreaklocale .................... 745
\XeTeXlastlinebreakpenalty ................... 746
\XeTeXlinebreaklocale ....................... 747
\XeTeXlinebreakpenalty ...................... 748
\XeTeXlinebreakskip ........................ 749
\XeTeXOTcountfeatures ....................... 750
\XeTeXOTcountlanguages ....................... 751
\XeTeXOTcountscripts ......................... 752
\XeTeXOTfeaturetag ............................ 753
\XeTeXOTlangtag .............................. 754
\XeTeXOTscripttag ............................. 755
\XeTeXpdfdoccount ............................ 756
\XeTeXpdfpagecount ............................ 757
\XeTeXpdfpagecount ......................... 758
\XeTeXpicfile ................................ 759
\XeTeXpdffile ............................... 760
\XeTeXpdfpicfile ............................... 761
\XeTeXrevision ............................... 762
\XeTeXrevision ................................ 763
\XeTeXselectorname ............................ 764
\XeTeXselectionfont ......................... 765
\XeTeXsinglefont ...................... 766
\XeTeXTupwardsmode ............................ 767
\XeTeXversion ............................... 768
\XeTeXversion ................................ 769
\XeTeXXeversion .............................. 770
\XeTeXXvariation ................................ 771
\XeTeXXvariationmax ............................ 772
\XeTeXXvariationmin ............................ 773
\XeTeXXvariationname .......................... 774
\XeTeXXversion ................................ 775
\xkanjiskip ............................... 1162
\xleaders ............................... 510
\xleaders ................................ 511
\xspc ...................... 1163
\ybaselineskip .............................. 1164
\year .................................. 512, 1327, 8806
\yoko .................................. 1165
\c_zero ............................... 36119