

# Babel

Version 3.36.1837

2019/11/26

*Original author*

Johannes L. Braams

*Current maintainer*

Javier Bezos

The standard distribution of  $\text{\LaTeX}$  contains a number of document classes that are meant to be used, but also serve as examples for other users to create their own document classes. These document classes have become very popular among  $\text{\LaTeX}$  users. But it should be kept in mind that they were designed for American tastes and typography. At one time they even contained a number of hard-wired texts.

This manual describes babel, a package that makes use of the capabilities of  $\text{\TeX}$ , xetex and luatex to provide an environment in which documents can be typeset in a language other than US English, or in more than one language or script.

Current development is focused on Unicode engines (Xe $\text{\TeX}$  and Lua $\text{\TeX}$ ) and the so-called *complex scripts*. New features related to font selection, bidi writing, line breaking and so on are being added incrementally.

Babel provides support (total or partial) for about 200 languages, either as a “classical” package option or as an ini file. Furthermore, new languages can be created from scratch easily.

# Contents

|           |  |           |
|-----------|--|-----------|
| <b>I</b>  | <b>User guide</b>  | <b>4</b>  |
| <b>1</b>  | <b>The user interface</b>  | <b>4</b>  |
| 1.1       | Monolingual documents . . . . .  | 4         |
| 1.2       | Multilingual documents . . . . .   | 5         |
| 1.3       | Modifiers . . . . .  | 6         |
| 1.4       | xelatex and luatex . . . . .   | 7         |
| 1.5       | Troubleshooting . . . . .  | 8         |
| 1.6       | Plain . . . . .  | 8         |
| 1.7       | Basic language selectors . . . . .   | 8         |
| 1.8       | Auxiliary language selectors . . . . .   | 9         |
| 1.9       | More on selection . . . . .  | 10        |
| 1.10      | Shorthands . . . . .   | 11        |
| 1.11      | Package options . . . . .  | 14        |
| 1.12      | The base option . . . . .  | 16        |
| 1.13      | ini files . . . . .  | 17        |
| 1.14      | Selecting fonts . . . . .  | 24        |
| 1.15      | Modifying a language . . . . .   | 26        |
| 1.16      | Creating a language . . . . .  | 27        |
| 1.17      | Digits . . . . .   | 29        |
| 1.18      | Getting the current language name . . . . .                                      | 30        |
| 1.19      | Hyphenation and line breaking . . . . .  | 30        |
| 1.20      | Selecting scripts . . . . .  | 32        |
| 1.21      | Selecting directions . . . . .   | 33        |
| 1.22      | Language attributes . . . . .  | 37        |
| 1.23      | Hooks . . . . .  | 37        |
| 1.24      | Languages supported by babel with ldf files . . . . .                            | 38        |
| 1.25      | Unicode character properties in luatex . . . . .                                 | 39        |
| 1.26      | Tweaking some features . . . . .   | 40        |
| 1.27      | Tips, workarounds, known issues and notes . . . . .                              | 40        |
| 1.28      | Current and future work . . . . .  | 41        |
| 1.29      | Tentative and experimental code . . . . .  | 42        |
| <b>2</b>  | <b>Loading languages with language.dat</b>                                       | <b>42</b> |
| 2.1       | Format . . . . .   | 43        |
| <b>3</b>  | <b>The interface between the core of babel and the language definition files</b> | <b>43</b> |
| 3.1       | Guidelines for contributed languages . . . . .                                   | 45        |
| 3.2       | Basic macros . . . . .   | 45        |
| 3.3       | Skeleton . . . . .   | 46        |
| 3.4       | Support for active characters . . . . .  | 47        |
| 3.5       | Support for saving macro definitions . . . . .                                   | 48        |
| 3.6       | Support for extending macros . . . . .   | 48        |
| 3.7       | Macros common to a number of languages . . . . .                                 | 48        |
| 3.8       | Encoding-dependent strings . . . . .   | 49        |
| <b>4</b>  | <b>Changes</b>   | <b>52</b> |
| 4.1       | Changes in babel version 3.9 . . . . .   | 52        |
| <b>II</b> | <b>Source code</b>   | <b>53</b> |
| <b>5</b>  | <b>Identification and loading of required files</b>                              | <b>53</b> |

|           |  |            |
|-----------|--|------------|
| <b>6</b>  | <b>locale directory</b>  | <b>53</b>  |
| <b>7</b>  | <b>Tools</b>   | <b>54</b>  |
| 7.1       | Multiple languages . . . . .                                       | 58         |
| <b>8</b>  | <b>The Package File (<math>\LaTeX</math>, babel.sty)</b>           | <b>59</b>  |
| 8.1       | base . . . . .   | 59         |
| 8.2       | key=value options and other general option . . . . .               | 61         |
| 8.3       | Conditional loading of shorthands . . . . .                        | 63         |
| 8.4       | Language options . . . . .   | 64         |
| <b>9</b>  | <b>The kernel of Babel (babel.def, common)</b>                     | <b>67</b>  |
| 9.1       | Tools . . . . .  | 67         |
| 9.2       | Hooks . . . . .  | 69         |
| 9.3       | Setting up language files . . . . .                                | 71         |
| 9.4       | Shorthands . . . . .   | 73         |
| 9.5       | Language attributes . . . . .                                      | 83         |
| 9.6       | Support for saving macro definitions . . . . .                     | 85         |
| 9.7       | Short tags . . . . .   | 86         |
| 9.8       | Hyphens . . . . .  | 86         |
| 9.9       | Multiencoding strings . . . . .                                    | 88         |
| 9.10      | Macros common to a number of languages . . . . .                   | 94         |
| 9.11      | Making glyphs available . . . . .                                  | 94         |
| 9.11.1    | Quotation marks . . . . .  | 94         |
| 9.11.2    | Letters . . . . .  | 95         |
| 9.11.3    | Shorthands for quotation marks . . . . .                           | 96         |
| 9.11.4    | Umlauts and tremas . . . . .                                       | 97         |
| 9.12      | Layout . . . . .   | 98         |
| 9.13      | Load engine specific macros . . . . .                              | 99         |
| 9.14      | Creating languages . . . . .                                       | 99         |
| <b>10</b> | <b>The kernel of Babel (babel.def for <math>\LaTeX</math>only)</b> | <b>110</b> |
| 10.1      | The redefinition of the style commands . . . . .                   | 110        |
| 10.2      | Cross referencing macros . . . . .                                 | 111        |
| 10.3      | Marks . . . . .  | 114        |
| 10.4      | Preventing clashes with other packages . . . . .                   | 115        |
| 10.4.1    | ifthen . . . . .   | 115        |
| 10.4.2    | varioref . . . . .   | 116        |
| 10.4.3    | hhline . . . . .   | 116        |
| 10.4.4    | hyperref . . . . .   | 117        |
| 10.4.5    | fancyhdr . . . . .   | 117        |
| 10.5      | Encoding and fonts . . . . .                                       | 118        |
| 10.6      | Basic bidi support . . . . .                                       | 119        |
| 10.7      | Local Language Configuration . . . . .                             | 122        |
| <b>11</b> | <b>Multiple languages (switch.def)</b>                             | <b>123</b> |
| 11.1      | Selecting the language . . . . .                                   | 124        |
| 11.2      | Errors . . . . .   | 133        |
| <b>12</b> | <b>Loading hyphenation patterns</b>                                | <b>135</b> |
| <b>13</b> | <b>Font handling with fontspec</b>                                 | <b>139</b> |

|           |   |            |
|-----------|---|------------|
| <b>14</b> | <b>Hooks for XeTeX and LuaTeX</b>                                 | <b>144</b> |
| 14.1      | XeTeX . . . . .   | 144        |
| 14.2      | Layout . . . . .  | 146        |
| 14.3      | LuaTeX . . . . .  | 148        |
| 14.4      | Southeast Asian scripts . . . . .                                 | 153        |
| 14.5      | CJK line breaking . . . . .                                       | 157        |
| 14.6      | Auto bidi with basic and basic-r . . . . .                        | 164        |
| <b>15</b> | <b>Data for CJK</b>   | <b>175</b> |
| <b>16</b> | <b>The ‘nil’ language</b>   | <b>175</b> |
| <b>17</b> | <b>Support for Plain T<sub>E</sub>X (plain.def)</b>               | <b>176</b> |
| 17.1      | Not renaming hyphen.tex . . . . .                                 | 176        |
| 17.2      | Emulating some L <sup>A</sup> T <sub>E</sub> X features . . . . . | 177        |
| 17.3      | General tools . . . . .   | 177        |
| 17.4      | Encoding related macros . . . . .                                 | 181        |
| <b>18</b> | <b>Acknowledgements</b>   | <b>184</b> |

## Troubleshooting

|  |    |
|--|----|
| Paragraph ended before \UTFviii@three@octets was complete . . . . .  | 4  |
| No hyphenation patterns were preloaded for (babel) the language ‘LANG’ into the format . . . . .   | 5  |
| You are loading directly a language style . . . . .  | 8  |
| Unknown language ‘LANG’ . . . . .  | 8  |
| Argument of \language@active@arg” has an extra } . . . . .   | 12 |
| Package fontspec Warning: ‘Language ‘LANG’ not available for font ‘FONT’ with script ‘SCRIPT’ ‘Default’ language used instead’ . . . . . | 26 |
| Package babel Warning: The following fonts are not babel standard families . .   | 26 |

## Part I

# User guide

- This user guide focuses on  $\LaTeX$ . There are also some notes on its use with Plain  $\TeX$ .
- Changes and new features with relation to version 3.8 are highlighted with **New X.XX**. The most recent features could be still unstable. Please, report any issues you find in <https://github.com/latex3/babel/issues>, which is better than just complaining on an e-mail list or a web forum.
- If you are interested in the  $\TeX$  multilingual support, please join the kadingira list on <http://tug.org/mailman/listinfo/kadingira>. You can follow the development of babel in <https://github.com/latex3/babel> (which provides some sample files, too).
- See section 3.1 for contributing a language.
- The first sections describe the traditional way of loading a language (with `ldf` files). The alternative way based on `ini` files, which complements the previous one (it will *not* replace it), is described below.

## 1 The user interface

### 1.1 Monolingual documents

In most cases, a single language is required, and then all you need in  $\LaTeX$  is to load the package using its standard mechanism for this purpose, namely, passing that language as an optional argument. In addition, you may want to set the font and input encodings.

**EXAMPLE** Here is a simple full example for “traditional”  $\TeX$  engines (see below for `xetex` and `luatex`). The packages `fontenc` and `inputenc` do not belong to babel, but they are included in the example because typically you will need them (however, the package `inputenc` may be omitted with  $\LaTeX \geq 2018-04-01$  if the encoding is UTF-8):

```
\documentclass{article}

\usepackage[T1]{fontenc}
%\usepackage[utf8]{inputenc}%\Uncomment_if_LaTeX<2018-04-01

\usepackage[french]{babel}

\begin{document}

Plus ça change, plus c'est la même chose!

\end{document}
```

**TROUBLESHOOTING** A common source of trouble is a wrong setting of the input encoding. Very often you will get the following somewhat cryptic error:

```
!_Paragraph_ended_before_\\UTFviii@three@octets_was_complete.
```

Make sure you set the encoding actually used by your editor.

Another approach is making the language (french in the example) a global option in order to let other packages detect and use it:

```
\documentclass[french]{article}
\usepackage{babel}
\usepackage{varioref}
```

In this last example, the package `varioref` will also see the option and will be able to use it.

**NOTE** Because of the way `babel` has evolved, “language” can refer to (1) a set of hyphenation patterns as preloaded into the format, (2) a package option, (3) an `ldf` file, and (4) a name used in the document to select a language or dialect. So, a package option refers to a language in a generic way – sometimes it is the actual language name used to select it, sometimes it is a file name loading a language with a different name, sometimes it is a file name loading several languages. Please, read the documentation for specific languages for further info.

**TROUBLESHOOTING** The following warning is about hyphenation patterns, which are not under the direct control of `babel`:

```
Package babel Warning: No hyphenation patterns were preloaded for
(babel) the language 'LANG' into the format.
(babel) Please, configure your TeX system to add them and
(babel) rebuild the format. Now I will use the patterns
(babel) preloaded for \language=0 instead on input line 57.
```

The document will be typeset, but very likely the text will not be correctly hyphenated. Some languages may be raising this warning wrongly (because they are not hyphenated); it is a bug to be fixed – just ignore it. See the manual of your distribution (MacTeX, MikTeX, T<sub>E</sub>XLive, etc.) for further info about how to configure it.

## 1.2 Multilingual documents

In multilingual documents, just use a list of the required languages as package or class options. The last language is considered the main one, activated by default. Sometimes, the main language changes the document layout (eg, spanish and french).

**EXAMPLE** In  $\text{\LaTeX}$ , the preamble of the document:

```
\documentclass{article}
\usepackage[dutch,english]{babel}
```

would tell  $\text{\LaTeX}$  that the document would be written in two languages, Dutch and English, and that English would be the first language in use, and the main one.

You can also set the main language explicitly, but it is discouraged except if there a real reason to do so:

```
\documentclass{article}
\usepackage[main=english,dutch]{babel}
```

Examples of cases where `main` is useful are the following.

**NOTE** Some classes load babel with a hardcoded language option. Sometimes, the main language could be overridden with something like that before `\documentclass`:

```
\PassOptionsToPackage{main=english}{babel}
```

**WARNING** Languages may be set as global and as package option at the same time, but in such a case you should set explicitly the main language with the package option `main`:

```
\documentclass[italian]{book}  
\usepackage[ngerman,main=italian]{babel}
```

**WARNING** In the preamble the main language has *not* been selected, except hyphenation patterns and the name assigned to `\language` (in particular, shorthands, captions and date are not activated). If you need to define boxes and the like in the preamble, you might want to use some of the language selectors described below.

To switch the language there are two basic macros, described below in detail: `\selectlanguage` is used for blocks of text, while `\foreignlanguage` is for chunks of text inside paragraphs.

**EXAMPLE** A full bilingual document follows. The main language is french, which is activated when the document begins. The package `inputenc` may be omitted with  $\text{\LaTeX} \geq 2018-04-01$  if the encoding is UTF-8.

```
\documentclass{article}  
  
\usepackage[T1]{fontenc}  
\usepackage[utf8]{inputenc}  
  
\usepackage[english,french]{babel}  
  
\begin{document}  
  
Plus ça change, plus c'est la même chose!  
  
\selectlanguage{english}  
  
And an English paragraph, with a short text in  
\foreignlanguage{french}{français}.  
  
\end{document}
```

### 1.3 Modifiers

**New 3.9c** The basic behavior of some languages can be modified when loading babel by means of *modifiers*. They are set after the language name, and are prefixed with a dot (only when the language is set as package option – neither global options nor the `main` key accepts them). An example is (spaces are not significant and they can be added or removed):<sup>1</sup>

---

<sup>1</sup>No predefined “axis” for modifiers are provided because languages and their scripts have quite different needs.

```
\usepackage[latin.medieval,spanish.notilde.lcroman,danish]{babel}
```

Attributes (described below) are considered modifiers, ie, you can set an attribute by including it in the list of modifiers. However, modifiers are a more general mechanism.

## 1.4 xelatex and lualatex

Many languages are compatible with xetex and luatex. With them you can use babel to localize the documents.

The Latin script is covered by default in current  $\text{\LaTeX}$  (provided the document encoding is UTF-8), because the font loader is preloaded and the font is switched to `lmroman`. Other scripts require loading `fontspec`. You may want to set the font attributes with `fontspec`, too.

**EXAMPLE** The following bilingual, single script document in UTF-8 encoding just prints a couple of ‘captions’ and `\today` in Danish and Vietnamese. No additional packages are required.

```
\documentclass{article}

\usepackage[vietnamese,danish]{babel}

\begin{document}

\prefacename{}--\alsoname{}--\today

\selectlanguage{vietnamese}

\prefacename{}--\alsoname{}--\today

\end{document}
```

**EXAMPLE** Here is a simple monolingual document in Russian (text from the Wikipedia). Note neither `fontenc` nor `inputenc` are necessary, but the document should be encoded in UTF-8 and a so-called Unicode font must be loaded (in this example `\babelfont` is used, described below).

```
\documentclass{article}

\usepackage[russian]{babel}

\babelfont{rm}{DejaVuSerif}

\begin{document}

Россия, находящаяся на пересечении множества культур, а также
с учётом многонационального характера её населения, отличается
высокой степенью этнокультурного многообразия и способностью к
межкультурному диалогу.

\end{document}
```



## 1.5 Troubleshooting

- Loading directly sty files in L<sup>A</sup>T<sub>E</sub>X (ie, `\usepackage{<language>}`) is deprecated and you will get the error:<sup>2</sup>

```
!Package babel Error: You are loading directly a language style.
(babel)~~~~~This syntax is deprecated and you must use
(babel)~~~~~\usepackage[language]{babel}.
```

- Another typical error when using babel is the following:<sup>3</sup>

```
!Package babel Error: Unknown language `#1'. Either you have
(babel)~~~~~misspelled its name, it has not been installed,
(babel)~~~~~or you requested it in a previous run. Fix its name,
(babel)~~~~~install it or just rerun the file, respectively. In
(babel)~~~~~some cases, you may need to remove the aux file
```

The most frequent reason is, by far, the latest (for example, you included spanish, but you realized this language is not used after all, and therefore you removed it from the option list). In most cases, the error vanishes when the document is typeset again, but in more severe ones you will need to remove the aux file.

## 1.6 Plain

In Plain, load languages styles with `\input` and then use `\begindocument` (the latter is defined by babel):

```
\input_estonian.sty
\begindocument
```

**WARNING** Not all languages provide a sty file and some of them are not compatible with Plain.<sup>4</sup>

## 1.7 Basic language selectors

This section describes the commands to be used in the document to switch the language in multilingual documents. In most cases, only the two basic macros `\selectlanguage` and `\foreignlanguage` are necessary. The environments `otherlanguage`, `otherlanguage*` and `hyphenrules` are auxiliary, and described in the next section.

The main language is selected automatically when the document environment begins.

`\selectlanguage` {<language>}

When a user wants to switch from one language to another he can do so using the macro `\selectlanguage`. This macro takes the language, defined previously by a language definition file, as its argument. It calls several macros that should be defined in the language definition files to activate the special definitions for the language chosen:

<sup>2</sup>In old versions the error read “You have used an old interface to call babel”, not very helpful.

<sup>3</sup>In old versions the error read “You haven’t loaded the language LANG yet”.

<sup>4</sup>Even in the babel kernel there were some macros not compatible with plain. Hopefully these issues have been fixed.

```
\selectlanguage{german}
```

This command can be used as environment, too.

**NOTE** For “historical reasons”, a macro name is converted to a language name without the leading \; in other words, `\selectlanguage{\german}` is equivalent to `\selectlanguage{german}`. Using a macro instead of a “real” name is deprecated.

**WARNING** If used inside braces there might be some non-local changes, as this would be roughly equivalent to:

```
{\selectlanguage{<inner-language>}\...}\selectlanguage{<outer-language>}
```

If you want a change which is really local, you must enclose this code with an additional grouping level.

**\foreignlanguage** `{\language}{\text}`

The command `\foreignlanguage` takes two arguments; the second argument is a phrase to be typeset according to the rules of the language named in its first one. This command (1) only switches the extra definitions and the hyphenation rules for the language, *not* the names and dates, (2) does not send information about the language to auxiliary files (i.e., the surrounding language is still in force), and (3) it works even if the language has not been set as package option (but in such a case it only sets the hyphenation patterns and a warning is shown). With the `bidi` option, it also enters in horizontal mode (this is not done always for backwards compatibility).

## 1.8 Auxiliary language selectors

**\begin{otherlanguage}** `{\language} ... \end{otherlanguage}`

The environment `otherlanguage` does basically the same as `\selectlanguage`, except that language change is (mostly) local to the environment.

Actually, there might be some non-local changes, as this environment is roughly equivalent to:

```
\begingroup
\selectlanguage{<inner-language>}
...
\endgroup
\selectlanguage{<outer-language>}
```

If you want a change which is really local, you must enclose this environment with an additional grouping, like braces {}.

Spaces after the environment are ignored.

**\begin{otherlanguage\*}** `{\language} ... \end{otherlanguage*}`

Same as `\foreignlanguage` but as environment. Spaces after the environment are *not* ignored.

This environment was originally intended for intermixing left-to-right typesetting with right-to-left typesetting in engines not supporting a change in the writing direction inside a

line. However, by default it never complied with the documented behavior and it is just a version as environment of `\foreignlanguage`, except when the option `bidi` is set – in this case, `\foreignlanguage` emits a `\leavevmode`, while `otherlanguage*` does not.

`\begin{hyphenrules}`  $\langle language \rangle$  ... `\end{hyphenrules}`

The environment `hyphenrules` can be used to select *only* the hyphenation rules to be used (it can be used as command, too). This can for instance be used to select ‘nohyphenation’, provided that in `language.dat` the ‘`language`’ `nohyphenation` is defined by loading `zerohyph.tex`. It deactivates language shorthands, too (but not user shorthands). Except for these simple uses, `hyphenrules` is discouraged and `otherlanguage*` (the starred version) is preferred, as the former does not take into account possible changes in encodings of characters like, say, ‘`’`’ done by some languages (eg, `italian`, `french`, `ukraineb`). To set hyphenation exceptions, use `\babelhyphenation` (see below).

## 1.9 More on selection

`\babeltags`  $\langle tag1 \rangle = \langle language1 \rangle, \langle tag2 \rangle = \langle language2 \rangle, \dots$

**New 3.9i** In multilingual documents with many language-switches the commands above can be cumbersome. With this tool shorter names can be defined. It adds nothing really new – it is just syntactical sugar.

It defines `\text<tag1>\{<text>\}` to be `\foreignlanguage{\langle language1 \rangle}\{<text>\}`, and `\begin{\langle tag1 \rangle}` to be `\begin{otherlanguage*}\{\langle language1 \rangle\}`, and so on. Note `\langle tag1 \rangle` is also allowed, but remember to set it locally inside a group.

**EXAMPLE** With

```
\babeltags{de=german}
```

you can write

```
text\textde{German}text
```

and

```
text
\begin{de}
  German
\end{de}
text
```

**NOTE** Something like `\babeltags{finnish = finnish}` is legitimate – it defines `\textfinnish` and `\finnish` (and, of course, `\begin{finnish}`).

**NOTE** Actually, there may be another advantage in the ‘short’ syntax `\text<tag>`, namely, it is not affected by `\MakeUppercase` (while `\foreignlanguage` is).

`\babelensure` [`include=<commands>`],`exclude=<commands>`],`fontenc=<encoding>`]{`<language>`}

**New 3.9i** Except in a few languages, like russian, captions and dates are just strings, and do not switch the language. That means you should set it explicitly if you want to use them, or hyphenation (and in some cases the text itself) will be wrong. For example:

```
\foreignlanguage{russian}{text}\foreignlanguage{polish}{\seename}\text}
```

Of course,  $\TeX$  can do it for you. To avoid switching the language all the while, `\babelensure` redefines the captions for a given language to wrap them with a selector:

```
\babelensure{polish}
```

By default only the basic captions and `\today` are redefined, but you can add further macros with the key `include` in the optional argument (without commas). Macros not to be modified are listed in `exclude`. You can also enforce a font encoding with `fontenc`.<sup>5</sup> A couple of examples:

```
\babelensure[include=\Today]{spanish}
\babelensure[fontenc=T5]{vietnamese}
```

They are activated when the language is selected (at the `afterextras` event), and it makes some assumptions which could not be fulfilled in some languages. Note also you should include only macros defined by the language, not global macros (eg,  $\TeX$  or `\dag`). With ini files (see below), captions are ensured by default.

## 1.10 Shorthands

A *shorthand* is a sequence of one or two characters that expands to arbitrary  $\TeX$  code. Shorthands can be used for different kinds of things, for example: (1) in some languages shorthands such as "a are defined to be able to hyphenate the word if the encoding is OT1; (2) in some languages shorthands such as ! are used to insert the right amount of white space; (3) several kinds of discretionary and breaks can be inserted easily with "-", "=", etc. The package `inputenc` as well as `xetex` and `luatex` have alleviated entering non-ASCII characters, but minority languages and some kinds of text can still require characters not directly available on the keyboards (and sometimes not even as separated or precomposed Unicode characters). As to the point 2, now `pdfTeX` provides `\knbcode`, and `luatex` can manipulate the glyph list. Tools for point 3 can be still very useful in general. There are three levels of shorthands: *user*, *language*, and *system* (by order of precedence). Version 3.9 introduces the *language user* level on top of the user level, as described below. In most cases, you will use only shorthands provided by languages.

**NOTE** Note the following:

1. Activated chars used for two-char shorthands cannot be followed by a closing brace } and the spaces following are gobbled. With one-char shorthands (eg, :), they are preserved.
2. If on a certain level (system, language, user) there is a one-char shorthand, two-char ones starting with that char and on the same level are ignored.
3. Since they are active, a shorthand cannot contain the same character in its definition (except if it is deactivated with, eg, `string`).

<sup>5</sup>With it encoded string may not work as expected.

A typical error when using shorthands is the following:

```
!_Argument_of_\language@active@arg" _has_an_extra_.
```

It means there is a closing brace just after a shorthand, which is not allowed (eg, "}"). Just add {} after (eg, "{}}").

**\shorthandon**     $\langle shorthand-list \rangle$   
**\shorthandoff**    $\ast \langle shorthand-list \rangle$

It is sometimes necessary to switch a shorthand character off temporarily, because it must be used in an entirely different way. For this purpose, the user commands `\shorthandoff` and `\shorthandon` are provided. They each take a list of characters as their arguments. The command `\shorthandoff` sets the `\catcode` for each of the characters in its argument to other (12); the command `\shorthandon` sets the `\catcode` to active (13). Both commands only work on ‘known’ shorthand characters.

**New 3.9a** However, `\shorthandoff` does not behave as you would expect with characters like `~` or `^`, because they usually are not “other”. For them `\shorthandoff*` is provided, so that with

```
\shorthandoff*{~^}
```

`~` is still active, very likely with the meaning of a non-breaking space, and `^` is the superscript character. The catcodes used are those when the shorthands are defined, usually when language files are loaded.

**\usesshorthands**    $\ast \langle char \rangle$

The command `\usesshorthands` initiates the definition of user-defined shorthand sequences. It has one argument, the character that starts these personal shorthands.

**New 3.9a** User shorthands are not always alive, as they may be deactivated by languages (for example, if you use `"` for your user shorthands and switch from german to french, they stop working). Therefore, a starred version `\usesshorthands*\langle char \rangle` is provided, which makes sure shorthands are always activated.

Currently, if the package option `shorthands` is used, you must include any character to be activated with `\usesshorthands`. This restriction will be lifted in a future release.

**\defineshorthand**    $[ \langle language \rangle , \langle language \rangle , \dots ] \langle shorthand \rangle \langle code \rangle$

The command `\defineshorthand` takes two arguments: the first is a one- or two-character shorthand sequence, and the second is the code the shorthand should expand to.

**New 3.9a** An optional argument allows to (re)define language and system shorthands (some languages do not activate shorthands, so you may want to add `\languageshorthands\langle lang \rangle` to the corresponding `\extras\langle lang \rangle`, as explained below). By default, user shorthands are (re)defined.

User shorthands override language ones, which in turn override system shorthands. Language-dependent user shorthands (new in 3.9) take precedence over “normal” user shorthands.

**EXAMPLE** Let’s assume you want a unified set of shorthand for discretionaries (languages do not define shorthands consistently, and `"`, `\`, `=` have different meanings). You could start with, say:

```
\usesshorthands*{"}
\defineshorthand{"*"}{\babelhyphen{soft}}
\defineshorthand{"-"}{\babelhyphen{hard}}
```

However, the behavior of hyphens is language-dependent. For example, in languages like Polish and Portuguese, a hard hyphen inside compound words are repeated at the beginning of the next line. You could then set:

```
\defineshorthand[*polish,*portugese]{"-"}{\babelhyphen{repeat}}
```

Here, options with `*` set a language-dependent user shorthand, which means the generic one above only applies for the rest of languages; without `*` they would (re)define the language shorthands instead, which are overridden by user ones.

Now, you have a single unified shorthand ("`-`"), with a content-based meaning ('compound word hyphen') whose visual behavior is that expected in each context.

**\aliasshorthand** `{⟨original⟩}{⟨alias⟩}`

The command `\aliasshorthand` can be used to let another character perform the same functions as the default shorthand character. If one prefers for example to use the character `/` over `"` in typing Polish texts, this can be achieved by entering `\aliasshorthand{"}{/}`.

**NOTE** The substitute character must *not* have been declared before as shorthand (in such a case, `\aliasshorthands` is ignored).

**EXAMPLE** The following example shows how to replace a shorthand by another

```
\aliasshorthand{~}{^}
\AtBeginDocument{\shorthandoff*{~}}
```

**WARNING** Shorthands remember somehow the original character, and the fallback value is that of the latter. So, in this example, if no shorthand is found, `^` expands to a non-breaking space, because this is the value of `~` (internally, `^` still calls `\active@char~` or `\normal@char~`). Furthermore, if you change the system value of `^` with `\defineshorthand` nothing happens.

**\languageshorthands** `{⟨language⟩}`

The command `\languageshorthands` can be used to switch the shorthands on the language level. It takes one argument, the name of a language or none (the latter does what its name suggests).<sup>6</sup> Note that for this to work the language should have been specified as an option when loading the babel package. For example, you can use in english the shorthands defined by ngerman with

```
\addto\extrasenglish{\languageshorthands{ngerman}}
```

<sup>6</sup>Actually, any name not corresponding to a language group does the same as none. However, follow this convention because it might be enforced in future releases of babel to catch possible errors.

(You may also need to activate them as user shorthands in the preamble with, for example, `\usesshorthands` or `\usesshorthands*`.)

Very often, this is a more convenient way to deactivate shorthands than `\shorthandoff`, for example if you want to define a macro to easy typing phonetic characters with tipa:

```
\newcommand{\myipa}[1]{\{\language shorthands{none}\tipaencoding#1}}
```

`\babelshorthand`  $\{\langle shorthand \rangle\}$

With this command you can use a shorthand even if (1) not activated in shorthands (in this case only shorthands for the current language are taken into account, ie, not user shorthands), (2) turned off with `\shorthandoff` or (3) deactivated with the internal `\bbl@deactivate`; for example, `\babelshorthand{"u}` or `\babelshorthand{:}`. (You can conveniently define your own macros, or even your own user shorthands provided they do not overlap.)

For your records, here is a list of shorthands, but you must double check them, as they may change:<sup>7</sup>

**Languages with no shorthands** Croatian, English (any variety), Indonesian, Hebrew, Interlingua, Irish, Lower Sorbian, Malaysian, North Sami, Romanian, Scottish, Welsh  
**Languages with only " as defined shorthand character** Albanian, Bulgarian, Danish, Dutch, Finnish, German (old and new orthography, also Austrian), Icelandic, Italian, Norwegian, Polish, Portuguese (also Brazilian), Russian, Serbian (with Latin script), Slovene, Swedish, Ukrainian, Upper Sorbian

**Basque** " ' ~  
**Breton** : ; ? !  
**Catalan** " ' `   
**Czech** " -  
**Esperanto** ^  
**Estonian** " ~  
**French** (all varieties) : ; ? !  
**Galician** " . ' ~ < >  
**Greek** ~  
**Hungarian** `  
**Kurmanji** ^  
**Latin** " ^ =  
**Slovak** " ^ ' -  
**Spanish** " . < > '  
**Turkish** : ! =

In addition, the babel core declares ~ as a one-char shorthand which is let, like the standard ~, to a non breaking space.<sup>8</sup>

`\ifbabelshorthand`  $\{\langle character \rangle\}\{\langle true \rangle\}\{\langle false \rangle\}$

**New 3.23** Tests if a character has been made a shorthand.

## 1.11 Package options

**New 3.9a** These package options are processed before language options, so that they are taken into account irrespective of its order. The first three options have been available in previous versions.

<sup>7</sup>Thanks to Enrico Gregorio

<sup>8</sup>This declaration serves to nothing, but it is preserved for backward compatibility.

|                             |   |
|-----------------------------|---|
| <b>KeepShorthandsActive</b> | Tells babel not to deactivate shorthands after loading a language file, so that they are also available in the preamble.  |
| <b>activeacute</b>          | For some languages babel supports this options to set ' as a shorthand in case it is not done by default.   |
| <b>activegrave</b>          | Same for `.   |
| <b>shorthands=</b>          | <p><math>\langle char \rangle \langle char \rangle \dots</math>   off</p> <p>The only language shorthands activated are those given, like, eg:</p> <pre style="background-color: #f0f0f0; padding: 10px;">\usepackage[esperanto,french,shorthands=:;!]{babel}</pre> <p>If ' is included, activeacute is set; if ` is included, activegrave is set. Active characters (like ~) should be preceded by \string (otherwise they will be expanded by <math>\LaTeX</math> before they are passed to the package and therefore they will not be recognized); however, t is provided for the common case of ~ (as well as c for not so common case of the comma). With shorthands=off no language shorthands are defined, As some languages use this mechanism for tools not available otherwise, a macro \babelshorthand is defined, which allows using them; see above.</p> |
| <b>safe=</b>                | <p>none   ref   bib</p> <p>Some <math>\LaTeX</math> macros are redefined so that using shorthands is safe. With safe=bib only \nocite, \bibtex and \bibitem are redefined. With safe=ref only \newlabel, \ref and \pageref are redefined (as well as a few macros from varioref and ifthen). With safe=none no macro is redefined. This option is strongly recommended, because a good deal of incompatibilities and errors are related to these redefinitions. As of <b>New 3.34</b>, in <math>\epsilon\TeX</math> based engines (ie, almost every engine except the oldest ones) shorthands can be used in these macros (formerly you could not).</p>   |
| <b>math=</b>                | <p>active   normal</p> <p>Shorthands are mainly intended for text, not for math. By setting this option with the value normal they are deactivated in math mode (default is active) and things like <math>\{a'\}</math> (a closing brace after a shorthand) are not a source of trouble anymore.</p>  |
| <b>config=</b>              | <p><math>\langle file \rangle</math></p> <p>Load <math>\langle file \rangle.cfg</math> instead of the default config file bblopts.cfg (the file is loaded even with noconfigs).</p>   |
| <b>main=</b>                | <p><math>\langle language \rangle</math></p> <p>Sets the main language, as explained above, ie, this language is always loaded last. If it is not given as package or global option, it is added to the list of requested languages.</p>  |
| <b>headfoot=</b>            | <p><math>\langle language \rangle</math></p> <p>By default, headlines and footlines are not touched (only marks), and if they contain language-dependent macros (which is not usual) there may be unexpected results. With this option you may set the language in heads and foots.</p>   |



- noconfigs** Global and language default config files are not loaded, so you can make sure your document is not spoiled by an unexpected .cfg file. However, if the key config is set, this file is loaded.
- showlanguages** Prints to the log the list of languages loaded when the format was created: number (remember dialects can share it), name, hyphenation file and exceptions file.
- nocase** **New 3.9l** Language settings for uppercase and lowercase mapping (as set by \SetCase) are ignored. Use only if there are incompatibilities with other packages.
- silent** **New 3.9l** No warnings and no *infos* are written to the log file.<sup>9</sup>
- strings=** generic | unicode | encoded | *<label>* | *<font encoding>*  
 Selects the encoding of strings in languages supporting this feature. Predefined labels are generic (for traditional T<sub>E</sub>X, LICR and ASCII strings), unicode (for engines like xetex and luatex) and encoded (for special cases requiring mixed encodings). Other allowed values are font encoding codes (T1, T2A, LGR, L7X...), but only in languages supporting them. Be aware with encoded captions are protected, but they work in \MakeUppercase and the like (this feature misuses some internal L<sup>A</sup>T<sub>E</sub>X tools, so use it only as a last resort).
- hyphenmap=** off | main | select | other | other\*  
**New 3.9g** Sets the behavior of case mapping for hyphenation, provided the language defines it.<sup>10</sup> It can take the following values:  
**off** deactivates this feature and no case mapping is applied;  
**first** sets it at the first switching commands in the current or parent scope (typically, when the aux file is first read and at \begin{document}, but also the first \selectlanguage in the preamble), and it's the default if a single language option has been stated;<sup>11</sup>  
**select** sets it only at \selectlanguage;  
**other** also sets it at otherlanguage;  
**other\*** also sets it at otherlanguage\* as well as in heads and foots (if the option headfoot is used) and in auxiliary files (ie, at \select@language), and it's the default if several language options have been stated. The option first can be regarded as an optimized version of other\* for monolingual documents.<sup>12</sup>
- bidi=** default | basic | basic-r | bidi-l | bidi-r  
**New 3.14** Selects the bidi algorithm to be used in luatex and xetex. See sec. 1.21.
- layout=** **New 3.16** Selects which layout elements are adapted in bidi documents. See sec. 1.21.

## 1.12 The base option

With this package option babel just loads some basic macros (those in switch.def), defines \AfterBabelLanguage and exits. It also selects the hyphenation patterns for the

<sup>9</sup>You can use alternatively the package silence.

<sup>10</sup>Turned off in plain.

<sup>11</sup>Duplicated options count as several ones.

<sup>12</sup>Providing foreign is pointless, because the case mapping applied is that at the end of the paragraph, but if either xetex or luatex change this behavior it might be added. On the other hand, other is provided even if I [JBL] think it isn't really useful, but who knows.

last language passed as option (by its name in `language.dat`). There are two main uses: classes and packages, and as a last resort in case there are, for some reason, incompatible languages. It can be used if you just want to select the hyphenation patterns of a single language, too.

`\AfterBabelLanguage`  $\{\langle option-name \rangle\}\{\langle code \rangle\}$

This command is currently the only provided by `base`. Executes  $\langle code \rangle$  when the file loaded by the corresponding package option is finished (at `\lbf@finish`). The setting is global. So

```
\AfterBabelLanguage{french}\{...}
```

does ... at the end of `french.ldf`. It can be used in `ldf` files, too, but in such a case the code is executed only if  $\langle option-name \rangle$  is the same as `\CurrentOption` (which could not be the same as the option name as set in `\usepackage!`).

**EXAMPLE** Consider two languages `foo` and `bar` defining the same `\macro` with `\newcommand`. An error is raised if you attempt to load both. Here is a way to overcome this problem:

```
\usepackage[base]{babel}
\AfterBabelLanguage{foo}{%
  \let\macroFoo\macro
  \let\macro\relax}
\usepackage[foo,bar]{babel}
```

## 1.13 ini files

An alternative approach to define a language is by means of an `ini` file. Currently `babel` provides about 200 of these files containing the basic data required for a language. Most of them set the date, and many also the captions (Unicode and LICR). They will be evolving with the time to add more features (something to keep in mind if backward compatibility is important). The following section shows how to make use of them currently (by means of `\babelprovide`), but a higher interface, based on package options, is under development (in other words, `\babelprovide` is mainly intended for auxiliary tasks).

**EXAMPLE** Although Georgian has its own `ldf` file, here is how to declare this language with an `ini` file in Unicode engines.

```
\documentclass{book}

\usepackage{babel}
\babelprovide[import,main]{georgian}

\babelfont{rm}{DejaVuSans}

\begin{document}

\tableofcontents

\chapter{სამზარეულოდასუფრისტრადიციები}

ქართულიტრადიციულისამზარეულოერთ-ერთიუმდიდრესიამთვედმსოფლიოში.
```

```
\end{document}
```

**NOTE** The ini files just define and set some parameters, but the corresponding behavior is not always implemented. Also, there are some limitations in the engines. A few remarks follows:

**Arabic** Monolingual documents mostly work in luatex, but it must be fine tuned, and a recent version of fontspec/loaotfload is required. In xetex babel resorts to the bidi package, which seems to work.

**Hebrew** Niqqud marks seem to work in both engines, but cantillation marks are misplaced (xetex seems better, but still problematic).

**Devanagari** In luatex many fonts work, but some others do not, the main issue being the ‘ra’. It is advisable to set explicitly the script to either deva or dev2, eg:

```
\newfontscript{Devanagari}{deva}
```

Other Indic scripts are still under development in luatex. On the other hand, xetex is better.

**Southeast scripts** Thai works in both luatex and xetex, but line breaking differs (rules can be modified in luatex; they are hardcoded in xetex). Lao seems to work, too, but there are no patterns for the latter in luatex. Some quick patterns could help, with something similar to:

```
\babelprovide[import,hyphenrules=+]{lao}
\babelpatterns[lao]{ໂດ, ມຸງ, ເລີ່ມ, ແຈ້ງ, ໄປ, ຈັບ, ປະສານ}
```

Khemer clusters are rendered wrongly.

**East Asia scripts** Settings for either Simplified or Traditional should work out of the box. `luatex` does basic line breaking, but currently `xetex` does not (you may load `zhspacing`). Although for a few words and short texts the `ini` files should be fine, CJK texts are best set with a dedicated framework (CJK, `luatexja`, `kotex`, `CtEx`, etc.). This is what the class `ltjbook` does with `luatex`, which can be used in conjunction with the `ldf` for `japanese`, because the following piece of code loads `luatexja`:

```
\documentclass{ltjbook}
\usepackage[japanese]{babel}
```

Here is the list (u means Unicode captions, and l means LICR captions):

|       |                         |         |                           |
|-------|-------------------------|---------|---------------------------|
| af    | Afrikaans <sup>ul</sup> | asa     | Asu                       |
| agq   | Aghem                   | ast     | Asturian <sup>ul</sup>    |
| ak    | Akan                    | az-Cyrl | Azerbaijani               |
| am    | Amharic <sup>ul</sup>   | az-Latn | Azerbaijani               |
| ar    | Arabic <sup>ul</sup>    | az      | Azerbaijani <sup>ul</sup> |
| ar-DZ | Arabic <sup>ul</sup>    | bas     | Basaa                     |
| ar-MA | Arabic <sup>ul</sup>    | be      | Belarusian <sup>ul</sup>  |
| ar-SY | Arabic <sup>ul</sup>    | bem     | Bemba                     |
| as    | Assamese                | bez     | Bena                      |

|         |                               |       |                             |
|---------|-------------------------------|-------|-----------------------------|
| bg      | Bulgarian <sup>ul</sup>       | gl    | Galician <sup>ul</sup>      |
| bm      | Bambara                       | gsw   | Swiss German                |
| bn      | Bangla <sup>ul</sup>          | gu    | Gujarati                    |
| bo      | Tibetan <sup>u</sup>          | guz   | Gusii                       |
| brx     | Bodo                          | gv    | Manx                        |
| bs-Cyrl | Bosnian                       | ha-GH | Hausa                       |
| bs-Latn | Bosnian <sup>ul</sup>         | ha-NE | Hausa <sup>l</sup>          |
| bs      | Bosnian <sup>ul</sup>         | ha    | Hausa                       |
| ca      | Catalan <sup>ul</sup>         | haw   | Hawaiian                    |
| ce      | Chechen                       | he    | Hebrew <sup>ul</sup>        |
| cgg     | Chiga                         | hi    | Hindi <sup>u</sup>          |
| chr     | Cherokee                      | hr    | Croatian <sup>ul</sup>      |
| ckb     | Central Kurdish               | hsb   | Upper Sorbian <sup>ul</sup> |
| cs      | Czech <sup>ul</sup>           | hu    | Hungarian <sup>ul</sup>     |
| cy      | Welsh <sup>ul</sup>           | hy    | Armenian                    |
| da      | Danish <sup>ul</sup>          | ia    | Interlingua <sup>ul</sup>   |
| dav     | Taita                         | id    | Indonesian <sup>ul</sup>    |
| de-AT   | German <sup>ul</sup>          | ig    | Igbo                        |
| de-CH   | German <sup>ul</sup>          | ii    | Sichuan Yi                  |
| de      | German <sup>ul</sup>          | is    | Icelandic <sup>ul</sup>     |
| dje     | Zarma                         | it    | Italian <sup>ul</sup>       |
| dsb     | Lower Sorbian <sup>ul</sup>   | ja    | Japanese                    |
| dua     | Duala                         | jgo   | Ngomba                      |
| dyo     | Jola-Fonyi                    | jmc   | Machame                     |
| dz      | Dzongkha                      | ka    | Georgian <sup>ul</sup>      |
| ebu     | Embu                          | kab   | Kabyle                      |
| ee      | Ewe                           | kam   | Kamba                       |
| el      | Greek <sup>ul</sup>           | kde   | Makonde                     |
| en-AU   | English <sup>ul</sup>         | kea   | Kabuverdianu                |
| en-CA   | English <sup>ul</sup>         | khq   | Koyra Chiini                |
| en-GB   | English <sup>ul</sup>         | ki    | Kikuyu                      |
| en-NZ   | English <sup>ul</sup>         | kk    | Kazakh                      |
| en-US   | English <sup>ul</sup>         | kkj   | Kako                        |
| en      | English <sup>ul</sup>         | kl    | Kalaallisut                 |
| eo      | Esperanto <sup>ul</sup>       | kln   | Kalenjin                    |
| es-MX   | Spanish <sup>ul</sup>         | km    | Khmer                       |
| es      | Spanish <sup>ul</sup>         | kn    | Kannada <sup>ul</sup>       |
| et      | Estonian <sup>ul</sup>        | ko    | Korean                      |
| eu      | Basque <sup>ul</sup>          | kok   | Konkani                     |
| ewo     | Ewondo                        | ks    | Kashmiri                    |
| fa      | Persian <sup>ul</sup>         | ksb   | Shambala                    |
| ff      | Fulah                         | ksf   | Bafia                       |
| fi      | Finnish <sup>ul</sup>         | ksh   | Colognian                   |
| fil     | Filipino                      | kw    | Cornish                     |
| fo      | Faroese                       | ky    | Kyrgyz                      |
| fr      | French <sup>ul</sup>          | lag   | Langi                       |
| fr-BE   | French <sup>ul</sup>          | lb    | Luxembourgish               |
| fr-CA   | French <sup>ul</sup>          | lg    | Ganda                       |
| fr-CH   | French <sup>ul</sup>          | lkt   | Lakota                      |
| fr-LU   | French <sup>ul</sup>          | ln    | Lingala                     |
| fur     | Friulian <sup>ul</sup>        | lo    | Lao <sup>ul</sup>           |
| fy      | Western Frisian               | lrc   | Northern Luri               |
| ga      | Irish <sup>ul</sup>           | lt    | Lithuanian <sup>ul</sup>    |
| gd      | Scottish Gaelic <sup>ul</sup> | lu    | Luba-Katanga                |

|         |                                 |            |                             |
|---------|---------------------------------|------------|-----------------------------|
| luo     | Luo                             | sa-Mlym    | Sanskrit                    |
| luy     | Luyia                           | sa-Telu    | Sanskrit                    |
| lv      | Latvian <sup>ul</sup>           | sa         | Sanskrit                    |
| mas     | Masai                           | sah        | Sakha                       |
| mer     | Meru                            | saq        | Samburu                     |
| mfe     | Morisyen                        | sbp        | Sangu                       |
| mg      | Malagasy                        | se         | Northern Sami <sup>ul</sup> |
| mgf     | Makhuwa-Meetto                  | seh        | Sena                        |
| mgo     | Meta'                           | ses        | Koyraboro Senni             |
| mk      | Macedonian <sup>ul</sup>        | sg         | Sango                       |
| ml      | Malayalam <sup>ul</sup>         | shi-Latn   | Tachelhit                   |
| mn      | Mongolian                       | shi-Tfng   | Tachelhit                   |
| mr      | Marathi <sup>ul</sup>           | shi        | Tachelhit                   |
| ms-BN   | Malay <sup>l</sup>              | si         | Sinhala                     |
| ms-SG   | Malay <sup>l</sup>              | sk         | Slovak <sup>ul</sup>        |
| ms      | Malay <sup>ul</sup>             | sl         | Slovenian <sup>ul</sup>     |
| mt      | Maltese                         | smn        | Inari Sami                  |
| mua     | Mundang                         | sn         | Shona                       |
| my      | Burmese                         | so         | Somali                      |
| mzn     | Mazanderani                     | sq         | Albanian <sup>ul</sup>      |
| naq     | Nama                            | sr-Cyrl-BA | Serbian <sup>ul</sup>       |
| nb      | Norwegian Bokmål <sup>ul</sup>  | sr-Cyrl-ME | Serbian <sup>ul</sup>       |
| nd      | North Ndebele                   | sr-Cyrl-XK | Serbian <sup>ul</sup>       |
| ne      | Nepali                          | sr-Cyrl    | Serbian <sup>ul</sup>       |
| nl      | Dutch <sup>ul</sup>             | sr-Latn-BA | Serbian <sup>ul</sup>       |
| nmg     | Kwasio                          | sr-Latn-ME | Serbian <sup>ul</sup>       |
| nn      | Norwegian Nynorsk <sup>ul</sup> | sr-Latn-XK | Serbian <sup>ul</sup>       |
| nnh     | Ngiemboon                       | sr-Latn    | Serbian <sup>ul</sup>       |
| nus     | Nuer                            | sr         | Serbian <sup>ul</sup>       |
| nyn     | Nyankole                        | sv         | Swedish <sup>ul</sup>       |
| om      | Oromo                           | sw         | Swahili                     |
| or      | Odia                            | ta         | Tamil <sup>u</sup>          |
| os      | Ossetic                         | te         | Telugu <sup>ul</sup>        |
| pa-Arab | Punjabi                         | teo        | Teso                        |
| pa-Guru | Punjabi                         | th         | Thai <sup>ul</sup>          |
| pa      | Punjabi                         | ti         | Tigrinya                    |
| pl      | Polish <sup>ul</sup>            | tk         | Turkmen <sup>ul</sup>       |
| pms     | Piedmontese <sup>ul</sup>       | to         | Tongan                      |
| ps      | Pashto                          | tr         | Turkish <sup>ul</sup>       |
| pt-BR   | Portuguese <sup>ul</sup>        | twq        | Tasawaq                     |
| pt-PT   | Portuguese <sup>ul</sup>        | tzm        | Central Atlas Tamazight     |
| pt      | Portuguese <sup>ul</sup>        | ug         | Uyghur                      |
| qu      | Quechua                         | uk         | Ukrainian <sup>ul</sup>     |
| rm      | Romansh <sup>ul</sup>           | ur         | Urdu <sup>ul</sup>          |
| rn      | Rundi                           | uz-Arab    | Uzbek                       |
| ro      | Romanian <sup>ul</sup>          | uz-Cyrl    | Uzbek                       |
| rof     | Rombo                           | uz-Latn    | Uzbek                       |
| ru      | Russian <sup>ul</sup>           | uz         | Uzbek                       |
| rw      | Kinyarwanda                     | vai-Latn   | Vai                         |
| rwk     | Rwa                             | vai-Vaii   | Vai                         |
| sa-Beng | Sanskrit                        | vai        | Vai                         |
| sa-Deva | Sanskrit                        | vi         | Vietnamese <sup>ul</sup>    |
| sa-Gujr | Sanskrit                        | vun        | Vunjo                       |
| sa-Knda | Sanskrit                        | wae        | Walser                      |

|            |                                |            |         |
|------------|--------------------------------|------------|---------|
| xog        | Soga                           | zh-Hans-MO | Chinese |
| yav        | Yangben                        | zh-Hans-SG | Chinese |
| yi         | Yiddish                        | zh-Hans    | Chinese |
| yo         | Yoruba                         | zh-Hant-HK | Chinese |
| yue        | Cantonese                      | zh-Hant-MO | Chinese |
| zgh        | Standard Moroccan<br>Tamazight | zh-Hant    | Chinese |
| zh-Hans-HK | Chinese                        | zh         | Chinese |
|            |                                | zu         | Zulu    |

---

In some contexts (currently `\babelfont`) an `ini` file may be loaded by its name. Here is the list of the names currently supported. With these languages, `\babelfont` loads (if not done before) the language and script names (even if the language is defined as a package option with an `ldf` file). These are also the names recognized by `\babelprovide` with a valueless `import`.

---

|                      |                                      |
|----------------------|--------------------------------------|
| aghem                | brazilian                            |
| akan                 | breton                               |
| albanian             | british                              |
| american             | bulgarian                            |
| amharic              | burmese                              |
| arabic               | canadian                             |
| arabic-algeria       | cantonese                            |
| arabic-DZ            | catalan                              |
| arabic-morocco       | centralatlastamazight                |
| arabic-MA            | centralkurdish                       |
| arabic-syria         | chechen                              |
| arabic-SY            | cherokee                             |
| armenian             | chiga                                |
| assamese             | chinese-hans-hk                      |
| asturian             | chinese-hans-mo                      |
| asu                  | chinese-hans-sg                      |
| australian           | chinese-hans                         |
| austrian             | chinese-hant-hk                      |
| azerbaijani-cyrillic | chinese-hant-mo                      |
| azerbaijani-cyrl     | chinese-hant                         |
| azerbaijani-latin    | chinese-simplified-hongkongsarchina  |
| azerbaijani-latn     | chinese-simplified-macausarchina     |
| azerbaijani          | chinese-simplified-singapore         |
| bafia                | chinese-simplified                   |
| bambara              | chinese-traditional-hongkongsarchina |
| basaa                | chinese-traditional-macausarchina    |
| basque               | chinese-traditional                  |
| belarusian           | chinese                              |
| bemba                | cognian                              |
| bena                 | cornish                              |
| bengali              | croatian                             |
| bodo                 | czech                                |
| bosnian-cyrillic     | danish                               |
| bosnian-cyrl         | duala                                |
| bosnian-latin        | dutch                                |
| bosnian-latn         | dzongkha                             |
| bosnian              | embu                                 |

english-au  
english-australia  
english-ca  
english-canada  
english-gb  
english-newzealand  
english-nz  
english-unitedkingdom  
english-unitedstates  
english-us  
english  
esperanto  
estonian  
ewe  
ewondo  
faroese  
filipino  
finnish  
french-be  
french-belgium  
french-ca  
french-canada  
french-ch  
french-lu  
french-luxembourg  
french-switzerland  
french  
friulian  
fulah  
galician  
ganda  
georgian  
german-at  
german-austria  
german-ch  
german-switzerland  
german  
greek  
gujarati  
gusii  
hausa-gh  
hausa-ghana  
hausa-ne  
hausa-niger  
hausa  
hawaiian  
hebrew  
hindi  
hungarian  
icelandic  
igbo  
inarisami  
indonesian  
interlingua

irish  
italian  
japanese  
jolafonyi  
kabuverdianu  
kabyle  
kako  
kalaallisut  
kalenjin  
kamba  
kannada  
kashmiri  
kazakh  
khmer  
kikuyu  
kinyarwanda  
konkani  
korean  
koyraborosenni  
koyrachiini  
kwasio  
kyrgyz  
lakota  
langi  
lao  
latvian  
lingala  
lithuanian  
lowersorbian  
lsorbian  
lubakatanga  
luo  
luxembourgish  
luyia  
macedonian  
machame  
makhuwameetto  
makonde  
malagasy  
malay-bn  
malay-brunei  
malay-sg  
malay-singapore  
malay  
malayalam  
maltese  
manx  
marathi  
masai  
mazanderani  
meru  
meta  
mexican  
mongolian

morisyen  
mundang  
nama  
nepali  
newzealand  
ngiemboon  
ngomba  
norsk  
northernluri  
northernsami  
northndebele  
norwegianbokmal  
norwegiannynorsk  
nswissgerman  
nuer  
nyankole  
nynorsk  
occitan  
oriya  
oromo  
ossetic  
pashto  
persian  
piedmontese  
polish  
portuguese-br  
portuguese-brazil  
portuguese-portugal  
portuguese-pt  
portuguese  
punjabi-arab  
punjabi-arabic  
punjabi-gurmukhi  
punjabi-guru  
punjabi  
quechua  
romanian  
romansh  
rombo  
rundi  
russian  
rwa  
sakha  
samburu  
samin  
sango  
sangu  
sanskrit-beng  
sanskrit-bengali  
sanskrit-deva  
sanskrit-devanagari  
sanskrit-gujarati  
sanskrit-gujr  
sanskrit-kannada

sanskrit-knda  
sanskrit-malayalam  
sanskrit-mlym  
sanskrit-telu  
sanskrit-telugu  
sanskrit  
scottishgaelic  
sena  
serbian-cyrillic-bosniaherzegovina  
serbian-cyrillic-kosovo  
serbian-cyrillic-montenegro  
serbian-cyrillic  
serbian-cyrl-ba  
serbian-cyrl-me  
serbian-cyrl-xk  
serbian-cyrl  
serbian-latin-bosniaherzegovina  
serbian-latin-kosovo  
serbian-latin-montenegro  
serbian-latin  
serbian-latn-ba  
serbian-latn-me  
serbian-latn-xk  
serbian-latn  
serbian  
shambala  
shona  
sichuanyi  
sinhala  
slovak  
slovene  
slovenian  
soga  
somali  
spanish-mexico  
spanish-mx  
spanish  
standardmoroccantamazight  
swahili  
swedish  
swissgerman  
tachelhit-latin  
tachelhit-latn  
tachelhit-tfng  
tachelhit-tifinagh  
tachelhit  
taita  
tamil  
tasawaq  
telugu  
teso  
thai  
tibetan  
tigrinya



|                |                |
|----------------|----------------|
| tongan         | vai-latin      |
| turkish        | vai-latn       |
| turkmen        | vai-vai        |
| ukenglish      | vai-vaii       |
| ukrainian      | vai            |
| upporsorbian   | vietnam        |
| urdu           | vietnamese     |
| usenglish      | vunjo          |
| usorbian       | walser         |
| uyghur         | welsh          |
| uzbek-arab     | westernfrisian |
| uzbek-arabic   | yangben        |
| uzbek-cyrillic | yiddish        |
| uzbek-cyrl     | yoruba         |
| uzbek-latin    | zarma          |
| uzbek-latn     | zulu afrikaans |
| uzbek          |                |

## 1.14 Selecting fonts

**New 3.15** Babel provides a high level interface on top of fontspec to select fonts. There is no need to load fontspec explicitly – babel does it for you with the first `\babelfont`.<sup>13</sup>

`\babelfont` [*<language-list>*][*<font-family>*][*<font-options>*]{*<font-name>*}

Here *font-family* is *rm*, *sf* or *tt* (or newly defined ones, as explained below), and *font-name* is the same as in fontspec and the like.

If no language is given, then it is considered the default font for the family, activated when a language is selected. On the other hand, if there is one or more languages in the optional argument, the font will be assigned to them, overriding the default. Alternatively, you may set a font for a script – just precede its name (lowercase) with a star (eg, *\*devanagari*).

Babel takes care of the font language and the font script when languages are selected (as well as the writing direction); see the recognized languages above. In most cases, you will not need *font-options*, which is the same as in fontspec, but you may add further key/value pairs if necessary.

**EXAMPLE** Usage in most cases is very simple. Let us assume you are setting up a document in Swedish, with some words in Hebrew, with a font suited for both languages.

```
\documentclass{article}

\usepackage[swedish,␣bidi=default]{babel}

\babelprovide[import]{hebrew}

\babelfont{rm}{FreeSerif}

\begin{document}

Svenska␣\foreignlanguage{hebrew}{עִבְרִית}␣svenska.

\end{document}
```

<sup>13</sup>See also the package `combofont` for a complementary approach.

If on the other hand you have to resort to different fonts, you could replace the red line above with, say:

```
\babelfont{rm}{Iwona}  
\babelfont[hebrew]{rm}{FreeSerif}
```

`\babelfont` can be used to implicitly define a new font family. Just write its name instead of `rm`, `sf` or `tt`. This is the preferred way to select fonts in addition to the three basic families.

**EXAMPLE** Here is how to do it:

```
\babelfont{kai}{FandolKai}
```

Now, `\kaifamily` and `\kaidefault`, as well as `\textkai` are at your disposal.

**NOTE** You may load `fontspec` explicitly. For example:

```
\usepackage{fontspec}  
\newfontscript{Devanagari}{deva}  
\babelfont[hindi]{rm}{Shobhika}
```

This makes sure the OpenType script for Devanagari is `deva` and not `dev2` (luatex does not detect automatically the correct script<sup>14</sup>). You may also pass some options to `fontspec`: with `silent`, the warnings about unavailable scripts or languages are not shown (they are only really useful when the document format is being set up).

**NOTE** Directionality is a property affecting margins, indentation, column order, etc., not just text. Therefore, it is under the direct control of the language, which applies both the script and the direction to the text. As a consequence, there is no need to set `Script` when declaring a font (nor `Language`). In fact, it is even discouraged.

**NOTE** `\fontspec` is not touched at all, only the preset font families (`rm`, `sf`, `tt`, and the like). If a language is switched when an *ad hoc* font is active, or you select the font with this command, neither the script nor the language is passed. You must add them by hand. This is by design, for several reasons (for example, each font has its own set of features and a generic setting for several of them could be problematic, and also a “lower-level” font selection is useful).

**NOTE** The keys `Language` and `Script` just pass these values to the *font*, and do *not* set the script for the *language* (and therefore the writing direction). In other words, the `ini` file or `\babelprovide` provides default values for `\babelfont` if omitted, but the opposite is not true. See the note above for the reasons of this behavior.

**WARNING** Do not use `\setxxxxfont` and `\babelfont` at the same time. `\babelfont` follows the standard  $\TeX$  conventions to set the basic families – define `\xxdefault`, and activate it with `\xxfamily`. On the other hand, `\setxxxxfont` in `fontspec` takes a different approach, because `\xxfamily` is redefined with the family name hardcoded (so that `\xxdefault` becomes no-op). Of course, both methods are incompatible, and if you use `\setxxxxfont`, font switching with `\babelfont` just does *not* work (nor the standard `\xxdefault`, for that matter). As of [New 3.34](#) there is an attempt to make them compatible, but the language system will not be set by babel and should be set with `fontspec` if necessary.

<sup>14</sup>And even with the correct code some fonts could be rendered incorrectly by `fontspec`, so double-check the results. `xetex` fares better, but some fonts are still problematic.

**TROUBLESHOOTING** *Package fontspec Warning: 'Language 'LANG' not available for font 'FONT' with script 'SCRIPT' 'Default' language used instead'. This is **not** and error.* This warning is shown by fontspec, not by babel. It could be irrelevant for English, but not for many other languages, including Urdu and Turkish. This is a useful and harmless warning, and if everything is fine with your document the best thing you can do is just to ignore it altogether.

**TROUBLESHOOTING** *Package babel Warning: The following fonts are not babel standard families. This is **not** and error.* The main purpose of `\babel font` is to define at once in a multilingual document the fonts required by the different languages, with their corresponding language systems (script and language). So, if you load, say, 4 languages, `\babel font{rm}{FreeSerif}` defines 4 fonts (with their variants, of course), which are switched with the language by babel. It's just a tool to make things easier and transparent to the user.

There is no real need to use `\babel font` in a monolingual document, if you set the language system in `\setmainfont` (or not, depending on what you want).

babel assumes that if you are using `\babel font` for a family, very likely you want to define the rest of them. If you don't, you can find some inconsistencies between families. This checking is done at the beginning of the document, at a point where we cannot know which families will be used.

As the message explains, *there is nothing intrinsically wrong* with not defining all the families. In fact, there is nothing intrinsically wrong with not using `\babel font` at all. But you must be aware that this may lead to some problems. And this is the very reason of the warning.

## 1.15 Modifying a language

Modifying the behavior of a language (say, the chapter “caption”), is sometimes necessary, but not always trivial.

- The old way, still valid for many languages, to redefine a caption is the following:

```
\addto\captionenglish{%  
  \renewcommand\contentsname{Foo}%  
}
```

As of 3.15, there is no need to hide spaces with % (babel removes them), but it is advisable to do it.

- The new way, which is found in bulgarian, azerbaijani, spanish, french, turkish, icelandic, vietnamese and a few more, as well as in languages created with `\babel provide` and its key import, is:

```
\renewcommand\spanishchaptername{Foo}
```

- Macros to be run when a language is selected can be add to `\extras<lang>`:

```
\addto\extrasrussian{\mymacro}
```

There is a counterpart for code to be run when a language is unselected: `\noextras<lang>`.

**NOTE** Do *not* redefine a caption in the following way:

```
\AtBeginDocument{\renewcommand\contentsname{Foo}}
```

The changes may be discarded with a language selector, and the original value restored.

**NOTE** These macros (`\captions $\langle lang \rangle$` , `\extras $\langle lang \rangle$` ) may be redefined, but *must not* be used as such – they just pass information to babel, which executes them in the proper context.

Another way to modify a language loaded as a package or class option is by means of `\babelprovide`, described below in depth. So, something like:

```
\usepackage[danish]{babel}
\babelprovide[captions=da,hyphenrules=nohyphenation]{danish}
```

first loads `danish.ldf`, and then redefines the captions for danish (as provided by the `ini` file) and prevents hyphenation. The rest of the language definitions are not touched.

## 1.16 Creating a language

**New 3.10** And what if there is no style for your language or none fits your needs? You may then define quickly a language with the help of the following macro in the preamble (which may be used to modify an existing language, too, as explained in the previous subsection).

`\babelprovide` [ $\langle options \rangle$ ] { $\langle language-name \rangle$ }

If the language  $\langle language-name \rangle$  has not been loaded as class or package option and there are no  $\langle options \rangle$ , it creates an “empty” one with some defaults in its internal structure: the hyphen rules, if not available, are set to the current ones, left and right hyphen mins are set to 2 and 3. In either case, caption, date and language system are not defined.

If no `ini` file is imported with `import`,  $\langle language-name \rangle$  is relevant because in such a case the hyphenation rules (including those for South East Asian and CJK) are based on it as provided in the `ini` file corresponding to that name; the same applies to OpenType language and script.

Conveniently, some options allow to fill the language, and babel warns you about what to do if there is a missing string. Very likely you will find alerts like that in the log file:

```
Package_babel_Warning: mylangchaptername not set. Please, define
(babel) it in the preamble with something like:
(babel) \renewcommand\mylangchaptername{..}
(babel) Reported on input line 18.
```

In most cases, you will only need to define a few macros.

**EXAMPLE** If you need a language named `arhinish`:

```
\usepackage[danish]{babel}
\babelprovide{arhinish}
\renewcommand\arhinishchaptername{Chapitula}
\renewcommand\arhinishrefname{Refirenke}
\renewcommand\arhinishhyphenmins{22}
```

The main language is not changed (danish in this example). So, you must add `\selectlanguage{arhinish}` or other selectors where necessary. If the language has been loaded as an argument in `\documentclass` or `\usepackage`, then `\babelprovide` redefines the requested data.

**import=** *<language-tag>*

**New 3.13** Imports data from an ini file, including captions, date, and hyphenmins. For example:

```
\babelprovide[import=hu]{hungarian}
```

Unicode engines load the UTF-8 variants, while 8-bit engines load the LICR (ie, with macros like `\'` or `\ss`) ones.

**New 3.23** It may be used without a value. In such a case, the ini file set in the corresponding `babel-<language>.tex` (where `<language>` is the last argument in `\babelprovide`) is imported. See the list of recognized languages above. So, the previous example could be written:

```
\babelprovide[import]{hungarian}
```

There are about 200 ini files, with data taken from the ldf files and the CLDR provided by Unicode. Not all languages in the latter are complete, and therefore neither are the ini files. A few languages will show a warning about the current lack of suitability of the date format (hindi, french, breton, and occitan).

Besides `\today`, this option defines an additional command for dates: `\<language>date`, which takes three arguments, namely, year, month and day numbers. In fact, `\today` calls `\<language>today`, which in turn calls `\<language>date{\the\year}{\the\month}{\the\day}`.

**captions=** *<language-tag>*

Loads only the strings. For example:

```
\babelprovide[captions=hu]{hungarian}
```

**hyphenrules=** *<language-list>*

With this option, with a space-separated list of hyphenation rules, babel assigns to the language the first valid hyphenation rules in the list. For example:

```
\babelprovide[hyphenrules=chavacano_italian]{chavacano}
```

If none of the listed hyphenrules exist, the default behavior applies. Note in this example we set chavacano as first option – without it, it would select spanish even if chavacano exists.

A special value is `+`, which allocates a new language (in the TeX sense). It only makes sense as the last value (or the only one; the subsequent ones are silently ignored). It is mostly useful with luatex, because you can add some patterns with `\babelpatterns`, as for example:

```
\babelprovide[hyphenrules=+]{neo}
\babelpatterns[neo]{a1_e1_i1_o1_u1}
```

In other engines it just suppresses hyphenation (because the pattern list is empty).

**main** This valueless option makes the language the main one. Only in newly defined languages.

**script=**  $\langle script-name \rangle$

**New 3.15** Sets the script name to be used by fontspec (eg, Devanagari). Overrides the value in the ini file. If fontspec does not define it, then babel sets its tag to that provided by the ini file. This value is particularly important because it sets the writing direction, so you must use it if for some reason the default value is wrong.

**language=**  $\langle language-name \rangle$

**New 3.15** Sets the language name to be used by fontspec (eg, Hindi). Overrides the value in the ini file. If fontspec does not define it, then babel sets its tag to that provided by the ini file. Not so important, but sometimes still relevant.

A few options (only luatex) set some properties of the writing system used by the language. These properties are *always* applied to the script, no matter which language is active. Although somewhat inconsistent, this makes setting a language up easier in most typical cases.

**mapfont=** direction

Assigns the font for the writing direction of this language (only with bidi=basic).<sup>15</sup> More precisely, what mapfont=direction means is, ‘when a character has the same direction as the script for the “provided” language, then change its font to that set for this language’. There are 3 directions, following the bidi Unicode algorithm, namely, Arabic-like, Hebrew-like and left to right.<sup>16</sup> So, there should be at most 3 directives of this kind.

**intraspace=**  $\langle base \rangle \langle shrink \rangle \langle stretch \rangle$

Sets the interword space for the writing system of the language, in em units (so, 0 .1 0 is 0em plus .1em). Like \spaceskip, the em unit applied is that of the current text (more precisely, the previous glyph). Currently used only in Southeast Asian scripts, like Thai, and CJK.

**intrapenalty=**  $\langle penalty \rangle$

Sets the interword penalty for the writing system of this language. Currently used only in Southeast Asian scripts, like Thai. Ignored if 0 (which is the default value).

**NOTE** (1) If you need shorthands, you can define them with \usesshorthands and \defineshortand as described above. (2) Captions and \today are “ensured” with \babelensure (this is the default in ini-based languages).

## 1.17 Digits

**New 3.20** About thirty ini files define a field named digits.native. When it is present, two macros are created: \<language>digits and \<language>counter (only xetex and

<sup>15</sup>There will be another value, language, not yet implemented.

<sup>16</sup>In future releases a new value (script) will be added.

luatex). With the first, a string of ‘Latin’ digits are converted to the native digits of that language; the second takes a counter name as argument. With the option `maparabic` in `\babelprovide`, `\arabic` is redefined to produce the native digits (this is done *globally*, to avoid inconsistencies in, for example, page numbering, and note as well dates do not rely on `\arabic`.)

For example:

```
\babelprovide[import]{telugu}\%Telugu\better\with\XeTeX
\%Or\also,\if\you\want:
\% \babelprovide[import,\maparabic]{telugu}
\babelfont{rm}{Gautami}
\begin{document}
\telugudigits{1234}
\telugucounter{section}
\end{document}
```

Languages providing native digits in all or some variants are *ar*, *as*, *bn*, *bo*, *brx*, *ckb*, *dz*, *fa*, *gu*, *hi*, *km*, *kn*, *kok*, *ks*, *lo*, *lrc*, *ml*, *mr*, *my*, *mzn*, *ne*, *or*, *pa*, *ps*, *ta*, *te*, *th*, *ug*, *ur*, *uz*, *vai*, *yue*, *zh*.

**New 3.30** With luatex there is an alternative approach for mapping digits, namely, `mapdigits`. Conversion is based on the language and it is applied to the typeset text (not math, PDF bookmarks, etc.) before bidi and fonts are processed (ie, to the node list as generated by the T<sub>E</sub>X code). This means the local digits have the correct bidirectional behavior (unlike `Numbers=Arabic` in `fontspec`, which is not recommended).

## 1.18 Getting the current language name

`\language` The control sequence `\language` contains the name of the current language.

**WARNING** Due to some internal inconsistencies in catcodes, it should *not* be used to test its value. Use `iflang`, by Heiko Oberdiek.

`\iflanguage` `{⟨language⟩}{⟨true⟩}{⟨false⟩}`

If more than one language is used, it might be necessary to know which language is active at a specific time. This can be checked by a call to `\iflanguage`, but note here “language” is used in the T<sub>E</sub>X sense, as a set of hyphenation patterns, and *not* as its babel name. This macro takes three arguments. The first argument is the name of a language; the second and third arguments are the actions to take if the result of the test is true or false respectively.

**WARNING** The advice about `\language` also applies here – use `iflang` instead of `\iflanguage` if possible.

## 1.19 Hyphenation and line breaking

`\babelhyphen` `*{⟨type⟩}`

`\babelhyphen` `*{⟨text⟩}`

**New 3.9a** It is customary to classify hyphens in two types: (1) *explicit* or *hard hyphens*, which in T<sub>E</sub>X are entered as `-`, and (2) *optional* or *soft hyphens*, which are entered as `\-`. Strictly, a *soft hyphen* is not a hyphen, but just a breaking opportunity or, in T<sub>E</sub>X terms, a “discretionary”; a *hard hyphen* is a hyphen with a breaking opportunity after it. A further type is a *non-breaking hyphen*, a hyphen without a breaking opportunity.

In  $\TeX$ , - and \- forbid further breaking opportunities in the word. This is the desired behavior very often, but not always, and therefore many languages provide shorthands for these cases. Unfortunately, this has not been done consistently: for example, "- in Dutch, Portuguese, Catalan or Danish is a hard hyphen, while in German, Spanish, Norwegian, Slovak or Russian is a soft hyphen. Furthermore, some of them even redefine \-, so that you cannot insert a soft hyphen without breaking opportunities in the rest of the word. Therefore, some macros are provided with a set of basic “hyphens” which can be used by themselves, to define a user shorthand, or even in language files.

- `\babelhyphen{soft}` and `\babelhyphen{hard}` are self explanatory.
- `\babelhyphen{repeat}` inserts a hard hyphen which is repeated at the beginning of the next line, as done in languages like Polish, Portuguese and Spanish.
- `\babelhyphen{nobreak}` inserts a hard hyphen without a break after it (even if a space follows).
- `\babelhyphen{empty}` inserts a break opportunity without a hyphen at all.
- `\babelhyphen{<text>}` is a hard “hyphen” using `<text>` instead. A typical case is `\babelhyphen{/}`.

With all of them, hyphenation in the rest of the word is enabled. If you don’t want to enable it, there is a starred counterpart: `\babelhyphen*{soft}` (which in most cases is equivalent to the original \-), `\babelhyphen*{hard}`, etc.

Note `hard` is also good for isolated prefixes (eg, *anti-*) and `nobreak` for isolated suffixes (eg, *-ism*), but in both cases `\babelhyphen*{nobreak}` is usually better.

There are also some differences with  $\TeX$ : (1) the character used is that set for the current font, while in  $\TeX$  it is hardwired to - (a typical value); (2) the hyphen to be used in fonts with a negative `\hyphenchar` is -, like in  $\TeX$ , but it can be changed to another value by redefining `\babelnullhyphen`; (3) a break after the hyphen is forbidden if preceded by a glue  $>0$  pt (at the beginning of a word, provided it is not immediately preceded by, say, a parenthesis).

**`\babelhyphenation`** [`<language>` , `<language>` , ... ] { `<exceptions>` }

**New 3.9a** Sets hyphenation exceptions for the languages given or, without the optional argument, for *all* languages (eg, proper nouns or common loan words, and of course monolingual documents). Language exceptions take precedence over global ones. It can be used only in the preamble, and exceptions are set when the language is first selected, thus taking into account changes of `\lccodes`’s done in `\extras<lang>` as well as the language-specific encoding (not set in the preamble by default). Multiple `\babelhyphenation`’s are allowed. For example:

```
\babelhyphenation{Wal-hal-la_Dar-bhan-ga}
```

Listed words are saved expanded and therefore it relies on the LICR. Of course, it also works without the LICR if the input and the font encodings are the same, like in Unicode based engines.

**NOTE** Using `\babelhyphenation` with Southeast Asian scripts is mostly pointless. But with `\babelpatterns` (below) you may fine-tune line breaking (only `luatex`). Even if there are no patterns for the language, you can add at least some typical cases.



**\babelpatterns** [*<language>* , *<language>* , ... ] { *<patterns>* }

**New 3.9m** In *luatex* only,<sup>17</sup> adds or replaces patterns for the languages given or, without the optional argument, for *all* languages. If a pattern for a certain combination already exists, it gets replaced by the new one.

It can be used only in the preamble, and patterns are added when the language is first selected, thus taking into account changes of `\lccodes`'s done in `\extras<lang>` as well as the language-specific encoding (not set in the preamble by default). Multiple `\babelpatterns`'s are allowed.

Listed patterns are saved expanded and therefore it relies on the LICR. Of course, it also works without the LICR if the input and the font encodings are the same, like in Unicode based engines.

**New 3.31** (Only *luatex*.) With `\babelprovide` and imported CJK languages, a simple generic line breaking algorithm (push-out-first) is applied, based on a selection of the Unicode rules ( **New 3.32** it is disabled in verbatim mode, or more precisely when the `hyphenrules` are set to `nohyphenation`). It can be activated alternatively by setting explicitly the `intraspace`.

**New 3.27** Interword spacing for Thai, Lao and Khemer is activated automatically if a language with one of those scripts are loaded with `\babelprovide`. See the sample on the babel repository. With both Unicode engines, spacing is based on the “current” em unit (the size of the previous char in *luatex*, and the font size set by the last `\selectfont` in *xetex*).

## 1.20 Selecting scripts

Currently babel provides no standard interface to select scripts, because they are best selected with either `\fontencoding` (low-level) or a language name (high-level). Even the Latin script may require different encodings (ie, sets of glyphs) depending on the language, and therefore such a switch would be in a sense incomplete.<sup>18</sup>

Some languages sharing the same script define macros to switch it (eg, `\textcyrillic`), but be aware they may also set the language to a certain default. Even the babel core defined `\textlatin`, but it was somewhat buggy because in some cases it messed up encodings and fonts (for example, if the main Latin encoding was `LY1`), and therefore it has been deprecated.<sup>19</sup>

**\ensureascii** { *<text>* }

**New 3.9i** This macro makes sure *<text>* is typeset with a LICR-savvy encoding in the ASCII range. It is used to redefine `\TeX` and `\LaTeX` so that they are correctly typeset even with `LGR` or `X2` (the complete list is stored in `\BabelNonASCII`, which by default is `LGR`, `X2`, `OT2`, `OT3`, `OT6`, `LHE`, `LWN`, `LMA`, `LMC`, `LMS`, `LMU`, but you can modify it). So, in some sense it fixes the bug described in the previous paragraph.

If non-ASCII encodings are not loaded (or no encoding at all), it is no-op (also `\TeX` and `\LaTeX` are not redefined); otherwise, `\ensureascii` switches to the encoding at the beginning of the document if ASCII-savvy, or else the last ASCII-savvy encoding loaded. For example, if you load `LY1`, `LGR`, then it is set to `LY1`, but if you load `LY1`, `T2A` it is set to `T2A`. The symbol encodings `TS1`, `T3`, and `TS3` are not taken into account, since they are not used for “ordinary” text (they are stored in `\BabelNonText`, used in some special cases when no Latin encoding is explicitly set).

<sup>17</sup>With *luatex* exceptions and patterns can be modified almost freely. However, this is very likely a task for a separate package and babel only provides the most basic tools.

<sup>18</sup>The so-called Unicode fonts do not improve the situation either. So, a font suited for Vietnamese is not necessarily suited for, say, the romanization of Indic languages, and the fact it contains glyphs for Modern Greek does not mean it includes them for Classic Greek.

<sup>19</sup>But still defined for backwards compatibility.

The foregoing rules (which are applied “at begin document”) cover most of the cases. No assumption is made on characters above 127, which may not follow the LICR conventions – the goal is just to ensure most of the ASCII letters and symbols are the right ones.

## 1.21 Selecting directions

No macros to select the writing direction are provided, either – writing direction is intrinsic to each script and therefore it is best set by the language (which could be a dummy one). Furthermore, there are in fact two right-to-left modes, depending on the language, which differ in the way ‘weak’ numeric characters are ordered (eg, Arabic %123 vs Hebrew 123%).

**WARNING** The current code for **text** in luatex should be considered essentially stable, but, of course, it is not bug-free and there could be improvements in the future, because setting bidi text has many subtleties (see for example <https://www.w3.org/TR/html-bidi/>). A basic stable version for other engines must wait. This applies to text; there is a basic support for **graphical** elements, including the picture environment (with pict2e) and pfg/tikz. Also, indexes and the like are under study, as well as math (there is progress in the latter, too, but for example cases may fail).

An effort is being made to avoid incompatibilities in the future (this one of the reason currently bidi must be explicitly requested as a package option, with a certain bidi model, and also the layout options described below).

There are some package options controlling bidi writing.

**bidi=** default | basic | basic-r | bidi-l | bidi-r

**New 3.14** Selects the bidi algorithm to be used. With default the bidi mechanism is just activated (by default it is not), but every change must be marked up. In xetex and pdftex this is the only option.

In luatex, basic-r provides a simple and fast method for R text, which handles numbers and unmarked L text within an R context many in typical cases. **New 3.19** Finally, basic supports both L and R text, and it is the preferred method (support for basic-r is currently limited). (They are named basic mainly because they only consider the intrinsic direction of scripts and weak directionality.)

**New 3.29** In xetex, bidi-r and bidi-l resort to the package bidi (by Vafa Khalighi). Integration is still somewhat tentative, but it mostly works. For RL documents use the former, and for LR ones use the latter.

**New 3.32** There is some experimental support for harftex. Since it is based on luatex, the option basic mostly works. You may need to deactivate the rtlm or the rtla font features (besides loading harfload before babel and activating mode=harf; there is a sample in the GitHub repository).

There are samples on GitHub, under /required/babel/samples. See particularly lua-bidibasic.tex and lua-secenum.tex.

**EXAMPLE** The following text comes from the Arabic Wikipedia (article about Arabia). Copy-pasting some text from the Wikipedia is a good way to test this feature. Remember basic-r is available in luatex only.

```
\documentclass{article}

\usepackage[bidi=basic-r]{babel}
```

```

\babelprovide[import,main]{arabic}

\babelfont{rm}{FreeSerif}

\begin{document}

وقد عرف شبه جزيرة العرب طيلة العصر الهليني (الغريقي) بـ
Aravia أو Arabia (بالغريقية  $\text{Αραβία}$ )، استخدم الرومان ثلاث
بدايات بـ "Arabia" على ثلاث مناطق من شبه الجزيرة العربية، إلا أنها
حقيقةً كانت أكبر مما تعرف عليه اليوم.

\end{document}

```

**EXAMPLE** With `bidi=basic` both L and R text can be mixed without explicit markup (the latter will be only necessary in some special cases where the Unicode algorithm fails). It is used much like `bidi=basic-r`, but with R text inside L text you may want to map the font so that the correct features are in force. This is accomplished with an option in `\babelprovide`, as illustrated:

```

\documentclass{book}

\usepackage[english,bidi=basic]{babel}

\babelprovide[mapfont=direction]{arabic}

\babelfont{rm}{Crimson}
\babelfont[*arabic]{rm}{FreeSerif}

\begin{document}

Most Arabic speakers consider the two varieties to be two registers
of one language, although the two registers can be referred to in
Arabic as فصحى (MSA) and فصحى التراث (CA).

\end{document}

```

In this example, and thanks to `mapfont=direction`, any Arabic letter (because the language is arabic) changes its font to that set for this language (here defined via `*arabic`, because `Crimson` does not provide Arabic letters).

**NOTE** Boxes are “black boxes”. Numbers inside an `\hbox` (for example in a `\ref`) do not know anything about the surrounding chars. So, `\ref{A}-\ref{B}` are not rendered in the visual order A-B, but in the wrong one B-A (because the hyphen does not “see” the digits inside the `\hbox`’es). If you need `\ref` ranges, the best option is to define a dedicated macro like this (to avoid explicit direction changes in the body; here `\texthe` must be defined to select the main language):

```

\newcommand\refrange[2]{\babelsublr{\texthe{\ref{#1}}-\texthe{\ref{#2}}}}

```

In the future a more complete method, reading recursively boxed text, may be added.

**layout=** sectioning | counters | lists | contents | footnotes | captions | columns | graphics | extras

**New 3.16** *To be expanded.* Selects which layout elements are adapted in bidi documents, including some text elements (except with options loading the `bidi` package, which provides its own mechanism to control these elements). You may use several options with a dot-separated list (eg, `layout=counters.contents.sectioning`). This list will be expanded in future releases. Note not all options are required by all engines.

**sectioning** makes sure the sectioning macros are typeset in the main language, but with the title text in the current language (see below `\BabelPatchSection` for further details).

**counters** required in all engines (except `luatex` with `bidi=basic`) to reorder section numbers and the like (eg, `\subsection{<subsection>.<section>}`); required in `xetex` and `pdftex` for counters in general, as well as in `luatex` with `bidi=default`; required in `luatex` for numeric footnote marks  $>9$  with `bidi=basic-r` (but *not* with `bidi=basic`); note, however, it could depend on the counter format.

With counters, `\arabic` is not only considered L text always (with `\babelsublr`, see below), but also an “isolated” block which does not interact with the surrounding chars. So, while `1.2` in R text is rendered in that order with `bidi=basic` (as a decimal number), in `\arabic{c1}.\arabic{c2}` the visual order is `c2.c1`. Of course, you may always adjust the order by changing the language, if necessary.<sup>20</sup>

**lists** required in `xetex` and `pdftex`, but only in bidirectional (with both R and L paragraphs) documents in `luatex`.

**WARNING** As of April 2019 there is a bug with `\parshape` in `luatex` (a `TEX` primitive) which makes lists to be horizontally misplaced if they are inside a `\vbox` (like `minipage`) and the current direction is different from the main one. A workaround is to restore the main language before the box and then set the local one inside.

**contents** required in `xetex` and `pdftex`; in `luatex` toc entries are R by default if the main language is R.

**columns** required in `xetex` and `pdftex` to reverse the column order (currently only the standard two-column mode); in `luatex` they are R by default if the main language is R (including `multicol`).

**footnotes** not required in monolingual documents, but it may be useful in bidirectional documents (with both R and L paragraphs) in all engines; you may use alternatively `\BabelFootnote` described below (what this option does exactly is also explained there).

**captions** is similar to sectioning, but for `\caption`; not required in monolingual documents with `luatex`, but may be required in `xetex` and `pdftex` in some styles (support for the latter two engines is still experimental) **New 3.18** .

**tabular** required in `luatex` for R `tabular` (it has been tested only with simple tables, so expect some readjustments in the future); ignored in `pdftex` or `xetex` (which will not support a similar option in the short term). It patches an internal command, so it might be ignored by some packages and classes (or even raise an error). **New 3.18** .

**graphics** modifies the `picture` environment so that the whole figure is L but the text is R. It *does not* work with the standard `picture`, and `pict2e` is required if you want sloped lines. It attempts to do the same for `pgf/tikz`. Somewhat experimental. **New 3.32** .

**extras** is used for miscellaneous readjustments which do not fit into the previous groups. Currently redefines in `luatex` `\underline` and `\LaTeX2e` **New 3.19** .

**EXAMPLE** Typically, in an Arabic document you would need:

---

<sup>20</sup>Next on the roadmap are counters and numeral systems in general. Expect some minor readjustments.

```
\usepackage[bidi=basic,
            layout=counters.tabular]{babel}
```

**\babelsublr**  $\{\langle lr\text{-}text\rangle\}$

Digits in pdf<sub>tex</sub> must be marked up explicitly (unlike luat<sub>ex</sub> with `bidi=basic` or `bidi=basic-r` and, usually, xet<sub>ex</sub>). This command is provided to set  $\{\langle lr\text{-}text\rangle\}$  in L mode if necessary. It's intended for what Unicode calls weak characters, because words are best set with the corresponding language. For this reason, there is no `rl` counterpart. Any `\babelsublr` in *explicit* L mode is ignored. However, with `bidi=basic` and *implicit* L, it first returns to R and then switches to explicit L. To clarify this point, consider, in an R context:

```
RTL_A_ltr_text_\thechapter{}\and_still_ltr RTL_B
```

There are *three* R blocks and *two* L blocks, and the order is *RTL B and still ltr 1 ltr text RTL A*. This is by design to provide the proper behavior in the most usual cases — but if you need to use `\ref` in an L text inside R, the L text must be marked up explicitly; for example:

```
RTL_A_\foreignlanguage{english}{ltr_text_\thechapter{}\and_still_ltr} RTL_B
```

**\BabelPatchSection**  $\{\langle section\text{-}name\rangle\}$

Mainly for bidi text, but it could be useful in other cases. `\BabelPatchSection` and the corresponding option `layout=sectioning` takes a more logical approach (at least in many cases) because it applies the global language to the section format (including the `\chaptername` in `\chapter`), while the section text is still the current language. The latter is passed to tocs and marks, too, and with `sectioning` in `layout` they both reset the “global” language to the main one, while the text uses the “local” language. With `layout=sectioning` all the standard sectioning commands are redefined (it also “isolates” the page number in heads, for a proper bidi behavior), but with this command you can set them individually if necessary (but note then tocs and marks are not touched).

**\BabelFootnote**  $\{\langle cmd\rangle\}\{\langle local\text{-}language\rangle\}\{\langle before\rangle\}\{\langle after\rangle\}$

**New 3.17** Something like:

```
\BabelFootnote{\parsfootnote}{\language}\{()\}
```

defines `\parsfootnote` so that `\parsfootnote{note}` is equivalent to:

```
\footnote{(\foreignlanguage{\language}\note)}
```

but the footnote itself is typeset in the main language (to unify its direction). In addition, `\parsfootnotetext` is defined. The option `footnotes` just does the following:

```
\BabelFootnote{\footnote}{\language}\{()\}%
\BabelFootnote{\localfootnote}{\language}\{()\}%
\BabelFootnote{\mainfootnote}\{()\}
```

(which also redefine `\footnotetext` and define `\localfootnotetext` and `\mainfootnotetext`). If the language argument is empty, then no language is selected inside the argument of the footnote. Note this command is available always in bidi documents, even without `layout=footnotes`.

**EXAMPLE** If you want to preserve directionality in footnotes and there are many footnotes entirely in English, you can define:

```
\BabelFootnote{\enfootnote}{english}{.}{.}
```

It adds a period outside the English part, so that it is placed at the left in the last line. This means the dot the end of the footnote text should be omitted.

## 1.22 Language attributes

`\languageattribute`

This is a user-level command, to be used in the preamble of a document (after `\usepackage[...]{babel}`), that declares which attributes are to be used for a given language. It takes two arguments: the first is the name of the language; the second, a (list of) attribute(s) to be used. Attributes must be set in the preamble and only once – they cannot be turned on and off. The command checks whether the language is known in this document and whether the attribute(s) are known for this language. Very often, using a *modifier* in a package option is better. Several language definition files use their own methods to set options. For example, french uses `\frenchsetup`, magyar (1.5) uses `\magyarOptions`; modifiers provided by spanish have no attribute counterparts. Macros setting options are also used (eg, `\ProsodicMarksOn` in latin).

## 1.23 Hooks

**New 3.9a** A hook is a piece of code to be executed at certain events. Some hooks are predefined when `luatex` and `xetex` are used.

`\AddBabelHook`

```
[<lang>]{<name>}{<event>}{<code>}
```

The same name can be applied to several events. Hooks may be enabled and disabled for all defined events with `\EnableBabelHook{<name>}`, `\DisableBabelHook{<name>}`.

Names containing the string `babel` are reserved (they are used, for example, by `\useshortands*` to add a hook for the event `afterextras`). **New 3.33** They may be also applied to a specific language with the optional argument; language-specific settings are executed after global ones.

Current events are the following; in some of them you can use one to three  $\text{\TeX}$  parameters (`#1`, `#2`, `#3`), with the meaning given:

**addialect** (language name, dialect name) Used by `luababel.def` to load the patterns if not preloaded.

**patterns** (language name, language with encoding) Executed just after the `\language` has been set. The second argument has the patterns name actually selected (in the form of either `lang:ENC` or `lang`).

**hyphenation** (language name, language with encoding) Executed locally just before exceptions given in `\babelhyphenation` are actually set.

**defaultcommands** Used (locally) in `\StartBabelCommands`.

**encodedcommands** (input, font encodings) Used (locally) in `\StartBabelCommands`. Both `xetex` and `luatex` make sure the encoded text is read correctly.

**stopcommands** Used to reset the above, if necessary.

**write** This event comes just after the switching commands are written to the aux file.  
**beforeextras** Just before executing `\extras⟨language⟩`. This event and the next one should not contain language-dependent code (for that, add it to `\extras⟨language⟩`).  
**afterextras** Just after executing `\extras⟨language⟩`. For example, the following deactivates shorthands in all languages:

```
\AddBabelHook{noshort}{afterextras}{\languageshorthands{none}}
```

**stringprocess** Instead of a parameter, you can manipulate the macro `\BabelString` containing the string to be defined with `\SetString`. For example, to use an expanded version of the string in the definition, write:

```
\AddBabelHook{myhook}{stringprocess}{%  
  \protected@edef\BabelString{\BabelString}}
```

**initiateactive** (char as active, char as other, original char) **New 3.9i** Executed just after a shorthand has been ‘initiated’. The three parameters are the same character with different catcodes: active, other (`\string’ed`) and the original one.

**afterreset** **New 3.9i** Executed when selecting a language just after `\originalTeX` is run and reset to its base value, before executing `\captions⟨language⟩` and `\date⟨language⟩`.

Four events are used in `hyphen.cfg`, which are handled in a quite different way for efficiency reasons – unlike the precedent ones, they only have a single hook and replace a default definition.

**everylanguage** (language) Executed before every language patterns are loaded.

**loadkernel** (file) By default loads `switch.def`. It can be used to load a different version of this file or to load nothing.

**loadpatterns** (patterns file) Loads the patterns file. Used by `luababel.def`.

**loadexceptions** (exceptions file) Loads the exceptions file. Used by `luababel.def`.

**\BabelContentsFiles** **New 3.9a** This macro contains a list of “toc” types requiring a command to switch the language. Its default value is `toc, lof, lot`, but you may redefine it with `\renewcommand` (it’s up to you to make sure no toc type is duplicated).

## 1.24 Languages supported by babel with ldf files

In the following table most of the languages supported by babel with and `.ldf` file are listed, together with the names of the option which you can load babel with for each language. Note this list is open and the current options may be different. It does not include ini files.

**Afrikaans** afrikaans  
**Azerbaijani** azerbaijani  
**Basque** basque  
**Breton** breton  
**Bulgarian** bulgarian  
**Catalan** catalan  
**Croatian** croatian  
**Czech** czech  
**Danish** danish  
**Dutch** dutch

**English** english, USenglish, american, UKenglish, british, canadian, australian, newzealand  
**Esperanto** esperanto  
**Estonian** estonian  
**Finnish** finnish  
**French** french, francais, canadien, acadian  
**Galician** galician  
**German** austrian, german, germanb, ngerman, naustrian  
**Greek** greek, polutonikogreek  
**Hebrew** hebrew  
**Icelandic** icelandic  
**Indonesian** bahasa, indonesian, indon, bahasai  
**Interlingua** interlingua  
**Irish Gaelic** irish  
**Italian** italian  
**Latin** latin  
**Lower Sorbian** lowersorbian  
**Malay** bahasam, malay, melayu  
**North Sami** samin  
**Norwegian** norsk, nynorsk  
**Polish** polish  
**Portuguese** portuges, portuguese, brazilian, brazil  
**Romanian** romanian  
**Russian** russian  
**Scottish Gaelic** scottish  
**Spanish** spanish  
**Slovakian** slovak  
**Slovenian** slovene  
**Swedish** swedish  
**Serbian** serbian  
**Turkish** turkish  
**Ukrainian** ukrainian  
**Upper Sorbian** uppersorbian  
**Welsh** welsh

There are more languages not listed above, including hindi, thai, thaicjk, latvian, turkmen, magyar, mongolian, romansh, lithuanian, spanglish, vietnamese, japanese, pinyin, arabic, farsi, ibygreek, bgreek, serbianc, frenchle, ethiop and friulan.

Most of them work out of the box, but some may require extra fonts, encoding files, a preprocessor or even a complete framework (like CJK or luatexja). For example, if you have got the velthuis/devnag package, you can create a file with extension .dn:

```

\documentclass{article}
\usepackage[hindi]{babel}
\begin{document}
{\dn_devaanaa.m_priya.h}
\end{document}

```

Then you preprocess it with devnag  $\langle file \rangle$ , which creates  $\langle file \rangle$ .tex; you can then typeset the latter with  $\LaTeX$ .

## 1.25 Unicode character properties in luatex

**New 3.32** Part of the babel job is to apply Unicode rules to some script-specific features based on some properties. Currently, they are 3, namely, direction (ie, bidi class), mirroring



glyphs, and line breaking for CJK scripts. These properties are stored in lua tables, which you can modify with the following macro (for example, to set them for glyphs in the PUA).

`\babelcharproperty`  $\{\langle char-code \rangle\}[\langle to-char-code \rangle]\{\langle property \rangle\}\{\langle value \rangle\}$

**New 3.32** Here,  $\{\langle char-code \rangle\}$  is a number (with  $\TeX$  syntax). With the optional argument, you can set a range of values. There are three properties (with a short name, taken from Unicode): direction (bc), mirror (bmg), linebreak (lb). The settings are global. For example:

```
\babelcharproperty{`z}{mirror}{`?}
\babelcharproperty{`-}{direction}{l}%_or_al,_r,_en,_an,_on,_et,_cs
\babelcharproperty{`)}{linebreak}{cl}%_or_id,_op,_cl,_ns,_ex,_in,_hy
```

This command is allowed only in vertical mode (the preamble or between paragraphs).

## 1.26 Tweaking some features

`\babeladjust`  $\{\langle key-value-list \rangle\}$

**New 3.36** Sometimes you might need to disable some babel features. Currently this macro understands the following keys (and only for  $\text{luatex}$ ), with values on or off: `bidi.text`, `bidi.mirroring`, `bidi.mapdigits`, `layout.lists`, `layout.tabular`, `linebreak.sea`, `linebreak.cjk`. For example, you can set `\babeladjust{bidi.text=off}` if you are using an alternative algorithm or with large sections not requiring it. With  $\text{luahbtex}$  you may need `bidi.mirroring=off`. Use with care, because these options do not deactivate other related options (like paragraph direction with `bidi.text`).

## 1.27 Tips, workarounds, known issues and notes

- If you use the document class *book* and you use `\ref` inside the argument of `\chapter` (or just use `\ref` inside `\MakeUppercase`),  $\text{\LaTeX}$  will keep complaining about an undefined label. To prevent such problems, you could revert to using uppercase labels, you can use `\lowercase{\ref{foo}}` inside the argument of `\chapter`, or, if you will not use shorthands in labels, set the safe option to `none` or `bib`.
- Both  $\text{ltxdoc}$  and `babel` use `\AtBeginDocument` to change some catcodes, and `babel` reloads `hline` to make sure `:` has the right one, so if you want to change the catcode of `|` it has to be done using the same method at the proper place, with

```
\AtBeginDocument{\DeleteShortVerb{\|}}
```

*before* loading `babel`. This way, when the document begins the sequence is (1) make `|` active ( $\text{ltxdoc}$ ); (2) make it unactive (your settings); (3) make `babel` shorthands active (`babel`); (4) reload `hline` (`babel`, now with the correct catcodes for `|` and `:`).

- Documents with several input encodings are not frequent, but sometimes are useful. You can set different encodings for different languages as the following example shows:

```
\addto\extrasfrench{\inputencoding{latin1}}
\addto\extrarussian{\inputencoding{koi8-r}}
```

(A recent version of `inputenc` is required.)

- For the hyphenation to work correctly, lccodes cannot change, because T<sub>E</sub>X only takes into account the values when the paragraph is hyphenated, i.e., when it has been finished.<sup>21</sup> So, if you write a chunk of French text with `\foreignlanguage`, the apostrophes might not be taken into account. This is a limitation of T<sub>E</sub>X, not of babel. Alternatively, you may use `\usesorthands` to activate ' and `\definesorthand`, or redefine `\textquoteright` (the latter is called by the non-ASCII right quote).
- `\bibitem` is out of sync with `\selectlanguage` in the .aux file. The reason is `\bibitem` uses `\immediate` (and others, in fact), while `\selectlanguage` doesn't. There is no known workaround.
- Babel does not take into account `\normalsfcodes` and (non-)French spacing is not always properly (un)set by languages. However, problems are unlikely to happen and therefore this part remains untouched in version 3.9 (but it is in the 'to do' list).
- Using a character mathematically active (ie, with math code "8000) as a shorthand can make T<sub>E</sub>X enter in an infinite loop in some rare cases. (Another issue in the 'to do' list, although there is a partial solution.)

The following packages can be useful, too (the list is still far from complete):

**csquotes** Logical markup for quotes.

**iflang** Tests correctly the current language.

**hyphsubst** Selects a different set of patterns for a language.

**translator** An open platform for packages that need to be localized.

**siunitx** Typesetting of numbers and physical quantities.

**biblatex** Programmable bibliographies and citations.

**bicaption** Bilingual captions.

**babelbib** Multilingual bibliographies.

**microtype** Adjusts the typesetting according to some languages (kerning and spacing).

Ligatures can be disabled.

**substitutefont** Combines fonts in several encodings.

**mkpattern** Generates hyphenation patterns.

**tracklang** Tracks which languages have been requested.

**ucharclasses** (xetex) Switches fonts when you switch from one Unicode block to another.

**zhspacing** Spacing for CJK documents in xetex.

## 1.28 Current and future work

The current work is focused on the so-called complex scripts in luatex. In 8-bit engines, babel provided a basic support for bidi text as part of the style for Hebrew, but it is somewhat unsatisfactory and internally replaces some hardwired commands by other hardwired commands (generic changes would be much better).

Useful additions would be, for example, time, currency, addresses and personal names.<sup>22</sup> But that is the easy part, because they don't require modifying the L<sup>A</sup>T<sub>E</sub>X internals.

Calendars (Arabic, Persian, Indic, etc.) are under study.

Also interesting are differences in the sentence structure or related to it. For example, in Basque the number precedes the name (including chapters), in Hungarian "from (1)" is "(1)-ből", but "from (3)" is "(3)-ből", in Spanish an item labelled "3.<sup>o</sup>" may be referred to as either "ítem 3.<sup>o</sup>" or "3.<sup>er</sup> ítem", and so on.

<sup>21</sup>This explains why L<sup>A</sup>T<sub>E</sub>X assumes the lowercase mapping of T1 and does not provide a tool for multiple mappings. Unfortunately, `\savingsphcodes` is not a solution either, because lccodes for hyphenation are frozen in the format and cannot be changed.

<sup>22</sup>See for example POSIX, ISO 14652 and the Unicode Common Locale Data Repository (CLDR). Those systems, however, have limited application to T<sub>E</sub>X because their aim is just to display information and not fine typesetting.

An option to manage bidirectional document layout in luatex (lists, footnotes, etc.) is almost finished, but xetex required more work. Unfortunately, proper support for xetex requires patching somehow lots of macros and packages (and some issues related to \specials remain, like color and hyperlinks), so babel resorts to the bidi package (by Vafa Khalighi). See the babel repository for a small example (xe-bidi).

## 1.29 Tentative and experimental code

See the code section for \foreignlanguage\* (a new starred version of \foreignlanguage).

### Old stuff

A couple of tentative macros were provided by babel ( $\geq 3.9g$ ) with a partial solution for “Unicode” fonts. These macros are now deprecated — use \babelfont. A short description follows, for reference:

- \babelFSstore{*babel-language*} sets the current three basic families (rm, sf, tt) as the default for the language given.
- \babelFSdefault{*babel-language*}{*fontspec-features*} patches \fontspec so that the given features are always passed as the optional argument or added to it (not an ideal solution).

So, for example:

```
\setmainfont[Language=Turkish]{Minion_Pro}
\babelFSstore{turkish}
\setmainfont{Minion_Pro}
\babelFSfeatures{turkish}{Language=Turkish}
```

### Modifying values of ini files

**New 3.36** There is a way to modify the values of ini files when they get loaded with \babelprovide. To set, say, digits.native in the numbers section, use something like numbers..digits.native=abcdefghijklm (note the double dot between the section and the key name). The syntax may change, and currently it only redefines existing keys.

## 2 Loading languages with language.dat

TeX and most engines based on it (pdfTeX, xetex,  $\epsilon$ -TeX, the main exception being luatex) require hyphenation patterns to be preloaded when a format is created (eg,  $\LaTeX$ , Xe $\LaTeX$ , pdf $\LaTeX$ ). babel provides a tool which has become standard in many distributions and based on a “configuration file” named language.dat. The exact way this file is used depends on the distribution, so please, read the documentation for the latter (note also some distributions generate the file with some tool).

**New 3.9q** With luatex, however, patterns are loaded on the fly when requested by the language (except the “0th” language, typically english, which is preloaded always).<sup>23</sup> Until 3.9n, this task was delegated to the package luatex-hyphen, by Khaled Hosny, Élie Roux, and Manuel Pégourié-Gonnard, and required an extra file named language.dat.lua, but now a new mechanism has been devised based solely on language.dat. **You must rebuild the formats** if upgrading from a previous version. You may want to have a local language.dat for a particular project (for example, a book on Chemistry).<sup>24</sup>

<sup>23</sup>This feature was added to 3.9o, but it was buggy. Both 3.9o and 3.9p are deprecated.

<sup>24</sup>The loader for lua(e)tex is slightly different as it’s not based on babel but on etex.src. Until 3.9p it just didn’t work, but thanks to the new code it works by reloading the data in the babel way, i.e., with language.dat.

## 2.1 Format

In that file the person who maintains a T<sub>E</sub>X environment has to record for which languages he has hyphenation patterns *and* in which files these are stored<sup>25</sup>. When hyphenation exceptions are stored in a separate file this can be indicated by naming that file *after* the file with the hyphenation patterns.

The file can contain empty lines and comments, as well as lines which start with an equals (=) sign. Such a line will instruct L<sup>A</sup>T<sub>E</sub>X that the hyphenation patterns just processed have to be known under an alternative name. Here is an example:

```
%_File_:_language.dat
%_Purpose_:_tell_!niTeX_what_files_with_patterns_to_load.
english_english.hyphenations
=british

dutch_hyphen.dutch_exceptions.dutch_%_Nederlands
german_hyphen.ger
```

You may also set the font encoding the patterns are intended for by following the language name by a colon and the encoding code.<sup>26</sup> For example:

```
german:T1_hyphenT1.ger
german_hyphen.ger
```

With the previous settings, if the encoding when the language is selected is T1 then the patterns in hyphenT1.ger are used, but otherwise use those in hyphen.ger (note the encoding could be set in `\extras{lang}`).

A typical error when using babel is the following:

```
No_hyphenation_patterns_were_preloaded_for
the_language`<lang>'into_the_format.
Please,_configure_your_TeX_system_to_add_them_and
rebuild_the_format._Now_I_will_use_the_patterns
preloaded_for_english_instead}}
```

It simply means you must reconfigure language.dat, either by hand or with the tools provided by your distribution.

## 3 The interface between the core of babel and the language definition files

The *language definition files* (ldf) must conform to a number of conventions, because these files have to fill in the gaps left by the common code in babel.def, i. e., the definitions of the macros that produce texts. Also the language-switching possibility which has been built into the babel system has its implications.

The following assumptions are made:

- Some of the language-specific definitions might be used by plain T<sub>E</sub>X users, so the files have to be coded so that they can be read by both L<sup>A</sup>T<sub>E</sub>X and plain T<sub>E</sub>X. The current format can be checked by looking at the value of the macro `\fmtname`.

<sup>25</sup>This is because different operating systems sometimes use very different file-naming conventions.

<sup>26</sup>This is not a new feature, but in former versions it didn't work correctly.

- The common part of the babel system redefines a number of macros and environments (defined previously in the document style) to put in the names of macros that replace the previously hard-wired texts. These macros have to be defined in the language definition files.
- The language definition files must define five macros, used to activate and deactivate the language-specific definitions. These macros are `\langle lang \rangle hyphenmins`, `\captions\langle lang \rangle`, `\date\langle lang \rangle`, `\extras\langle lang \rangle` and `\noextras\langle lang \rangle` (the last two may be left empty); where `\langle lang \rangle` is either the name of the language definition file or the name of the `\LaTeX` option that is to be used. These macros and their functions are discussed below. You must define all or none for a language (or a dialect); defining, say, `\date\langle lang \rangle` but not `\captions\langle lang \rangle` does not raise an error but can lead to unexpected results.
- When a language definition file is loaded, it can define `\l@\langle lang \rangle` to be a dialect of `\language0` when `\l@\langle lang \rangle` is undefined.
- Language names must be all lowercase. If an unknown language is selected, babel will attempt setting it after lowercasing its name.
- The semantics of modifiers is not defined (on purpose). In most cases, they will just be simple separated options (eg, spanish), but a language might require, say, a set of options organized as a tree with suboptions (in such a case, the recommended separator is `/`).

Some recommendations:

- The preferred shorthand is `"`, which is not used in `\LaTeX` (quotes are entered as ``` and `'`). Other good choices are characters which are not used in a certain context (eg, `=` in an ancient language). Note however `=`, `<`, `>`, `:` and the like can be dangerous, because they may be used as part of the syntax of some elements (numeric expressions, key/value pairs, etc.).
- Captions should not contain shorthands or encoding-dependent commands (the latter is not always possible, but should be clearly documented). They should be defined using the LICR. You may also use the new tools for encoded strings, described below.
- Avoid adding things to `\noextras\langle lang \rangle` except for `umlauthigh` and `friends`, `\bbl@deactivate`, `\bbl@(non)frenchspacing`, and language-specific macros. Use always, if possible, `\bbl@save` and `\bbl@savevariable` (except if you still want to have access to the previous value). Do not reset a macro or a setting to a hardcoded value. Never. Instead save its value in `\extras\langle lang \rangle`.
- Do not switch scripts. If you want to make sure a set of glyphs is used, switch either the font encoding (low-level) or the language (high-level, which in turn may switch the font encoding). Usage of things like `\latintext` is deprecated.<sup>27</sup>
- Please, for “private” internal macros do not use the `\bbl@` prefix. It is used by babel and it can lead to incompatibilities.

There are no special requirements for documenting your language files. Now they are not included in the base babel manual, so provide a standalone document suited for your needs, as well as other files you think can be useful. A PDF and a “readme” are strongly recommended.

---

<sup>27</sup> But not removed, for backward compatibility.

### 3.1 Guidelines for contributed languages

Now language files are “outsourced” and are located in a separate directory (/macros/latex/contrib/babel-contrib), so that they are contributed directly to CTAN (please, do not send to me language styles just to upload them to CTAN).

Of course, placing your style files in this directory is not mandatory, but if you want to do it, here are a few guidelines.

- Do not hesitate stating on the file heads you are the author and the maintainer, if you actually are. There is no need to state the babel maintainer(s) as authors if they have not contributed significantly to your language files.
- Fonts are not strictly part of a language, so they are best placed in the corresponding TeX tree. This includes not only tfm, vf, ps1, ot f, mf files and the like, but also fd ones.
- Font and input encodings are usually best placed in the corresponding tree, too, but sometimes they belong more naturally to the babel style. Note you may also need to define a LICR.
- Babel ldf files may just interface a framework, as it happens often with Oriental languages/scripts. This framework is best placed in its own directory.

The following page provides a starting point: <http://www.texnia.com/incubator.html>. If you need further assistance and technical advice in the development of language styles, I am willing to help you. And of course, you can make any suggestion you like.

### 3.2 Basic macros

In the core of the babel system, several macros are defined for use in language definition files. Their purpose is to make a new language known. The first two are related to hyphenation patterns.

`\addlanguage` The macro `\addlanguage` is a non-outer version of the macro `\newlanguage`, defined in `plain.tex` version 3.x. For older versions of `plain.tex` and `lplain.tex` a substitute definition is used. Here “language” is used in the TeX sense of set of hyphenation patterns.

`\adddialect` The macro `\adddialect` can be used when two languages can (or must) use the same hyphenation patterns. This can also be useful for languages for which no patterns are preloaded in the format. In such cases the default behavior of the babel system is to define this language as a ‘dialect’ of the language for which the patterns were loaded as `\language0`. Here “language” is used in the TeX sense of set of hyphenation patterns.

`\<lang>hyphenmins` The macro `\<lang>hyphenmins` is used to store the values of the `\lefthyphenmin` and `\righthyphenmin`. Redefine this macro to set your own values, with two numbers corresponding to these two parameters. For example:

```
\renewcommand\spanishhyphenmins{34}
```

(Assigning `\lefthyphenmin` and `\righthyphenmin` directly in `\extras<lang>` has no effect.)

`\providehyphenmins` The macro `\providehyphenmins` should be used in the language definition files to set `\lefthyphenmin` and `\righthyphenmin`. This macro will check whether these parameters were provided by the hyphenation file before it takes any action. If these values have been already set, this command is ignored (currently, default pattern files do *not* set them).

`\captions<lang>` The macro `\captions<lang>` defines the macros that hold the texts to replace the original hard-wired texts.

`\date<lang>` The macro `\date<lang>` defines `\today`.

`\extras<lang>` The macro `\extras<lang>` contains all the extra definitions needed for a specific language.

|                                    |   |
|------------------------------------|---|
|                                    | This macro, like the following, is a hook – you can add things to it, but it must not be used directly.   |
| <code>\noextras&lt;lang&gt;</code> | Because we want to let the user switch between languages, but we do not know what state $\TeX$ might be in after the execution of <code>\extras&lt;lang&gt;</code> , a macro that brings $\TeX$ into a predefined state is needed. It will be no surprise that the name of this macro is <code>\noextras&lt;lang&gt;</code> .                                       |
| <code>\bbl@declare@ttribute</code> | This is a command to be used in the language definition files for declaring a language attribute. It takes three arguments: the name of the language, the attribute to be defined, and the code to be executed when the attribute is to be used.  |
| <code>\main@language</code>        | To postpone the activation of the definitions needed for a language until the beginning of a document, all language definition files should use <code>\main@language</code> instead of <code>\selectlanguage</code> . This will just store the name of the language, and the proper language will be activated at the start of the document.                        |
| <code>\ProvidesLanguage</code>     | The macro <code>\ProvidesLanguage</code> should be used to identify the language definition files. Its syntax is similar to the syntax of the $\LaTeX$ command <code>\ProvidesPackage</code> .  |
| <code>\LdfInit</code>              | The macro <code>\LdfInit</code> performs a couple of standard checks that must be made at the beginning of a language definition file, such as checking the category code of the <code>@</code> -sign, preventing the <code>.ldf</code> file from being processed twice, etc.   |
| <code>\ldf@quit</code>             | The macro <code>\ldf@quit</code> does work needed if a <code>.ldf</code> file was processed earlier. This includes resetting the category code of the <code>@</code> -sign, preparing the language to be activated at <code>\begin{document}</code> time, and ending the input stream.  |
| <code>\ldf@finish</code>           | The macro <code>\ldf@finish</code> does work needed at the end of each <code>.ldf</code> file. This includes resetting the category code of the <code>@</code> -sign, loading a local configuration file, and preparing the language to be activated at <code>\begin{document}</code> time.   |
| <code>\loadlocalcfg</code>         | After processing a language definition file, $\LaTeX$ can be instructed to load a local configuration file. This file can, for instance, be used to add strings to <code>\captions&lt;lang&gt;</code> to support local document classes. The user will be informed that this configuration file has been loaded. This macro is called by <code>\ldf@finish</code> . |
| <code>\substitutefontfamily</code> | (Deprecated.) This command takes three arguments, a font encoding and two font family names. It creates a font description file for the first font in the given encoding. This <code>.fd</code> file will instruct $\LaTeX$ to use a font from the second family when a font from the first family in the given encoding seems to be needed.                        |

### 3.3 Skeleton

Here is the basic structure of an `ldf` file, with a language, a dialect and an attribute. Strings are best defined using the method explained in sec. 3.8 (babel 3.9 and later).

```

\ProvidesLanguage{<language>}
\l[2016/04/23]v0.0<Language>_support_from_the_babel_system]
\LdfInit{<language>}{captions<language>}

\ifx\undefined\l@<language>
\l\@nopatterns{<Language>}
\l\adddialect\l@<language>0
\fi

\adddialect\l@<dialect>\l@<language>

\bbl@declare@ttribute{<language>}{<attrib>}{%
\l\expandafter\addto\expandafter\extras<language>
\l\expandafter{\extras<attrib><language>}%
\l\let\captions<language>\captions<attrib><language>}
```

```

\providehyphenmins{<language>}{\tw@\thr@@}

\StartBabelCommands*{<language>}{captions}
\SetString\chaptername{<chapter_name>}
%_More_strings

\StartBabelCommands*{<language>}{date}
\SetString\monthname{<name_of_first_month>}
%_More_strings

\StartBabelCommands*{<dialect>}{captions}
\SetString\chaptername{<chapter_name>}
%_More_strings

\StartBabelCommands*{<dialect>}{date}
\SetString\monthname{<name_of_first_month>}
%_More_strings

\EndBabelCommands

\addto\extras<language>{}
\addto\noextras<language>{}
\let\extras<dialect>\extras<language>
\let\noextras<dialect>\noextras<language>

\ldf@finish{<language>}

```

**NOTE** If for some reason you want to load a package in your style, you should be aware it cannot be done directly in the ldf file, but it can be delayed with `\AtEndOfPackage`. Macros from external packages can be used *inside* definitions in the ldf itself (for example, `\extras<language>`), but if executed directly, the code must be placed inside `\AtEndOfPackage`. A trivial example illustrating these points is:

```

\AtEndOfPackage{%
  \RequirePackage{dingbat}%_Delay_package
  \savebox{\myeye}{\eye}}%_And_direct_usage
\newsavebox{\myeye}
\newcommand\myanchor{\anchor}%_But_OK_inside_command

```

### 3.4 Support for active characters

In quite a number of language definition files, active characters are introduced. To facilitate this, some support macros are provided.

`\initiate@active@char`

The internal macro `\initiate@active@char` is used in language definition files to instruct  $\TeX$  to give a character the category code ‘active’. When a character has been made active it will remain that way until the end of the document. Its definition may vary.

`\bbl@activate`  
`\bbl@deactivate`

The command `\bbl@activate` is used to change the way an active character expands. `\bbl@activate` ‘switches on’ the active behavior of the character. `\bbl@deactivate` lets the active character expand to its former (mostly) non-active self.

`\declare@shorthand`

The macro `\declare@shorthand` is used to define the various shorthands. It takes three arguments: the name for the collection of shorthands this definition belongs to; the character (sequence) that makes up the shorthand, i.e. `~` or `"a`; and the code to be executed when the shorthand is encountered. (It does *not* raise an error if the shorthand character has not been “initiated”.)



`\bbl@add@special`  
`\bbl@remove@special`

The T<sub>E</sub>Xbook states: “Plain T<sub>E</sub>X includes a macro called `\dospecials` that is essentially a set macro, representing the set of all characters that have a special category code.” [2, p. 380] It is used to set text ‘verbatim’. To make this work if more characters get a special category code, you have to add this character to the macro `\dospecials`. L<sup>A</sup>T<sub>E</sub>X adds another macro called `\@sanitize` representing the same character set, but without the curly braces. The macros `\bbl@add@special⟨char⟩` and `\bbl@remove@special⟨char⟩` add and remove the character `⟨char⟩` to these two sets.

### 3.5 Support for saving macro definitions

Language definition files may want to *redefine* macros that already exist. Therefore a mechanism for saving (and restoring) the original definition of those macros is provided. We provide two macros for this<sup>28</sup>.

`\babel@save`

To save the current meaning of any control sequence, the macro `\babel@save` is provided. It takes one argument, `⟨cname⟩`, the control sequence for which the meaning has to be saved.

`\babel@savevariable`

A second macro is provided to save the current value of a variable. In this context, anything that is allowed after the `\the` primitive is considered to be a variable. The macro takes one argument, the `⟨variable⟩`.

The effect of the preceding macros is to append a piece of code to the current definition of `\originalTeX`. When `\originalTeX` is expanded, this code restores the previous definition of the control sequence or the previous value of the variable.

### 3.6 Support for extending macros

`\addto`

The macro `\addto{⟨control sequence⟩}{⟨TEX code⟩}` can be used to extend the definition of a macro. The macro need not be defined (ie, it can be undefined or `\relax`). This macro can, for instance, be used in adding instructions to a macro like `\extrasenglish`. Be careful when using this macro, because depending on the case the assignment could be either global (usually) or local (sometimes). That does not seem very consistent, but this behavior is preserved for backward compatibility. If you are using `etoolbox`, by Philipp Lehman, consider using the tools provided by this package instead of `\addto`.

### 3.7 Macros common to a number of languages

`\bbl@allowhyphens`

In several languages compound words are used. This means that when T<sub>E</sub>X has to hyphenate such a compound word, it only does so at the ‘-’ that is used in such words. To allow hyphenation in the rest of such a compound word, the macro `\bbl@allowhyphens` can be used.

`\allowhyphens`

Same as `\bbl@allowhyphens`, but does nothing if the encoding is T1. It is intended mainly for characters provided as real glyphs by this encoding but constructed with `\accent` in OT1.

Note the previous command (`\bbl@allowhyphens`) has different applications (hyphens and discretionaries) than this one (composite chars). Note also prior to version 3.7, `\allowhyphens` had the behavior of `\bbl@allowhyphens`.

`\set@low@box`

For some languages, quotes need to be lowered to the baseline. For this purpose the macro `\set@low@box` is available. It takes one argument and puts that argument in an `\hbox`, at the baseline. The result is available in `\box0` for further processing.

`\save@sfc@q`

Sometimes it is necessary to preserve the `\spacefactor`. For this purpose the macro `\save@sfc@q` is available. It takes one argument, saves the current spacefactor, executes the argument, and restores the spacefactor.

`\bbl@frenchspacing`  
`\bbl@nonfrenchspacing`

The commands `\bbl@frenchspacing` and `\bbl@nonfrenchspacing` can be used to properly switch French spacing on and off.

<sup>28</sup>This mechanism was introduced by Bernd Raichle.

### 3.8 Encoding-dependent strings

**New 3.9a** Babel 3.9 provides a way of defining strings in several encodings, intended mainly for luatex and xetex. This is the only new feature requiring changes in language files if you want to make use of it.

Furthermore, it must be activated explicitly, with the package option `strings`. If there is no `strings`, these blocks are ignored, except `\SetCases` (and except if forced as described below). In other words, the old way of defining/switching strings still works and it's used by default.

It consists of a series of blocks started with `\StartBabelCommands`. The last block is closed with `\EndBabelCommands`. Each block is a single group (ie, local declarations apply until the next `\StartBabelCommands` or `\EndBabelCommands`). An ldf may contain several series of this kind.

Thanks to this new feature, string values and string language switching are not mixed any more. No need of `\addto`. If the language is french, just redefine `\frenchchaptername`.

`\StartBabelCommands`  $\langle\textit{language-list}\rangle\{\langle\textit{category}\rangle\}[\langle\textit{selector}\rangle]$

The  $\langle\textit{language-list}\rangle$  specifies which languages the block is intended for. A block is taken into account only if the `\CurrentOption` is listed here. Alternatively, you can define `\BabelLanguages` to a comma-separated list of languages to be defined (if undefined, `\StartBabelCommands` sets it to `\CurrentOption`). You may write `\CurrentOption` as the language, but this is discouraged – a explicit name (or names) is much better and clearer. A “selector” is a name to be used as value in package option `strings`, optionally followed by extra info about the encodings to be used. The name `unicode` must be used for xetex and luatex (the key `strings` has also other two special values: `generic` and `encoded`). If a string is set several times (because several blocks are read), the first one takes precedence (ie, it works much like `\providecommand`).

Encoding info is `charset=` followed by a charset, which if given sets how the strings should be translated to the internal representation used by the engine, typically `utf8`, which is the only value supported currently (default is no translations). Note `charset` is applied by luatex and xetex when reading the file, not when the macro or string is used in the document.

A list of font encodings which the strings are expected to work with can be given after `fontenc=` (separated with spaces, if two or more) – recommended, but not mandatory, although blocks without this key are not taken into account if you have requested `strings=encoded`.

Blocks without a selector are read always if the key `strings` has been used. They provide fallback values, and therefore must be the last blocks; they should be provided always if possible and all strings should be defined somehow inside it; they can be the only blocks (mainly LGC scripts using the LICR). Blocks without a selector can be activated explicitly with `strings=generic` (no block is taken into account except those). With `strings=encoded`, strings in those blocks are set as default (internally, `?`). With `strings=encoded` strings are protected, but they are correctly expanded in `\MakeUppercase` and the like. If there is no key `strings`, string definitions are ignored, but `\SetCases` are still honored (in an encoded way).

The  $\langle\textit{category}\rangle$  is either `captions`, `date` or `extras`. You must stick to these three categories, even if no error is raised when using other name.<sup>29</sup> It may be empty, too, but in such a case using `\SetString` is an error (but not `\SetCase`).

```
\StartBabelCommands{language}{captions}
  \[unicode,fontenc=TU_EU1_EU2,charset=utf8]
```

<sup>29</sup>In future releases further categories may be added.

```

\SetString{\chaptername}{utf8-string}

\StartBabelCommands{language}{captions}
\SetString{\chaptername}{ascii-maybe-LICR-string}

\EndBabelCommands

```

A real example is:

```

\StartBabelCommands{austrian}{date}
  \[unicode,fontenc=TU,EU1,EU2,charset=utf8]
  \SetString\monthiname{Jänner}

\StartBabelCommands{german,austrian}{date}
  \[unicode,fontenc=TU,EU1,EU2,charset=utf8]
  \SetString\monthiiname{März}

\StartBabelCommands{austrian}{date}
  \SetString\monthiname{J\{a}nner}

\StartBabelCommands{german}{date}
  \SetString\monthiname{Januar}


\StartBabelCommands{german,austrian}{date}
  \SetString\monthiiname{Februar}
  \SetString\monthiiname{M\{a}rz}
  \SetString\monthivname{April}
  \SetString\monthvname{Mai}
  \SetString\monthviname{Juni}
  \SetString\monthviiname{Juli}
  \SetString\monthviiiname{August}
  \SetString\monthixname{September}
  \SetString\monthxname{Oktober}
  \SetString\monthxiname{November}
  \SetString\monthxiiname{Dezenber}
  \SetString\today{\number\day.~%
    \csname\month\romannumeral\month\name\endcsname\space
    \number\year}

\StartBabelCommands{german,austrian}{captions}
  \SetString\prefacename{Vorwort}
  \[etc.]

\EndBabelCommands

```

When used in ldf files, previous values of  $\langle category \rangle \langle language \rangle$  are overridden, which means the old way to define strings still works and used by default (to be precise, is first set to undefined and then strings are added). However, when used in the preamble or in a package, new settings are added to the previous ones, if the language exists (in the babel sense, ie, if  $\backslash date \langle language \rangle$  exists).

**$\backslash StartBabelCommands$**    $\{ \langle language-list \rangle \} \{ \langle category \rangle \} [ \langle selector \rangle ]$

The starred version just forces strings to take a value – if not set as package option, then the default for the engine is used. This is not done by default to prevent backward incompatibilities, but if you are creating a new language this version is better. It's up to the

maintainers of the current languages to decide if using it is appropriate.<sup>30</sup>

**\EndBabelCommands** Marks the end of the series of blocks.

**\AfterBabelCommands**  $\{\langle code \rangle\}$   
The code is delayed and executed at the global scope just after \EndBabelCommands.

**\SetString**  $\{\langle macro-name \rangle\}\{\langle string \rangle\}$   
Adds  $\langle macro-name \rangle$  to the current category, and defines globally  $\langle lang-macro-name \rangle$  to  $\langle code \rangle$  (after applying the transformation corresponding to the current charset or defined with the hook stringprocess).  
Use this command to define strings, without including any “logic” if possible, which should be a separated macro. See the example above for the date.

**\SetStringLoop**  $\{\langle macro-name \rangle\}\{\langle string-list \rangle\}$   
A convenient way to define several ordered names at once. For example, to define \abmoniname, \abmoniiname, etc. (and similarly with abday):

```
\SetStringLoop{abmon#1name}{en,fb,mr,ab,my,jn,jl,ag,sp,oc,nv,dc}  
\SetStringLoop{abday#1name}{lu,ma,mi,ju,vi,sa,do}
```

#1 is replaced by the roman numeral.

**\SetCase**  $[\langle map-list \rangle]\{\langle toupper-code \rangle\}\{\langle tolower-code \rangle\}$   
Sets globally code to be executed at \MakeUppercase and \MakeLowercase. The code would typically be things like \let\BB\bb and \uccode or \lccode (although for the reasons explained above, changes in lc/uc codes may not work). A  $\langle map-list \rangle$  is a series of macros using the internal format of \@uclclist (eg, \bb\BB\cc\CC). The mandatory arguments take precedence over the optional one. This command, unlike \SetString, is executed always (even without strings), and it is intended for minor readjustments only. For example, as T1 is the default case mapping in L<sup>A</sup>T<sub>E</sub>X, we could set for Turkish:

```
\StartBabelCommands{turkish}{}[ot1enc,fontenc=OT1]  
\SetCase  
  \uccode"10=`I\relax  
  \lccode`I="10\relax  
  
\StartBabelCommands{turkish}{}[unicode,fontenc=TU_EU1_EU2,charset=utf8]  
\SetCase  
  \uccode`i=`İ\relax  
  \uccode`ı=`I\relax  
  \lccode`İ=`i\relax  
  \lccode`I=`ı\relax  
  
\StartBabelCommands{turkish}{}  
\SetCase  
  \uccode`i="9D\relax  
  \uccode"19=`I\relax  
  \lccode"9D=`i\relax  
  \lccode`I="19\relax
```

<sup>30</sup>This replaces in 3.9g a short-lived \UseStrings which has been removed because it did not work.

```
\EndBabelCommands
```

(Note the mapping for OT1 is not complete.)

`\SetHyphenMap` `{\to-lower-macros}`

**New 3.9g** Case mapping serves in T<sub>E</sub>X for two unrelated purposes: case transforms (upper/lower) and hyphenation. `\SetCase` handles the former, while hyphenation is handled by `\SetHyphenMap` and controlled with the package option `hyphenmap`. So, even if internally they are based on the same T<sub>E</sub>X primitive (`\lccode`), `babel` sets them separately. There are three helper macros to be used inside `\SetHyphenMap`:

- `\BabelLower{\uccode}{\lccode}` is similar to `\lccode` but it's ignored if the char has been set and saves the original `\lccode` to restore it when switching the language (except with `hyphenmap=first`).
- `\BabelLowerMM{\uccode-from}{\uccode-to}{\step}{\lccode-from}` loops through the given uppercase codes, using the step, and assigns them the `\lccode`, which is also increased (MM stands for *many-to-many*).
- `\BabelLowerMO{\uccode-from}{\uccode-to}{\step}{\lccode}` loops through the given uppercase codes, using the step, and assigns them the `\lccode`, which is fixed (MO stands for *many-to-one*).

An example is (which is redundant, because these assignments are done by both `luatex` and `xetex`):

```
\SetHyphenMap{\BabelLowerMM{"100}{\lccode}{\lccode}{\lccode}}
```

This macro is not intended to fix wrong mappings done by Unicode (which are the default in both `xetex` and `luatex`) – if an assignment is wrong, fix it directly.

## 4 Changes

### 4.1 Changes in `babel` version 3.9

Most of the changes in version 3.9 were related to bugs, either to fix them (there were lots), or to provide some alternatives. Even new features like `\babelhyphen` are intended to solve a certain problem (in this case, the lacking of a uniform syntax and behavior for shorthands across languages). These changes are described in this manual in the corresponding place. A selective list follows:

- `\select@language` did not set `\language`. This meant the language in force when auxiliary files were loaded was the one used in, for example, shorthands – if the language was `german`, a `\select@language{spanish}` had no effect.
- `\foreignlanguage` and `otherlanguage*` messed up `\extras<language>`. Scripts, encodings and many other things were not switched correctly.
- The `:ENC` mechanism for hyphenation patterns used the encoding of the *previous* language, not that of the language being selected.
- `'` (with `activeacute`) had the original value when writing to an auxiliary file, and things like an infinite loop could happen. It worked incorrectly with `^` (if activated) and also if deactivated.

- Active chars were not reset at the end of language options, and that lead to incompatibilities between languages.
- `\textormath` raised an error with a conditional.
- `\aliasshorthand` didn't work (or only in a few and very specific cases).
- `\l@english` was defined incorrectly (using `\let` instead of `\chardef`).
- `ldf` files not bundled with babel were not recognized when called as global options.

## Part II

# Source code

babel is being developed incrementally, which means parts of the code are under development and therefore incomplete. Only documented features are considered complete. In other words, use babel only as documented (except, of course, if you want to explore and test them – you can post suggestions about multilingual issues to [kadingira@tug.org](mailto:kadingira@tug.org) on <http://tug.org/mailman/listinfo/kadingira>).

## 5 Identification and loading of required files

*Code documentation is still under revision.*

The babel package after unpacking consists of the following files:

**switch.def** defines macros to set and switch languages.

**babel.def** defines the rest of macros. It has two parts: a generic one and a second one only for LaTeX.

**babel.sty** is the  $\TeX$  package, which sets options and loads language styles.

**plain.def** defines some  $\TeX$  macros required by `babel.def` and provides a few tools for Plain.

**hyphen.cfg** is the file to be used when generating the formats to load hyphenation patterns. By default it also loads `switch.def`.

The babel installer extends `docstrip` with a few “pseudo-guards” to set “variables” used at installation time. They are used with `<@name@>` at the appropriated places in the source code and shown below with `<<name>>`. That brings a little bit of literate programming.

## 6 locale directory

A required component of babel is a set of `ini` files with basic definitions for about 200 languages. They are distributed as a separate zip file, not packed as `dtx`. With them, babel will fully support Unicode engines.

Most of them are essentially finished (except bugs and mistakes, of course). Some of them are still incomplete (but they will be usable), and there are some omissions (eg, Latin and polytonic Greek, and there are no geographic areas in Spanish). Hindi, French, Occitan and Breton will show a warning related to dates. Not all include LICR variants.

This is a preliminary documentation.

`ini` files contain the actual data; `tex` files are currently just proxies to the corresponding `ini` files.

Most keys are self-explanatory.

**charset** the encoding used in the `ini` file.

**version** of the ini file

**level** “version” of the ini specification . which keys are available (they may grow in a compatible way) and how they should be read.

**encodings** a descriptive list of font encodings.

**[captions]** section of captions in the file charset

**[captions.licr]** same, but in pure ASCII using the LICR

**date.long** fields are as in the CLDR, but the syntax is different. Anything inside brackets is a date field (eg, MMMM for the month name) and anything outside is text. In addition, [ ] is a non breakable space and [ . ] is an abbreviation dot.

Keys may be further qualified in a particular language with a suffix starting with a uppercase letter. It can be just a letter (eg, babel.name.A, babel.name.B) or a name (eg, date.long.Nominative, date.long.Formal, but no language is currently using the latter). Multi-letter qualifiers are forward compatible in the sense they won’t conflict with new “global” keys (all lowercase).

## 7 Tools

```
1 <<version=3.36.1837>>
2 <<date=2019/11/26>>
```

**Do not use the following macros in ldf files. They may change in the future.** This applies mainly to those recently added for replacing, trimming and looping. The older ones, like `\bbl@afterfi`, will not change.

We define some basic macros which just make the code cleaner. `\bbl@add` is now used internally instead of `\addto` because of the unpredictable behavior of the latter. Used in `babel.def` and in `babel.sty`, which means in  $\LaTeX$  is executed twice, but we need them when defining options and `babel.def` cannot be load until options have been defined. This does not hurt, but should be fixed somehow.

```
3 <<*Basic macros>> ≡
4 \bbl@trace{Basic macros}
5 \def\bbl@stripslash{\expandafter\@gobble\string}
6 \def\bbl@add#1#2{%
7   \bbl@ifunset{\bbl@stripslash#1}%
8   {\def#1{#2}}%
9   {\expandafter\def\expandafter#1\expandafter{#1#2}}}
10 \def\bbl@xin@{\@expandtwoargs\in@}
11 \def\bbl@csarg#1#2{\expandafter#1\csname bbl@#2\endcsname}%
12 \def\bbl@cs#1{\csname bbl@#1\endcsname}
13 \def\bbl@loop#1#2#3{\bbl@loop#1{#3}#2,\@nnil,}
14 \def\bbl@loopx#1#2{\expandafter\bbl@loop\expandafter#1\expandafter{#2}}
15 \def\bbl@loop#1#2#3,{%
16   \ifx\@nnil#3\relax\else
17     \def#1{#3}#2\bbl@afterfi\bbl@loop#1{#2}%
18   \fi}
19 \def\bbl@for#1#2#3{\bbl@loopx#1{#2}{\ifx#1\@empty\else#3\fi}}
```

`\bbl@add@list` This internal macro adds its second argument to a comma separated list in its first argument. When the list is not defined yet (or empty), it will be initiated. It presumes expandable character strings.

```
20 \def\bbl@add@list#1#2{%
21   \edef#1{%
22     \bbl@ifunset{\bbl@stripslash#1}%
23     {}%
24     {\ifx#1\@empty\else#1,\fi}%
25     #2}}
```

`\bbl@afterelse` Because the code that is used in the handling of active characters may need to look ahead, we take extra care to ‘throw’ it over the `\else` and `\fi` parts of an `\if`-statement<sup>31</sup>. These macros will break if another `\if... \fi` statement appears in one of the arguments and it is not enclosed in braces.

```
26 \long\def\bbl@afterelse#1\else#2\fi{\fi#1}
27 \long\def\bbl@afterfi#1\fi{\fi#1}
```

`\bbl@exp` Now, just syntactical sugar, but it makes partial expansion of some code a lot more simple and readable. Here `\` stands for `\noexpand` and `\<. .>` for `\noexpand` applied to a built macro name (the latter does not define the macro if undefined to `\relax`, because it is created locally). The result may be followed by extra arguments, if necessary.

```
28 \def\bbl@exp#1{%
29   \begingroup
30   \let\ \noexpand
31   \def\<##1>{\expandafter\noexpand\csname##1\endcsname}%
32   \edef\bbl@exp@aux{\endgroup#1}%
33   \bbl@exp@aux}
```

`\bbl@trim` The following piece of code is stolen (with some changes) from `keyval`, by David Carlisle. It defines two macros: `\bbl@trim` and `\bbl@trim@def`. The first one strips the leading and trailing spaces from the second argument and then applies the first argument (a macro, `\toks@` and the like). The second one, as its name suggests, defines the first argument as the stripped second argument.

```
34 \def\bbl@tempa#1{%
35   \long\def\bbl@trim##1##2{%
36     \futurelet\bbl@trim@a\bbl@trim@c##2\@nil\@nil#1\@nil\relax{##1}}%
37   \def\bbl@trim@c{%
38     \ifx\bbl@trim@a\@sptoken
39       \expandafter\bbl@trim@b
40     \else
41       \expandafter\bbl@trim@b\expandafter#1%
42     \fi}%
43   \long\def\bbl@trim@b#1##1 \@nil{\bbl@trim@i##1}}
44 \bbl@tempa{ }
45 \long\def\bbl@trim@i#1\@nil#2\relax#3{#3{#1}}
46 \long\def\bbl@trim@def#1{\bbl@trim{\def#1}}
```

`\bbl@ifunset` To check if a macro is defined, we create a new macro, which does the same as `\ifundefined`. However, in an  $\epsilon$ -tex engine, it is based on `\ifcsname`, which is more efficient, and do not waste memory.

```
47 \begingroup
48 \gdef\bbl@ifunset#1{%
49   \expandafter\ifx\csname#1\endcsname\relax
50     \expandafter\@firstoftwo
51   \else
52     \expandafter\@secondoftwo
53   \fi}
54 \bbl@ifunset{ifcsname}%
55 {}%
56 {\gdef\bbl@ifunset#1{%
57   \ifcsname#1\endcsname
58     \expandafter\ifx\csname#1\endcsname\relax
59     \bbl@afterelse\expandafter\@firstoftwo
60   \else
```

<sup>31</sup>This code is based on code presented in TUGboat vol. 12, no2, June 1991 in “An expansion Power Lemma” by Sonja Maus.



```

61      \bbl@afterfi\expandafter\@secondoftwo
62      \fi
63      \else
64      \expandafter\@firstoftwo
65      \fi}}
66 \endgroup

```

`\bbl@ifblank` A tool from url, by Donald Arseneau, which tests if a string is empty or space.

```

67 \def\bbl@ifblank#1{%
68   \bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
69 \long\def\bbl@ifblank@i#1#2\@nil#3#4#5\@nil{#4}

```

For each element in the comma separated `<key>=<value>` list, execute `<code>` with #1 and #2 as the key and the value of current item (trimmed). In addition, the item is passed verbatim as #3. With the `<key>` alone, it passes `\@empty` (ie, the macro thus named, not an empty argument, which is what you get with `<key>=` and no value).

```

70 \def\bbl@forkv#1#2{%
71   \def\bbl@kvcmd##1##2##3{#2}%
72   \bbl@kvnext#1,\@nil,}
73 \def\bbl@kvnext#1,{%
74   \ifx\@nil#1\relax\else
75     \bbl@ifblank{#1}{\bbl@forkv@eq#1=\@empty=\@nil{#1}}%
76     \expandafter\bbl@kvnext
77   \fi}
78 \def\bbl@forkv@eq#1=#2=#3\@nil#4{%
79   \bbl@trim\def\bbl@forkv@a{#1}%
80   \bbl@trim{\expandafter\bbl@kvcmd\expandafter{\bbl@forkv@a}}{#2}{#4}}

```

A *for* loop. Each item (trimmed), is #1. It cannot be nested (it's doable, but we don't need it).

```

81 \def\bbl@vforeach#1#2{%
82   \def\bbl@forcmd##1{#2}%
83   \bbl@fornext#1,\@nil,}
84 \def\bbl@fornext#1,{%
85   \ifx\@nil#1\relax\else
86     \bbl@ifblank{#1}{\bbl@trim\bbl@forcmd{#1}}%
87     \expandafter\bbl@fornext
88   \fi}
89 \def\bbl@foreach#1{\expandafter\bbl@vforeach\expandafter{#1}}

```

`\bbl@replace`

```

90 \def\bbl@replace#1#2#3{% in #1 -> repl #2 by #3
91   \toks@{}}%
92   \def\bbl@replace@aux##1#2##2#2{%
93     \ifx\bbl@nil##2%
94       \toks@\expandafter{\the\toks@##1}%
95     \else
96       \toks@\expandafter{\the\toks@##1#3}%
97       \bbl@afterfi
98       \bbl@replace@aux##2#2%
99     \fi}%
100   \expandafter\bbl@replace@aux#1#2\bbl@nil#2%
101   \edef#1{\the\toks@}}

```

An extension to the previous macro. It takes into account the parameters, and it is string based (ie, if you replace `\elax` by `ho`, then `\relax` becomes `\rho`). No checking is done at all, because it is not a general purpose macro, and it is used by babel only when it works (an example where it does *not* work is in `\bbl@TG@date`, and also fails if there are macros

with spaces, because they retokenized). It may change! (or even merged with `\bbl@replace`; I'm not sure ckecking the replacement is really necessary or just paranoia).

```

102 \bbl@exp{\def\\bbl@parsedef##1\detokenize{macro:}}#2->#3\relax{%
103   \def\bbl@tempa{#1}%
104   \def\bbl@tempb{#2}%
105   \def\bbl@tempe{#3}}
106 \def\bbl@sreplace#1#2#3{%
107   \begingroup
108     \expandafter\bbl@parsedef\meaning#1\relax
109     \def\bbl@tempc{#2}%
110     \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
111     \def\bbl@tempd{#3}%
112     \edef\bbl@tempd{\expandafter\strip@prefix\meaning\bbl@tempd}%
113     \bbl@xin@{\bbl@tempc}{\bbl@tempe}% If not in macro, do nothing
114     \ifin@
115       \bbl@exp{\\bbl@replace\\bbl@tempe{\bbl@tempc}{\bbl@tempd}}%
116       \def\bbl@tempc{%      Expanded an executed below as 'uplevel'
117         \\makeatletter % "internal" macros with @ are assumed
118         \\scantokens{%
119           \bbl@tempa\\@namedef{\bbl@stripslash#1}\bbl@tempb{\bbl@tempe}}%
120           \catcode64=\the\catcode64\relax}% Restore @
121       \else
122         \let\bbl@tempc\@empty % Not \relax
123       \fi
124       \bbl@exp{%      For the 'uplevel' assignments
125     \endgroup
126     \bbl@tempc}} % empty or expand to set #1 with changes

```

Two further tools. `\bbl@samestring` first expand its arguments and then compare their expansion (sanitized, so that the catcodes do not matter). `\bbl@engine` takes the following values: 0 is pdf $\TeX$ , 1 is luatex, and 2 is xetex. You may use the latter it in your language style if you want.

```

127 \def\bbl@ifsamestring#1#2{%
128   \begingroup
129     \protected@edef\bbl@tempb{#1}%
130     \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
131     \protected@edef\bbl@tempc{#2}%
132     \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
133     \ifx\bbl@tempb\bbl@tempc
134       \aftergroup\@firstoftwo
135     \else
136       \aftergroup\@secondoftwo
137     \fi
138   \endgroup}
139 \chardef\bbl@engine=%
140 \ifx\directlua\undefined
141   \ifx\XeTeXinputencoding\undefined
142     \z@
143   \else
144     \tw@
145   \fi
146 \else
147   \@ne
148 \fi
149 <</Basic macros>>

```

Some files identify themselves with a  $\LaTeX$  macro. The following code is placed before them to define (and then undefine) if not in  $\LaTeX$ .

```

150 <<*Make sure ProvidesFile is defined>> ≡
151 \ifx\ProvidesFile\@undefined
152   \def\ProvidesFile#1[#2 #3 #4]{%
153     \wlog{File: #1 #4 #3 <#2>}%
154     \let\ProvidesFile\@undefined}
155 \fi
156 <</Make sure ProvidesFile is defined>>

```

The following code is used in `babel.sty` and `babel.def`, and loads (only once) the data in `language.dat`.

```

157 <<*Load patterns in luatex>> ≡
158 \ifx\directlua\@undefined\else
159   \ifx\bbl@luapatterns\@undefined
160     \input luababel.def
161   \fi
162 \fi
163 <</Load patterns in luatex>>

```

The following code is used in `babel.def` and `switch.def`.

```

164 <<*Load macros for plain if not LaTeX>> ≡
165 \ifx\AtBeginDocument\@undefined
166   \input plain.def\relax
167 \fi
168 <</Load macros for plain if not LaTeX>>

```

## 7.1 Multiple languages

`\language` Plain T<sub>E</sub>X version 3.0 provides the primitive `\language` that is used to store the current language. When used with a pre-3.0 version this function has to be implemented by allocating a counter. The following block is used in `switch.def` and `hyphen.cfg`; the latter may seem redundant, but remember `babel` doesn't require loading `switch.def` in the format.

```

169 <<*Define core switching macros>> ≡
170 \ifx\language\@undefined
171   \csname newcount\endcsname\language
172 \fi
173 <</Define core switching macros>>

```

`\last@language` Another counter is used to store the last language defined. For pre-3.0 formats an extra counter has to be allocated.

`\addlanguage` To add languages to T<sub>E</sub>X's memory plain T<sub>E</sub>X version 3.0 supplies `\newlanguage`, in a pre-3.0 environment a similar macro has to be provided. For both cases a new macro is defined here, because the original `\newlanguage` was defined to be `\outer`. For a format based on plain version 2.x, the definition of `\newlanguage` can not be copied because `\count 19` is used for other purposes in these formats. Therefore `\addlanguage` is defined using a definition based on the macros used to define `\newlanguage` in plain T<sub>E</sub>X version 3.0. For formats based on plain version 3.0 the definition of `\newlanguage` can be simply copied, removing `\outer`. Plain T<sub>E</sub>X version 3.0 uses `\count 19` for this purpose.

```

174 <<*Define core switching macros>> ≡
175 \ifx\newlanguage\@undefined
176   \csname newcount\endcsname\last@language
177   \def\addlanguage#1{%
178     \global\advance\last@language\@ne
179     \ifnum\last@language<\@ccclvi
180       \else

```

```

181 \errmessage{No room for a new \string\language!}%
182 \fi
183 \global\chardef#1\last@language
184 \wlog{\string#1 = \string\language\the\last@language}}
185 \else
186 \countdef\last@language=19
187 \def\addlanguage{\alloc@9\language\chardef\@cclvi}
188 \fi
189 <</Define core switching macros>>

```

Now we make sure all required files are loaded. When the command `\AtBeginDocument` doesn't exist we assume that we are dealing with a plain-based format or  $\text{\LaTeX}2.09$ . In that case the file `plain.def` is needed (which also defines `\AtBeginDocument`, and therefore it is not loaded twice). We need the first part when the format is created, and `\orig@dump` is used as a flag. Otherwise, we need to use the second part, so `\orig@dump` is not defined (`plain.def` undefines it).

Check if the current version of `switch.def` has been previously loaded (mainly, `hyphen.cfg`). If not, load it now. We cannot load `babel.def` here because we first need to declare and process the package options.

## 8 The Package File ( $\text{\LaTeX}$ , `babel.sty`)

In order to make use of the features of  $\text{\LaTeX}2\epsilon$ , the `babel` system contains a package file, `babel.sty`. This file is loaded by the `\usepackage` command and defines all the language options whose name is different from that of the `.ldf` file (like variant spellings). It also takes care of a number of compatibility issues with other packages and defines a few additional package options.

Apart from all the language options below we also have a few options that influence the behavior of language definition files.

Many of the following options don't do anything themselves, they are just defined in order to make it possible for `babel` and language definition files to check if one of them was specified by the user.

### 8.1 base

The first option to be processed is `base`, which sets the hyphenation patterns then resets `ver@babel.sty` so that  $\text{\LaTeX}$  forgets about the first loading. After `switch.def` has been loaded (above) and `\AfterBabelLanguage` defined, exits.

```

190 (*package)
191 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
192 \ProvidesPackage{babel}[\langle\langle date \rangle\rangle \langle\langle version \rangle\rangle] The Babel package]
193 \@ifpackagewith{babel}{debug}
194 {\providecommand\bbl@trace[1]{\message{^^J[ #1 ]}}}%
195 \let\bbl@debug\@firstofone}
196 {\providecommand\bbl@trace[1]{}%
197 \let\bbl@debug\@gobble}
198 \ifx\bbl@switchflag@undefined % Prevent double input
199 \let\bbl@switchflag\relax
200 \input switch.def\relax
201 \fi
202 <<Load patterns in luatex>>
203 <<Basic macros>>
204 \def\AfterBabelLanguage#1{%
205 \global\expandafter\bbl@add\csname#1.ldf-h@@k\endcsname}%

```

If the format created a list of loaded languages (in \bbl@languages), get the name of the 0-th to show the actual language used.

```

206 \ifx\bbl@languages\undefined\else
207   \begingroup
208     \catcode`\^^I=12
209     \@ifpackagewith{babel}{showlanguages}{%
210       \begingroup
211         \def\bbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}%
212         \wlog{<*languages>}%
213         \bbl@languages
214         \wlog{</languages>}%
215       \endgroup}{%
216     \endgroup
217     \def\bbl@elt#1#2#3#4{%
218       \ifnum#2=\z@
219         \gdef\bbl@nulllanguage{#1}%
220         \def\bbl@elt##1##2##3##4{%
221           \fi}%
222       \bbl@languages
223     \fi
224     \ifodd\bbl@engine
225       % Harftex is evolving, so the callback is not hardcoded, just in case
226       \def\bbl@harfpreamble{Harf pre_linebreak_filter callback}%
227       \def\bbl@activate@preotf{%
228         \let\bbl@activate@preotf\relax % only once
229         \directlua{
230           Babel = Babel or {}
231           %
232           function Babel.pre_otfload_v(head)
233             if Babel.numbers and Babel.digits_mapped then
234               head = Babel.numbers(head)
235             end
236             if Babel.bidi_enabled then
237               head = Babel.bidi(head, false, dir)
238             end
239             return head
240           end
241           %
242           function Babel.pre_otfload_h(head, gc, sz, pt, dir)
243             if Babel.numbers and Babel.digits_mapped then
244               head = Babel.numbers(head)
245             end
246             if Babel.bidi_enabled then
247               head = Babel.bidi(head, false, dir)
248             end
249             return head
250           end
251           %
252           luatexbase.add_to_callback('pre_linebreak_filter',
253             Babel.pre_otfload_v,
254             'Babel.pre_otfload_v',
255             luatexbase.priority_in_callback('pre_linebreak_filter',
256               '\bbl@harfpreamble')
257             or luatexbase.priority_in_callback('pre_linebreak_filter',
258               'luaotfload.node_processor')
259             or nil)
260           %
261           luatexbase.add_to_callback('hpack_filter',

```

```

262     Babel.pre_otfload_h,
263     'Babel.pre_otfload_h',
264     luatexbase.priority_in_callback('hpack_filter',
265     '\bbl@harfpreline')
266     or luatexbase.priority_in_callback('hpack_filter',
267     'luaotfload.node_processor')
268     or nil)
269 }%
270 \@ifpackageloaded{harfload}%
271   {\directlua{ Babel.mirroring_enabled = false }}%
272   {}%
273 \let\bbl@tempa\relax
274 \@ifpackagewith{babel}{bidi=basic}%
275   {\def\bbl@tempa{basic}}%
276   {\@ifpackagewith{babel}{bidi=basic-r}%
277     {\def\bbl@tempa{basic-r}}%
278     {}%
279   \ifx\bbl@tempa\relax\else
280     \let\bbl@beforeforeign\leavevmode
281     \AtEndOfPackage{\EnableBabelHook{babel-bidi}}%
282     \RequirePackage{luatexbase}%
283     \directlua{
284       require('babel-data-bidi.lua')
285       require('babel-bidi-\bbl@tempa.lua')
286     }
287     \bbl@activate@preotf
288   \fi
289 \fi

```

Now the base option. With it we can define (and load, with luatex) hyphenation patterns, even if we are not interested in the rest of babel. Useful for old versions of polyglossia, too.

```

290 \bbl@trace{Defining option 'base'}
291 \@ifpackagewith{babel}{base}{%
292   \ifx\directlua\@undefined
293     \DeclareOption*{\bbl@patterns{\CurrentOption}}%
294   \else
295     \DeclareOption*{\bbl@patterns@lua{\CurrentOption}}%
296   \fi
297   \DeclareOption{base}{}%
298   \DeclareOption{showlanguages}{}%
299   \ProcessOptions
300   \global\expandafter\let\csname opt@babel.sty\endcsname\relax
301   \global\expandafter\let\csname ver@babel.sty\endcsname\relax
302   \global\let\@ifl@ter@@\@ifl@ter
303   \def\@ifl@ter#1#2#3#4#5{\global\let\@ifl@ter\@ifl@ter@@}%
304   \endinput}{}%

```

## 8.2 key=value options and other general option

The following macros extract language modifiers, and only real package options are kept in the option list. Modifiers are saved and assigned to `\BabelModifiers` at `\bbl@load@language`; when no modifiers have been given, the former is `\relax`. How modifiers are handled are left to language styles; they can use `\in@`, loop them with `\@for` or `\loadkeyval`, for example.

```

305 \bbl@trace{key=value and another general options}
306 \bbl@csarg\let{tempa\expandafter}\csname opt@babel.sty\endcsname
307 \def\bbl@tempb#1.#2{%
308   #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%

```

```

309 \def\bbl@tempd#1.#2\@nnil{%
310   \ifx\@empty#2%
311     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
312   \else
313     \in@{=}{#1}\ifin@
314     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
315   \else
316     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
317     \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
318   \fi
319 \fi}
320 \let\bbl@tempc\@empty
321 \bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty\@nnil}
322 \expandafter\let\csname opt@babel.sty\endcsname\bbl@tempc

```

The next option tells babel to leave shorthand characters active at the end of processing the package. This is *not* the default as it can cause problems with other packages, but for those who want to use the shorthand characters in the preamble of their documents this can help.

```

323 \DeclareOption{KeepShorthandsActive}{}
324 \DeclareOption{activeacute}{}
325 \DeclareOption{activegrave}{}
326 \DeclareOption{debug}{}
327 \DeclareOption{noconfigs}{}
328 \DeclareOption{showlanguages}{}
329 \DeclareOption{silent}{}
330 \DeclareOption{mono}{}
331 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
332 % Don't use. Experimental:
333 \newif\ifbbl@single
334 \DeclareOption{selectors=off}{\bbl@singletrue}
335 <<More package options>>

```

Handling of package options is done in three passes. (I [JBL] am not very happy with the idea, anyway.) The first one processes options which has been declared above or follow the syntax <key>=<value>, the second one loads the requested languages, except the main one if set with the key main, and the third one loads the latter. First, we “flag” valid keys with a nil value.

```

336 \let\bbl@opt@shorthands\@nnil
337 \let\bbl@opt@config\@nnil
338 \let\bbl@opt@main\@nnil
339 \let\bbl@opt@headfoot\@nnil
340 \let\bbl@opt@layout\@nnil

```

The following tool is defined temporarily to store the values of options.

```

341 \def\bbl@tempa#1=#2\bbl@tempa{%
342   \bbl@csarg\ifx{opt@#1}\@nnil
343     \bbl@csarg\edef{opt@#1}{#2}%
344   \else
345     \bbl@error{%
346       Bad option `#1=#2'. Either you have misspelled the\\%
347       key or there is a previous setting of `#1'}{%
348       Valid keys are `shorthands', `config', `strings', `main',\\%
349       `headfoot', `safe', `math', among others.}
350   \fi}

```

Now the option list is processed, taking into account only currently declared options (including those declared with a =), and <key>=<value> options (the former take

precedence). Unrecognized options are saved in `\bbl@language@opts`, because they are language options.

```

351 \let\bbl@language@opts\@empty
352 \DeclareOption*{%
353   \bbl@xin@\string=\CurrentOption}%
354   \ifin@
355     \expandafter\bbl@tempa\CurrentOption\bbl@tempa
356   \else
357     \bbl@add@list\bbl@language@opts{\CurrentOption}%
358   \fi}

```

Now we finish the first pass (and start over).

```

359 \ProcessOptions*

```

### 8.3 Conditional loading of shorthands

If there is no `shorthands=<chars>`, the original babel macros are left untouched, but if there is, these macros are wrapped (in `babel.def`) to define only those given. A bit of optimization: if there is no `shorthands=`, then `\bbl@ifshorthand` is always true, and it is always false if `shorthands` is empty. Also, some code makes sense only with `shorthands=...`

```

360 \bbl@trace{Conditional loading of shorthands}
361 \def\bbl@sh@string#1{%
362   \ifx#1\@empty\else
363     \ifx#1t\string~%
364     \else\ifx#1c\string,%
365     \else\string#1%
366   \fi\fi
367   \expandafter\bbl@sh@string
368 \fi}
369 \ifx\bbl@opt@shorthands\@nnil
370   \def\bbl@ifshorthand#1#2#3{#2}%
371 \else\ifx\bbl@opt@shorthands\@empty
372   \def\bbl@ifshorthand#1#2#3{#3}%
373 \else

```

The following macro tests if a shorthand is one of the allowed ones.

```

374 \def\bbl@ifshorthand#1{%
375   \bbl@xin@\string#1}{\bbl@opt@shorthands}%
376   \ifin@
377     \expandafter\@firstoftwo
378   \else
379     \expandafter\@secondoftwo
380   \fi}

```

We make sure all chars in the string are ‘other’, with the help of an auxiliary macro defined above (which also zaps spaces).

```

381 \edef\bbl@opt@shorthands{%
382   \expandafter\bbl@sh@string\bbl@opt@shorthands\@empty}%

```

The following is ignored with `shorthands=off`, since it is intended to take some additional actions for certain chars.

```

383 \bbl@ifshorthand{'}%
384   {\PassOptionsToPackage{activeacute}{babel}}{}
385 \bbl@ifshorthand{`}%
386   {\PassOptionsToPackage{activegrave}{babel}}{}
387 \fi\fi

```



With `headfoot=lang` we can set the language used in heads/foots. For example, in `babel/3796` just adds `headfoot=english`. It misuses `\@resetactivechars` but seems to work.

```
388 \ifx\bbbl@opt@headfoot\@nnil\else
389   \g@addto@macro\@resetactivechars{%
390     \set@typeset@protect
391     \expandafter\select@language@x\expandafter{\bbbl@opt@headfoot}%
392     \let\protect\noexpand}
393 \fi
```

For the option `safe` we use a different approach – `\bbbl@opt@safe` says which macros are redefined (B for bibs and R for refs). By default, both are set.

```
394 \ifx\bbbl@opt@safe\@undefined
395   \def\bbbl@opt@safe{BR}
396 \fi
397 \ifx\bbbl@opt@main\@nnil\else
398   \edef\bbbl@language@opts{%
399     \ifx\bbbl@language@opts\@empty\else\bbbl@language@opts,\fi
400     \bbbl@opt@main}
401 \fi
```

For layout an auxiliary macro is provided, available for packages and language styles.

```
402 \bbbl@trace{Defining IfBabelLayout}
403 \ifx\bbbl@opt@layout\@nnil
404   \newcommand\IfBabelLayout[3]{#3}%
405 \else
406   \newcommand\IfBabelLayout[1]{%
407     \@expandtwoargs\in@{.#1.}{.\bbbl@opt@layout.}%
408     \ifin@
409       \expandafter\@firstoftwo
410     \else
411       \expandafter\@secondoftwo
412     \fi}
413 \fi
```

## 8.4 Language options

Languages are loaded when processing the corresponding option *except* if a main language has been set. In such a case, it is not loaded until all options has been processed. The following macro inputs the `ldf` file and does some additional checks (`\input` works, too, but possible errors are not caught).

```
414 \bbbl@trace{Language options}
415 \let\bbbl@afterlang\relax
416 \let\BabelModifiers\relax
417 \let\bbbl@loaded\@empty
418 \def\bbbl@load@language#1{%
419   \InputIfFileExists{#1.ldf}%
420   {\edef\bbbl@loaded{\CurrentOption
421     \ifx\bbbl@loaded\@empty\else,\bbbl@loaded\fi}%
422     \expandafter\let\expandafter\bbbl@afterlang
423       \csname\CurrentOption.ldf-h@@k\endcsname
424     \expandafter\let\expandafter\BabelModifiers
425       \csname bbl@mod@\CurrentOption\endcsname}%
426   {\bbbl@error{%
427     Unknown option '\CurrentOption'. Either you misspelled it\\%
428     or the language definition file \CurrentOption.ldf was not found}}%
429   Valid options are: shorthands=, KeepShorthandsActive,\\%
```

```

430     activeacute, activegrave, noconfigs, safe=, main=, math=\\%
431     headfoot=, strings=, config=, hyphenmap=, or a language name.}}

```

Now, we set language options whose names are different from ldf files.

```

432 \def\bbl@try@load@lang#1#2#3{%
433     \IfFileExists{\CurrentOption.ldf}%
434     {\bbl@load@language{\CurrentOption}}%
435     {\#1\bbl@load@language{\#2}\#3}}
436 \DeclareOption{afrikaans}{\bbl@try@load@lang{}{dutch}{}}
437 \DeclareOption{brazil}{\bbl@try@load@lang{}{portuges}{}}
438 \DeclareOption{brazilian}{\bbl@try@load@lang{}{portuges}{}}
439 \DeclareOption{hebrew}{%
440     \input{rlbabel.def}%
441     \bbl@load@language{hebrew}}
442 \DeclareOption{hungarian}{\bbl@try@load@lang{}{magyar}{}}
443 \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}{}}
444 \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}{}}
445 \DeclareOption{polutonikogreek}{%
446     \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}}
447 \DeclareOption{portuguese}{\bbl@try@load@lang{}{portuges}{}}
448 \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}{}}
449 \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}{}}
450 \DeclareOption{uppersorbian}{\bbl@try@load@lang{}{usorbian}{}}

```

Another way to extend the list of ‘known’ options for babel was to create the file `bblopts.cfg` in which one can add option declarations. However, this mechanism is deprecated – if you want an alternative name for a language, just create a new `.ldf` file loading the actual one. You can also set the name of the file with the package option `config=<name>`, which will load `<name>.cfg` instead.

```

451 \ifx\bbl@opt@config\@nnil
452   \@ifpackagewith{babel}{noconfigs}{}%
453   {\InputIfFileExists{bblopts.cfg}%
454     {\typeout{*****^J%
455               * Local config file bblopts.cfg used^^J%
456               *}}}%
457   {}}%
458 \else
459   \InputIfFileExists{\bbl@opt@config.cfg}%
460   {\typeout{*****^J%
461             * Local config file \bbl@opt@config.cfg used^^J%
462             *}}}%
463   {\bbl@error{%
464     Local config file '\bbl@opt@config.cfg' not found}{%
465     Perhaps you misspelled it.}}%
466 \fi

```

Recognizing global options in packages not having a closed set of them is not trivial, as for them to be processed they must be defined explicitly. So, package options not yet taken into account and stored in `bbl@language@opts` are assumed to be languages (note this list also contains the language given with `main`). If not declared above, the names of the option and the file are the same.

```

467 \bbl@for\bbl@tempa\bbl@language@opts{%
468   \bbl@ifunset{ds@\bbl@tempa}%
469   {\edef\bbl@tempb{%
470     \noexpand\DeclareOption
471     {\bbl@tempa}%
472     {\noexpand\bbl@load@language{\bbl@tempa}}}%
473   \bbl@tempb}%
474   \@empty}

```

Now, we make sure an option is explicitly declared for any language set as global option, by checking if an ldf exists. The previous step was, in fact, somewhat redundant, but that way we minimize accessing the file system just to see if the option could be a language.

```

475 \bbl@foreach\@classoptionslist{%
476   \bbl@ifunset{ds@#1}%
477   {\IfFileExists{#1.ldf}%
478    {\DeclareOption{#1}{\bbl@load@language{#1}}}%
479    {}}%
480  {}}

```

If a main language has been set, store it for the third pass.

```

481 \ifx\bbl@opt@main\@nnil\else
482   \expandafter
483   \let\expandafter\bbl@loadmain\csname ds@\bbl@opt@main\endcsname
484   \DeclareOption{\bbl@opt@main}{}
485 \fi

```

And we are done, because all options for this pass has been declared. Those already processed in the first pass are just ignored.

The options have to be processed in the order in which the user specified them (except, of course, global options, which  $\LaTeX$  processes before):

```

486 \def\AfterBabelLanguage#1{%
487   \bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang{}}
488   \DeclareOption*{}
489   \ProcessOptions*

```

This finished the second pass. Now the third one begins, which loads the main language set with the key main. A warning is raised if the main language is not the same as the last named one, or if the value of the key main is not a language. Then execute directly the option (because it could be used only in main). After loading all languages, we deactivate \AfterBabelLanguage.

```

490 \ifx\bbl@opt@main\@nnil
491   \edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
492   \let\bbl@tempc\@empty
493   \bbl@for\bbl@tempb\bbl@tempa{%
494     \bbl@xin@{\bbl@tempb,}{,\bbl@loaded,}%
495     \ifin@\edef\bbl@tempc{\bbl@tempb}\fi
496   \def\bbl@tempa#1,#2\@nnil{\def\bbl@tempb{#1}}
497   \expandafter\bbl@tempa\bbl@loaded,\@nnil
498   \ifx\bbl@tempb\bbl@tempc\else
499     \bbl@warning{%
500       Last declared language option is '\bbl@tempc',\%
501       but the last processed one was '\bbl@tempb'.\%
502       The main language cannot be set as both a global\%
503       and a package option. Use 'main=\bbl@tempc' as\%
504       option. Reported}%
505   \fi
506 \else
507   \DeclareOption{\bbl@opt@main}{\bbl@loadmain}
508   \ExecuteOptions{\bbl@opt@main}
509   \DeclareOption*{}
510   \ProcessOptions*
511 \fi
512 \def\AfterBabelLanguage{%
513   \bbl@error
514   {Too late for \string\AfterBabelLanguage}%
515   {Languages have been loaded, so I can do nothing}}

```

In order to catch the case where the user forgot to specify a language we check whether `\bbl@main@language`, has become defined. If not, no language has been loaded and an error message is displayed.

```

516 \ifx\bbl@main@language\@undefined
517   \bbl@info{%
518     You haven't specified a language. I'll use 'nil'\%
519     as the main language. Reported}
520   \bbl@load@language{nil}
521 \fi
522 \end{package}
523 \end{*core}

```

## 9 The kernel of Babel (babel.def, common)

The kernel of the babel system is stored in either `hyphen.cfg` or `switch.def` and `babel.def`. The file `babel.def` contains most of the code, while `switch.def` defines the language-switching commands; both can be read at run time. The file `hyphen.cfg` is a file that can be loaded into the format, which is necessary when you want to be able to switch hyphenation patterns (by default, it also inputs `switch.def`, for “historical reasons”, but it is not necessary). When `babel.def` is loaded it checks if the current version of `switch.def` is in the format; if not, it is loaded. A further file, `babel.sty`, contains  $\text{\LaTeX}$ -specific stuff. Because plain  $\text{\TeX}$  users might want to use some of the features of the babel system too, care has to be taken that plain  $\text{\TeX}$  can process the files. For this reason the current format will have to be checked in a number of places. Some of the code below is common to plain  $\text{\TeX}$  and  $\text{\LaTeX}$ , some of it is for the  $\text{\LaTeX}$  case only.

Plain formats based on `etex` (`etex`, `xetex`, `luatex`) don’t load `hyphen.cfg` but `etex.src`, which follows a different naming convention, so we need to define the babel names. It presumes `language.def` exists and it is the same file used when formats were created.

### 9.1 Tools

```

524 \ifx\ldf@quit\@undefined
525 \else
526   \expandafter\endinput
527 \fi
528 \langle\textit{Make sure ProvidesFile is defined}\rangle
529 \ProvidesFile{babel.def}[\langle\textit{date}\rangle] \langle\textit{version}\rangle Babel common definitions]
530 \langle\textit{Load macros for plain if not LaTeX}\rangle

```

The file `babel.def` expects some definitions made in the  $\text{\LaTeX} 2_{\epsilon}$  style file. So, In  $\text{\LaTeX} 2.09$  and Plain we must provide at least some predefined values as well some tools to set them (even if not all options are available). There in no package options, and therefore and alternative mechanism is provided. For the moment, only `\babeloptionstrings` and `\babeloptionmath` are provided, which can be defined before loading babel. `\BabelModifiers` can be set too (but not sure it works).

```

531 \ifx\bbl@ifshorthand\@undefined
532   \let\bbl@opt@shorthands\@nnil
533   \def\bbl@ifshorthand#1#2#3{#2}%
534   \let\bbl@language@opts\@empty
535   \ifx\babeloptionstrings\@undefined
536     \let\bbl@opt@strings\@nnil
537   \else
538     \let\bbl@opt@strings\babeloptionstrings
539   \fi
540   \def\BabelStringsDefault{generic}

```

```

541 \def\bbl@tempa{normal}
542 \ifx\babeloptionmath\bbl@tempa
543 \def\bbl@mathnormal{\noexpand\textormath}
544 \fi
545 \def\AfterBabelLanguage#1#2{}
546 \ifx\BabelModifiers\@undefined\let\BabelModifiers\relax\fi
547 \let\bbl@afterlang\relax
548 \def\bbl@opt@safe{BR}
549 \ifx\@uclclist\@undefined\let\@uclclist\@empty\fi
550 \ifx\bbl@trace\@undefined\def\bbl@trace#1{}\fi
551 \expandafter\newif\csname ifbbl@single\endcsname
552 \fi

And continue.
553 \ifx\bbl@switchflag\@undefined % Prevent double input
554 \let\bbl@switchflag\relax
555 \input switch.def\relax
556 \fi
557 \bbl@trace{Compatibility with language.def}
558 \ifx\bbl@languages\@undefined
559 \ifx\directlua\@undefined
560 \openin1 = language.def
561 \ifeof1
562 \closein1
563 \message{I couldn't find the file language.def}
564 \else
565 \closein1
566 \begingroup
567 \def\addlanguage#1#2#3#4#5{%
568 \expandafter\ifx\csname lang@#1\endcsname\relax\else
569 \global\expandafter\let\csname l@#1\expandafter\endcsname
570 \csname lang@#1\endcsname
571 \fi}%
572 \def\uselanguage#1{}%
573 \input language.def
574 \endgroup
575 \fi
576 \fi
577 \chardef\l@english\z@
578 \fi
579 <<Load patterns in luatex>>
580 <<Basic macros>>

```

\addto For each language four control sequences have to be defined that control the language-specific definitions. To be able to add something to these macro once they have been defined the macro \addto is introduced. It takes two arguments, a *<control sequence>* and T<sub>E</sub>X-code to be added to the *<control sequence>*.

If the *<control sequence>* has not been defined before it is defined now. The control sequence could also expand to \relax, in which case a circular definition results. The net result is a stack overflow. Otherwise the replacement text for the *<control sequence>* is expanded and stored in a token register, together with the T<sub>E</sub>X-code to be added. Finally the *<control sequence>* is redefined, using the contents of the token register.

```

581 \def\addto#1#2{%
582 \ifx#1\@undefined
583 \def#1{#2}%
584 \else
585 \ifx#1\relax
586 \def#1{#2}%
587 \else

```

```

588      {\toks@\expandafter{#1#2}%
589       \xdef#1{\the\toks@}}}%
590      \fi
591      \fi}

```

The macro `\initiate@active@char` takes all the necessary actions to make its argument a shorthand character. The real work is performed once for each character.

```

592 \def\bbl@withactive#1#2{%
593   \begingroup
594   \lccode`~=`#2\relax
595   \lowercase{\endgroup#1~}}

```

`\bbl@redefine` To redefine a command, we save the old meaning of the macro. Then we redefine it to call the original macro with the ‘sanitized’ argument. The reason why we do it this way is that we don’t want to redefine the  $\LaTeX$  macros completely in case their definitions change (they have changed in the past).

Because we need to redefine a number of commands we define the command `\bbl@redefine` which takes care of this. It creates a new control sequence, `\org@...`

```

596 \def\bbl@redefine#1{%
597   \edef\bbl@tempa{\bbl@stripslash#1}%
598   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
599   \expandafter\def\csname\bbl@tempa\endcsname}

```

This command should only be used in the preamble of the document.

```
600 \@onlypreamble\bbl@redefine
```

`\bbl@redefine@long` This version of `\babel@redefine` can be used to redefine `\long` commands such as `\ifthenelse`.

```

601 \def\bbl@redefine@long#1{%
602   \edef\bbl@tempa{\bbl@stripslash#1}%
603   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
604   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
605 \@onlypreamble\bbl@redefine@long

```

`\bbl@redefineroobust` For commands that are redefined, but which *might* be robust we need a slightly more intelligent macro. A robust command `foo` is defined to expand to `\protect\foo_`. So it is necessary to check whether `\foo_` exists. The result is that the command that is being redefined is always robust afterwards. Therefore all we need to do now is define `\foo_`.

```

606 \def\bbl@redefineroobust#1{%
607   \edef\bbl@tempa{\bbl@stripslash#1}%
608   \bbl@ifunset{\bbl@tempa\space}%
609   {\expandafter\let\csname org@\bbl@tempa\endcsname#1%
610    \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
611   {\bbl@exp{\let\<org@\bbl@tempa>\<\bbl@tempa\space>}}}%
612   \@namedef{\bbl@tempa\space}}

```

This command should only be used in the preamble of the document.

```
613 \@onlypreamble\bbl@redefineroobust
```

## 9.2 Hooks

Note they are loaded in `babel.def`. `switch.def` only provides a “hook” for hooks (with a default value which is a no-op, below). Admittedly, the current implementation is a somewhat simplistic and does very little to catch errors, but it is intended for developers, after all. `\bbl@usehooks` is the commands used by `babel` to execute hooks defined for an event.

```

614 \bbl@trace{Hooks}
615 \newcommand\AddBabelHook[3][{}]{%
616   \bbl@ifunset{bbl@hk@#2}{\EnableBabelHook{#2}}{}%
617   \def\bbl@tempa##1,#3=##2,##3\@empty{\def\bbl@tempb{##2}}%
618   \expandafter\bbl@tempa\bbl@evargs,#3=,\@empty
619   \bbl@ifunset{bbl@ev@#2@#3@#1}%
620     {\bbl@csarg\bbl@add{ev@#3@#1}{\bbl@elt{#2}}}%
621     {\bbl@csarg\let{ev@#2@#3@#1}\relax}%
622   \bbl@csarg\newcommand{ev@#2@#3@#1}[\bbl@tempb]}
623 \newcommand\EnableBabelHook[1]{\bbl@csarg\let{hk@#1}\@firstofone}
624 \newcommand\DisableBabelHook[1]{\bbl@csarg\let{hk@#1}\@gobble}
625 \def\bbl@usehooks#1#2{%
626   \def\bbl@elt##1{%
627     \@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1@#2}}%
628     \@nameuse{bbl@ev@#1@}%
629     \ifx\language\undefined\else % Test required for Plain (?)
630       \def\bbl@elt##1{%
631         \@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1@\language}#2}}%
632         \@nameuse{bbl@ev@#1@\language}%
633       \fi}

```

To ensure forward compatibility, arguments in hooks are set implicitly. So, if a further argument is added in the future, there is no need to change the existing code. Note events intended for hyphen.cfg are also loaded (just in case you need them for some reason).

```

634 \def\bbl@evargs{,% <- don't delete this comma
635   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
636   addialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
637   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
638   hyphenation=2,initiateactive=3,afterreset=0,foreign=0,foreign*=0,%
639   beforestart=0}

```

`\babelensure` The user command just parses the optional argument and creates a new macro named `\bbl@e@<language>`. We register a hook at the `afterextras` event which just executes this macro in a “complete” selection (which, if undefined, is `\relax` and does nothing). This part is somewhat involved because we have to make sure things are expanded the correct number of times.

The macro `\bbl@e@<language>` contains `\bbl@ensure{<include>}{<exclude>}{<fontenc>}`, which in turn loops over the macros names in `\bbl@captionslist`, excluding (with the help of `\in@`) those in the `exclude` list. If the `fontenc` is given (and not `\relax`), the `\fontencoding` is also added. Then we loop over the `include` list, but if the macro already contains `\foreignlanguage`, nothing is done. Note this macro (1) is not restricted to the preamble, and (2) changes are local.

```

640 \bbl@trace{Defining babelensure}
641 \newcommand\babelensure[2][{}]{% TODO - revise test files
642   \AddBabelHook{babel-ensure}{afterextras}{%
643     \ifcase\bbl@select@type
644       \@nameuse{bbl@e@\language}%
645     \fi}%
646   \begingroup
647     \let\bbl@ens@include\@empty
648     \let\bbl@ens@exclude\@empty
649     \def\bbl@ens@fontenc{\relax}%
650     \def\bbl@tempb##1{%
651       \ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}%
652     \edef\bbl@tempa{\bbl@tempb#1\@empty}%
653     \def\bbl@tempb##1=##2\@{ \@namedef{bbl@ens@##1}{##2}}%
654     \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
655     \def\bbl@tempc{\bbl@ensure}%

```

```

656 \expandafter\bb1@add\expandafter\bb1@tempc\expandafter{%
657 \expandafter{\bb1@ens@include}}%
658 \expandafter\bb1@add\expandafter\bb1@tempc\expandafter{%
659 \expandafter{\bb1@ens@exclude}}%
660 \toks@\expandafter{\bb1@tempc}%
661 \bb1@exp{%
662 \endgroup
663 \def\<bb1@e@#2>{\the\toks@{\bb1@ens@fontenc}}}%
664 \def\bb1@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
665 \def\bb1@tempb##1{% elt for (excluding) \bb1@captionslist list
666 \ifx##1\undefined % 3.32 - Don't assume the macros exists
667 \edef##1{\noexpand\bb1@nocaption
668 {\bb1@stripslash##1}{\language\bb1@stripslash##1}}%
669 \fi
670 \ifx##1\@empty\else
671 \in@{##1}{#2}%
672 \ifin@ \else
673 \bb1@ifunset{\bb1@ensure@\language}%
674 {\bb1@exp{%
675 \\\DeclareRobustCommand\<bb1@ensure@[1]{%
676 \\\foreignlanguage{\language}%
677 {\ifx\relax#3\else
678 \\\fontencoding{#3}\selectfont
679 \fi
680 #####1}}}%
681 {}%
682 \toks@\expandafter{##1}%
683 \edef##1{%
684 \bb1@csarg\noexpand{ensure@\language}%
685 {\the\toks@}}%
686 \fi
687 \expandafter\bb1@tempb
688 \fi}%
689 \expandafter\bb1@tempb\bb1@captionslist\today\@empty
690 \def\bb1@tempa##1{% elt for include list
691 \ifx##1\@empty\else
692 \bb1@csarg\in@{ensure@\language\expandafter}\expandafter{##1}%
693 \ifin@ \else
694 \bb1@tempb##1\@empty
695 \fi
696 \expandafter\bb1@tempa
697 \fi}%
698 \bb1@tempa#1\@empty}
699 \def\bb1@captionslist{%
700 \prefacename\refname\abstractname\bibname\chaptername\appendixname
701 \contentsname\listfigurename\listtablename\indexname\figurename
702 \tablename\partname\enclname\ccname\headtoname\pagename\seename
703 \alsoname\proofname\glossaryname}

```

### 9.3 Setting up language files

`\LdfInit` The second version of `\LdfInit` macro takes two arguments. The first argument is the name of the language that will be defined in the language definition file; the second argument is either a control sequence or a string from which a control sequence should be constructed. The existence of the control sequence indicates that the file has been processed before.

At the start of processing a language definition file we always check the category code of the at-sign. We make sure that it is a ‘letter’ during the processing of the file. We also save



its name as the last called option, even if not loaded.

Another character that needs to have the correct category code during processing of language definition files is the equals sign, ‘=’, because it is sometimes used in constructions with the `\let` primitive. Therefore we store its current catcode and restore it later on. Now we check whether we should perhaps stop the processing of this file. To do this we first need to check whether the second argument that is passed to `\LdfInit` is a control sequence. We do that by looking at the first token after passing #2 through string. When it is equal to `\@backslashchar` we are dealing with a control sequence which we can compare with `\@undefined`.

If so, we call `\ldf@quit` to set the main language, restore the category code of the @-sign and call `\endinput`

When #2 was *not* a control sequence we construct one and compare it with `\relax`.

Finally we check `\originalTeX`.

```
704 \bbl@trace{Macros for setting language files up}
705 \def\bbl@ldfinit{%
706   \let\bbl@screset\@empty
707   \let\BabelStrings\bbl@opt@string
708   \let\BabelOptions\@empty
709   \let\BabelLanguages\relax
710   \ifx\originalTeX\@undefined
711     \let\originalTeX\@empty
712   \else
713     \originalTeX
714   \fi}
715 \def\LdfInit#1#2{%
716   \chardef\atcatcode=\catcode`\@
717   \catcode`\@=11\relax
718   \chardef\eqcatcode=\catcode`\=
719   \catcode`\==12\relax
720   \expandafter\if\expandafter\@backslashchar
721     \expandafter\@car\string#2\@nil
722   \ifx#2\@undefined\else
723     \ldf@quit{#1}%
724   \fi
725 \else
726   \expandafter\ifx\csname#2\endcsname\relax\else
727     \ldf@quit{#1}%
728   \fi
729 \fi
730 \bbl@ldfinit}
```

`\ldf@quit` This macro interrupts the processing of a language definition file.

```
731 \def\ldf@quit#1{%
732   \expandafter\main@language\expandafter{#1}%
733   \catcode`\@=\atcatcode \let\atcatcode\relax
734   \catcode`\==\eqcatcode \let\eqcatcode\relax
735   \endinput}
```

`\ldf@finish` This macro takes one argument. It is the name of the language that was defined in the language definition file.

We load the local configuration file if one is present, we set the main language (taking into account that the argument might be a control sequence that needs to be expanded) and reset the category code of the @-sign.

```
736 \def\bbl@afterldf#1{%
737   \bbl@afterlang
738   \let\bbl@afterlang\relax
```

```

739 \let\BabelModifiers\relax
740 \let\bbl@screset\relax}%
741 \def\ldf@finish#1{%
742   \loadlocalcfg{#1}%
743   \bbl@afterldf{#1}%
744   \expandafter\main@language\expandafter{#1}%
745   \catcode`\@=\atcatcode \let\atcatcode\relax
746   \catcode`\==\eqcatcode \let\eqcatcode\relax}

```

After the preamble of the document the commands `\LdfInit`, `\ldf@quit` and `\ldf@finish` are no longer needed. Therefore they are turned into warning messages in  $\LaTeX$ .

```

747 \onlypreamble\LdfInit
748 \onlypreamble\ldf@quit
749 \onlypreamble\ldf@finish

```

`\main@language` This command should be used in the various language definition files. It stores its argument in `\bbl@main@language`; to be used to switch to the correct language at the beginning of the document.

```

750 \def\main@language#1{%
751   \def\bbl@main@language{#1}%
752   \let\language\name\bbl@main@language
753   \bbl@id@assign
754   \bbl@patterns{\language}

```

We also have to make sure that some code gets executed at the beginning of the document. Languages do not set `\pagedir`, so we set here for the whole document to the main `\bodydir`.

```

755 \def\bbl@beforestart{%
756   \bbl@usehooks{beforestart}}}%
757 \global\let\bbl@beforestart\relax}
758 \AtBeginDocument{%
759   \@nameuse{bbl@beforestart}}%
760 \if@filesw
761   \immediate\write\@mainaux{\string\@nameuse{bbl@beforestart}}}%
762 \fi
763 \expandafter\selectlanguage\expandafter{\bbl@main@language}%
764 \ifbbl@single % must go after the line above
765   \renewcommand\selectlanguage[1]{}%
766   \renewcommand\foreignlanguage[2]{#2}%
767   \global\let\babel@aux\@gobbletwo % Also as flag
768 \fi
769 \ifcase\bbl@engine\or\pagedir\bodydir\fi} % TODO - a better place

```

A bit of optimization. Select in heads/foots the language only if necessary.

```

770 \def\select@language@x#1{%
771   \ifcase\bbl@select@type
772     \bbl@ifsamestring\language\name{#1}{\select@language{#1}}%
773   \else
774     \select@language{#1}%
775   \fi}

```

## 9.4 Shorthands

`\bbl@add@special` The macro `\bbl@add@special` is used to add a new character (or single character control sequence) to the macro `\dospecials` (and `\@sanitize` if  $\LaTeX$  is used). It is used only at one place, namely when `\initiate@active@char` is called (which is ignored if the char

has been made active before). Because \@sanitize can be undefined, we put the definition inside a conditional.

Items are added to the lists without checking its existence or the original catcode. It does not hurt, but should be fixed. It's already done with \nfss@catcodes, added in 3.10.

```

776 \bbl@trace{Shorhands}
777 \def\bbl@add@special#1{% 1:a macro like \", \?, etc.
778   \bbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
779   \bbl@ifunset{@sanitize}{\bbl@add@sanitize{\@makeother#1}}%
780   \ifx\nfss@catcodes\@undefined\else % TODO - same for above
781     \begingroup
782       \catcode`#1\active
783       \nfss@catcodes
784       \ifnum\catcode`#1=\active
785         \endgroup
786         \bbl@add\nfss@catcodes{\@makeother#1}%
787       \else
788         \endgroup
789       \fi
790   \fi}

```

`\bbl@remove@special` The companion of the former macro is `\bbl@remove@special`. It removes a character from the set macros `\dospecials` and `\@sanitize`, but it is not used at all in the babel core.

```

791 \def\bbl@remove@special#1{%
792   \begingroup
793   \def\x##1##2{\ifnum`#1=##2\noexpand\@empty
794     \else\noexpand##1\noexpand##2\fi}%
795   \def\do{\x\do}%
796   \def\@makeother{\x\@makeother}%
797   \edef\x{\endgroup
798     \def\noexpand\dospecials{\dospecials}%
799     \expandafter\ifx\csname @sanitize\endcsname\relax\else
800       \def\noexpand\@sanitize{\@sanitize}%
801     \fi}%
802   \x}

```

`\initiate@active@char` A language definition file can call this macro to make a character active. This macro takes one argument, the character that is to be made active. When the character was already active this macro does nothing. Otherwise, this macro defines the control sequence `\normal@char⟨char⟩` to expand to the character in its ‘normal state’ and it defines the active character to expand to `\normal@char⟨char⟩` by default (`⟨char⟩` being the character to be made active). Later its definition can be changed to expand to `\active@char⟨char⟩` by calling `\bbl@activate{⟨char⟩}`.

For example, to make the double quote character active one could have `\initiate@active@char{"}` in a language definition file. This defines `"` as `\active@prefix "\active@char"` (where the first `"` is the character with its original catcode, when the shorthand is created, and `\active@char` is a single token). In protected contexts, it expands to `\protect "` or `\noexpand "` (ie, with the original `"`); otherwise `\active@char` is executed. This macro in turn expands to `\normal@char` in “safe” contexts (eg, `\label`), but `\user@active` in normal “unsafe” ones. The latter search a definition in the user, language and system levels, in this order, but if none is found, `\normal@char` is used. However, a deactivated shorthand (with `\bbl@deactivate` is defined as `\active@prefix "\normal@char`).

The following macro is used to define shorthands in the three levels. It takes 4 arguments: the (string’ed) character, `\<level>@group`, `<level>@active` and `<next-level>@active` (except in system).

```

803 \def\bbl@active@def#1#2#3#4{%

```

```

804 \namedef{#3#1}{%
805 \expandafter\ifx\csname#2@sh@#1@endcsname\relax
806 \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
807 \else
808 \bbl@afterfi\csname#2@sh@#1@endcsname
809 \fi}%

```

When there is also no current-level shorthand with an argument we will check whether there is a next-level defined shorthand for this active character.

```

810 \long\namedef{#3@arg#1}##1{%
811 \expandafter\ifx\csname#2@sh@#1\string##1@endcsname\relax
812 \bbl@afterelse\csname#4#1@endcsname##1%
813 \else
814 \bbl@afterfi\csname#2@sh@#1\string##1@endcsname
815 \fi}%

```

\initiate@active@char calls \@initiate@active@char with 3 arguments. All of them are the same character with different catcodes: active, other (\string'ed) and the original one. This trick simplifies the code a lot.

```

816 \def\initiate@active@char#1{%
817 \bbl@ifunset{active@char\string#1}%
818 {\bbl@withactive
819 {\expandafter\@initiate@active@char\expandafter}#1\string#1#1}%
820 {}}

```

The very first thing to do is saving the original catcode and the original definition, even if not active, which is possible (undefined characters require a special treatment to avoid making them \relax).

```

821 \def\@initiate@active@char#1#2#3{%
822 \bbl@csarg\edef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%
823 \ifx#1\@undefined
824 \bbl@csarg\edef{oridef@#2}{\let\noexpand#1\noexpand\@undefined}%
825 \else
826 \bbl@csarg\let{oridef@#2}#1%
827 \bbl@csarg\edef{oridef@#2}{%
828 \let\noexpand#1%
829 \expandafter\noexpand\csname bbl@oridef@@#2@endcsname}%
830 \fi

```

If the character is already active we provide the default expansion under this shorthand mechanism. Otherwise we write a message in the transcript file, and define \normal@char⟨char⟩ to expand to the character in its default state. If the character is mathematically active when babel is loaded (for example ') the normal expansion is somewhat different to avoid an infinite loop (but it does not prevent the loop if the mathcode is set to "8000 *a posteriori*").

```

831 \ifx#1#3\relax
832 \expandafter\let\csname normal@char#2@endcsname#3%
833 \else
834 \bbl@info{Making #2 an active character}%
835 \ifnum\mathcode`#2=\ifodd\bbl@engine"1000000 \else"8000 \fi
836 \namedef{normal@char#2}{%
837 \textormath{#3}{\csname bbl@oridef@@#2@endcsname}}%
838 \else
839 \namedef{normal@char#2}{#3}%
840 \fi

```

To prevent problems with the loading of other packages after babel we reset the catcode of the character to the original one at the end of the package and of each language file (except

with `KeepShorthandsActive`). It is re-activate again at `\begin{document}`. We also need to make sure that the shorthands are active during the processing of the `.aux` file. Otherwise some citations may give unexpected results in the printout when a shorthand was used in the optional argument of `\bibitem` for example. Then we make it active (not strictly necessary, but done for backward compatibility).

```

841 \bbl@restoreactive{#2}%
842 \AtBeginDocument{%
843   \catcode`#2\active
844   \if@filesw
845     \immediate\write\@mainaux{\catcode`\string#2\active}%
846   \fi}%
847 \expandafter\bbl@add@special\csname#2\endcsname
848 \catcode`#2\active
849 \fi

```

Now we have set `\normal@char⟨char⟩`, we must define `\active@char⟨char⟩`, to be executed when the character is activated. We define the first level expansion of `\active@char⟨char⟩` to check the status of the `@safe@actives` flag. If it is set to true we expand to the ‘normal’ version of this character, otherwise we call `\user@active⟨char⟩` to start the search of a definition in the user, language and system levels (or eventually `normal@char⟨char⟩`).

```

850 \let\bbl@tempa\@firstoftwo
851 \if\string^#2%
852   \def\bbl@tempa{\noexpand\textormath}%
853 \else
854   \ifx\bbl@mathnormal\@undefined\else
855     \let\bbl@tempa\bbl@mathnormal
856   \fi
857 \fi
858 \expandafter\edef\csname active@char#2\endcsname{%
859   \bbl@tempa
860   {\noexpand\if@safe@actives
861     \noexpand\expandafter
862     \expandafter\noexpand\csname normal@char#2\endcsname
863   \noexpand\else
864     \noexpand\expandafter
865     \expandafter\noexpand\csname bbl@doactive#2\endcsname
866   \noexpand\fi}%
867   {\expandafter\noexpand\csname normal@char#2\endcsname}}%
868 \bbl@csarg\edef{doactive#2}{%
869   \expandafter\noexpand\csname user@active#2\endcsname}%

```

We now define the default values which the shorthand is set to when activated or deactivated. It is set to the deactivated form (globally), so that the character expands to

`\active@prefix⟨char⟩\normal@char⟨char⟩`

(where `\active@char⟨char⟩` is *one* control sequence!).

```

870 \bbl@csarg\edef{active@#2}{%
871   \noexpand\active@prefix\noexpand#1%
872   \expandafter\noexpand\csname active@char#2\endcsname}%
873 \bbl@csarg\edef{normal@#2}{%
874   \noexpand\active@prefix\noexpand#1%
875   \expandafter\noexpand\csname normal@char#2\endcsname}%
876 \expandafter\let\expandafter#1\csname bbl@normal@#2\endcsname

```

The next level of the code checks whether a user has defined a shorthand for himself with this character. First we check for a single character shorthand. If that doesn’t exist we check for a shorthand with an argument.

```

877 \bbl@active@def#2\user@group{user@active}{language@active}%
878 \bbl@active@def#2\language@group{language@active}{system@active}%
879 \bbl@active@def#2\system@group{system@active}{normal@char}%

```

In order to do the right thing when a shorthand with an argument is used by itself at the end of the line we provide a definition for the case of an empty argument. For that case we let the shorthand character expand to its non-active self. Also, When a shorthand combination such as ' ' ends up in a heading  $\TeX$  would see `\protect'\protect'`. To prevent this from happening a couple of shorthand needs to be defined at user level.

```

880 \expandafter\edef\csname\user@group @sh@#2@@\endcsname
881   {\expandafter\noexpand\csname normal@char#2\endcsname}%
882 \expandafter\edef\csname\user@group @sh@#2@\string\protect\endcsname
883   {\expandafter\noexpand\csname user@active#2\endcsname}%

```

Finally, a couple of special cases are taken care of. (1) If we are making the right quote (') active we need to change `\pr@m@s` as well. Also, make sure that a single ' in math mode 'does the right thing'. (2) If we are using the caret (^) as a shorthand character special care should be taken to make sure math still works. Therefore an extra level of expansion is introduced with a check for math mode on the upper level.

```

884 \if\string'#2%
885   \let\prim@s\bbl@prim@s
886   \let\active@math@prime#1%
887 \fi
888 \bbl@usehooks{initiateactive}{{#1}{#2}{#3}}

```

The following package options control the behavior of shorthands in math mode.

```

889 <<(*More package options)>> ≡
890 \DeclareOption{math=active}{}
891 \DeclareOption{math=normal}{{\def\bbl@mathnormal{\noexpand\textormath}}}
892 <</More package options>>

```

Initiating a shorthand makes active the char. That is not strictly necessary but it is still done for backward compatibility. So we need to restore the original catcode at the end of package *and* and the end of the *ldf*.

```

893 \ifpackagewith{babel}{KeepShorthandsActive}%
894   {\let\bbl@restoreactive@gobble}%
895   {\def\bbl@restoreactive#1{%
896     \bbl@exp{%
897       \\\AfterBabelLanguage\\\CurrentOption
898       {\catcode`#1=\the\catcode`#1\relax}%
899       \\\AtEndOfPackage
900       {\catcode`#1=\the\catcode`#1\relax}}}%
901   \AtEndOfPackage{\let\bbl@restoreactive@gobble}}

```

`\bbl@sh@select` This command helps the shorthand supporting macros to select how to proceed. Note that this macro needs to be expandable as do all the shorthand macros in order for them to work in expansion-only environments such as the argument of `\hyphenation`. This macro expects the name of a group of shorthands in its first argument and a shorthand character in its second argument. It will expand to either `\bbl@firstcs` or `\bbl@scndcs`. Hence two more arguments need to follow it.

```

902 \def\bbl@sh@select#1#2{%
903   \expandafter\ifx\csname#1@sh@#2@sel\endcsname\relax
904     \bbl@afterelse\bbl@scndcs
905   \else
906     \bbl@afterfi\csname#1@sh@#2@sel\endcsname
907   \fi}

```

`\active@prefix` The command `\active@prefix` which is used in the expansion of active characters has a function similar to `\OT1-cmd` in that it `\protects` the active character whenever `\protect` is *not* `\@typeset@protect`. The `\@gobble` is needed to remove a token such as `\activechar:` (when the double colon was the active character to be dealt with). There are two definitions, depending of `\ifincsname` is available. If there is, the expansion will be more robust.

```

908 \begingroup
909 \bbl@ifunset{ifincsname}%
910 {\gdef\active@prefix#1{%
911   \ifx\protect\@typeset@protect
912   \else
913     \ifx\protect\@unexpandable@protect
914       \noexpand#1%
915     \else
916       \protect#1%
917     \fi
918     \expandafter\@gobble
919   \fi}}
920 {\gdef\active@prefix#1{%
921   \ifincsname
922     \string#1%
923     \expandafter\@gobble
924   \else
925     \ifx\protect\@typeset@protect
926     \else
927       \ifx\protect\@unexpandable@protect
928         \noexpand#1%
929       \else
930         \protect#1%
931       \fi
932       \expandafter\expandafter\expandafter\@gobble
933     \fi
934   \fi}}
935 \endgroup

```

`\if@safe@actives` In some circumstances it is necessary to be able to change the expansion of an active character on the fly. For this purpose the switch `@safe@actives` is available. The setting of this switch should be checked in the first level expansion of `\active@char⟨char⟩`.

```

936 \newif\if@safe@actives
937 \@safe@activesfalse

```

`\bbl@restore@actives` When the output routine kicks in while the active characters were made “safe” this must be undone in the headers to prevent unexpected typeset results. For this situation we define a command to make them “unsafe” again.

```

938 \def\bbl@restore@actives{\if@safe@actives\@safe@activesfalse\fi}

```

`\bbl@activate` Both macros take one argument, like `\initiate@active@char`. The macro is used to  
`\bbl@deactivate` change the definition of an active character to expand to `\active@char⟨char⟩` in the case of `\bbl@activate`, or `\normal@char⟨char⟩` in the case of `\bbl@deactivate`.

```

939 \def\bbl@activate#1{%
940   \bbl@withactive{\expandafter\let\expandafter}#1%
941   \csname bbl@active@\string#1\endcsname}
942 \def\bbl@deactivate#1{%
943   \bbl@withactive{\expandafter\let\expandafter}#1%
944   \csname bbl@normal@\string#1\endcsname}

```

|  |   |
|--|---|
| <code>\bbl@firstcs</code><br><code>\bbl@scndcs</code>                                  | <p>These macros have two arguments. They use one of their arguments to build a control sequence from.</p> <pre> 945 \def\bbl@firstcs#1#2{\csname#1\endcsname} 946 \def\bbl@scndcs#1#2{\csname#2\endcsname} </pre>   |
| <code>\declare@shorthand</code>  | <p>The command <code>\declare@shorthand</code> is used to declare a shorthand on a certain level. It takes three arguments:</p> <ol style="list-style-type: none"> <li>1. a name for the collection of shorthands, i.e. ‘system’, or ‘dutch’;</li> <li>2. the character (sequence) that makes up the shorthand, i.e. <code>~</code> or <code>"a</code>;</li> <li>3. the code to be executed when the shorthand is encountered.</li> </ol> <pre> 947 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil} 948 \def\@decl@short#1#2#3\@nil#4{% 949   \def\bbl@tempa{#3}% 950   \ifx\bbl@tempa\@empty 951     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs 952     \bbl@ifunset{#1@sh@\string#2@}\}% 953     {\def\bbl@tempa{#4}% 954       \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa 955       \else 956         \bbl@info 957           {Redefining #1 shorthand \string#2\\ 958            in language \CurrentOption}% 959         \fi}% 960     \@namedef{#1@sh@\string#2@}{#4}% 961   \else 962     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs 963     \bbl@ifunset{#1@sh@\string#2@\string#3@}\}% 964     {\def\bbl@tempa{#4}% 965       \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa 966       \else 967         \bbl@info 968           {Redefining #1 shorthand \string#2\string#3\\ 969            in language \CurrentOption}% 970         \fi}% 971     \@namedef{#1@sh@\string#2@\string#3@}{#4}% 972   \fi} </pre> |
| <code>\textormath</code>   | <p>Some of the shorthands that will be declared by the language definition files have to be usable in both text and mathmode. To achieve this the helper macro <code>\textormath</code> is provided.</p> <pre> 973 \def\textormath{% 974   \ifmmode 975     \expandafter\@secondoftwo 976   \else 977     \expandafter\@firstoftwo 978   \fi} </pre>  |
| <code>\user@group</code><br><code>\language@group</code><br><code>\system@group</code> | <p>The current concept of ‘shorthands’ supports three levels or groups of shorthands. For each level the name of the level or group is stored in a macro. The default is to have a user group; use language group ‘english’ and have a system group called ‘system’.</p> <pre> 979 \def\user@group{user} 980 \def\language@group{english} 981 \def\system@group{system} </pre>  |



`\useshortands` This is the user level command to tell  $\TeX$  that user level shorthands will be used in the document. It takes one argument, the character that starts a shorthand. First note that this is user level, and then initialize and activate the character for use as a shorthand character (ie, it's active in the preamble). Languages can deactivate shorthands, so a starred version is also provided which activates them always after the language has been switched.

```

982 \def\useshortands{%
983   \@ifstar\bb1@usesesh@s{\bb1@usesesh@x{}}
984 \def\bb1@usesesh@s#1{%
985   \bb1@usesesh@x
986   {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bb1@activate{#1}}}%
987   {#1}}
988 \def\bb1@usesesh@x#1#2{%
989   \bb1@ifshorthand{#2}%
990   {\def\user@group{user}%
991     \initiate@active@char{#2}%
992     #1%
993     \bb1@activate{#2}}%
994   {\bb1@error
995     {Cannot declare a shorthand turned off (\string#2)}
996     {Sorry, but you cannot use shorthands which have been\%
997       turned off in the package options}}}
```

`\defineshorthand` Currently we only support two groups of user level shorthands, named internally `user` and `user@<lang>` (language-dependent user shorthands). By default, only the first one is taken into account, but if the former is also used (in the optional argument of `\defineshorthand`) a new level is inserted for it (`user@generic`, done by `\bb1@set@user@generic`); we make also sure `{}` and `\protect` are taken into account in this new top level.

```

998 \def\user@language@group{user@\language@group}
999 \def\bb1@set@user@generic#1#2{%
1000   \bb1@ifunset{user@generic@active#1}%
1001   {\bb1@active@def#1\user@language@group{user@active}{user@generic@active}%
1002     \bb1@active@def#1\user@group{user@generic@active}{language@active}%
1003     \expandafter\edef\csname#2@sh@#1@@\endcsname{%
1004       \expandafter\noexpand\csname normal@char#1\endcsname}%
1005     \expandafter\edef\csname#2@sh@#1@\string\protect\endcsname{%
1006       \expandafter\noexpand\csname user@active#1\endcsname}}%
1007   \@empty}
1008 \newcommand\defineshorthand[3][user]{%
1009   \edef\bb1@tempa{\zap@space#1 \@empty}%
1010   \bb1@for\bb1@tempb\bb1@tempa{%
1011     \if*\expandafter\car\bb1@tempb\@nil
1012       \edef\bb1@tempb{user\expandafter@gobble\bb1@tempb}%
1013       \@expandtwoargs
1014       \bb1@set@user@generic{\expandafter\string\car#2\@nil}\bb1@tempb
1015     \fi
1016     \declare@shorthand{\bb1@tempb}{#2}{#3}}}
```

`\languageshortands` A user level command to change the language from which shorthands are used. Unfortunately, babel currently does not keep track of defined groups, and therefore there is no way to catch a possible change in casing.

```

1017 \def\languageshortands#1{\def\language@group{#1}}
```

`\aliasshorthand` First the new shorthand needs to be initialized,

```

1018 \def\aliasshorthand#1#2{%
1019   \bb1@ifshorthand{#2}%
1020   {\expandafter\ifx\csname active@char\string#2\endcsname\relax
1021     \ifx\document\@notprerr
```

```

1022      \@notshorthand{#2}%
1023      \else
1024      \initiate@active@char{#2}%

Then, we define the new shorthand in terms of the original one, but note with
\aliasshorthands{"}{/} is \active@prefix / \active@char/, so we still need to let the
latest to \active@char".

1025      \expandafter\let\csname active@char\string#2\expandafter\endcsname
1026      \csname active@char\string#1\endcsname
1027      \expandafter\let\csname normal@char\string#2\expandafter\endcsname
1028      \csname normal@char\string#1\endcsname
1029      \bbl@activate{#2}%
1030      \fi
1031      \fi}%
1032      {\bbl@error
1033      {Cannot declare a shorthand turned off (\string#2)}
1034      {Sorry, but you cannot use shorthands which have been\\%
1035      turned off in the package options}}}

```

\@notshorthand

```

1036 \def\@notshorthand#1{%
1037   \bbl@error{%
1038     The character '\string #1' should be made a shorthand character;\\%
1039     add the command \string\usesshorthands\string{#1\string} to
1040     the preamble.\\%
1041     I will ignore your instruction}%
1042   {You may proceed, but expect unexpected results}}

```

\shorthandon The first level definition of these macros just passes the argument on to \bbl@switch@sh, adding \@nil at the end to denote the end of the list of characters.

```

1043 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
1044 \DeclareRobustCommand*\shorthandoff{%
1045   \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
1046 \def\bbl@shorthandoff#1#2{\bbl@switch@sh#1#2\@nnil}

```

\bbl@switch@sh The macro \bbl@switch@sh takes the list of characters apart one by one and subsequently switches the category code of the shorthand character according to the first argument of \bbl@switch@sh.

But before any of this switching takes place we make sure that the character we are dealing with is known as a shorthand character. If it is, a macro such as \active@char" should exist.

Switching off and on is easy – we just set the category code to ‘other’ (12) and \active. With the starred version, the original catcode and the original definition, saved in @initiate@active@char, are restored.

```

1047 \def\bbl@switch@sh#1#2{%
1048   \ifx#2\@nnil\else
1049     \bbl@ifunset{\bbl@active@\string#2}%
1050     {\bbl@error
1051       {I cannot switch '\string#2' on or off--not a shorthand}%
1052       {This character is not a shorthand. Maybe you made\\%
1053       a typing mistake? I will ignore your instruction}}%
1054     {\ifcase#1%
1055       \catcode`#2\relax
1056       \or
1057       \catcode`#2\active
1058       \or
1059       \csname bbl@oricat@\string#2\endcsname

```

```

1060      \csname bbl@oridef@string#2\endcsname
1061      \fi}%
1062      \bbl@afterfi\bbl@switch@sh#1%
1063      \fi}

```

Note the value is that at the expansion time, eg, in the preamble shorthands are usually deactivated.

```

1064 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
1065 \def\bbl@putsh#1{%
1066   \bbl@ifunset{\bbl@active@string#1}%
1067   {\bbl@putsh@i#1\@empty\@nnil}%
1068   {\csname bbl@active@string#1\endcsname}}
1069 \def\bbl@putsh@i#1#2\@nnil{%
1070   \csname\language @sh@string#1@%
1071     \ifx\@empty#2\else\string#2\fi\endcsname}
1072 \ifx\bbl@opt@shorthands\@nnil\else
1073   \let\bbl@s@initiate@active@char\initiate@active@char
1074   \def\initiate@active@char#1{%
1075     \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
1076   \let\bbl@s@switch@sh\bbl@switch@sh
1077   \def\bbl@switch@sh#1#2{%
1078     \ifx#2\@nnil\else
1079       \bbl@afterfi
1080       \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
1081       \fi}
1082   \let\bbl@s@activate\bbl@activate
1083   \def\bbl@activate#1{%
1084     \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
1085   \let\bbl@s@deactivate\bbl@deactivate
1086   \def\bbl@deactivate#1{%
1087     \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
1088   \fi

```

You may want to test if a character is a shorthand. Note it does not test whether the shorthand is on or off.

```

1089 \newcommand\ifbabelshorthand[3]{\bbl@ifunset{\bbl@active@string#1}{#3}{#2}}

```

**\bbl@prim@s** One of the internal macros that are involved in substituting \prime for each right quote in mathmode is \prim@s. This checks if the next character is a right quote. When the right quote is active, the definition of this macro needs to be adapted to look also for an active right quote; the hat could be active, too.

```

\bbl@pr@m@s
1090 \def\bbl@prim@s{%
1091   \prime\futurelet\@let@token\bbl@pr@m@s}
1092 \def\bbl@if@primes#1#2{%
1093   \ifx#1\@let@token
1094     \expandafter\@firstoftwo
1095   \else\ifx#2\@let@token
1096     \bbl@afterelse\expandafter\@firstoftwo
1097   \else
1098     \bbl@afterfi\expandafter\@secondoftwo
1099   \fi\fi}
1100 \begingroup
1101   \catcode`\^=7 \catcode`\*=\active \lccode`\*=\^
1102   \catcode`\'=12 \catcode`\"=\active \lccode`\"=\'
1103   \lowercase{%
1104     \gdef\bbl@pr@m@s{%
1105       \bbl@if@primes" '%
1106       \pr@@@s

```

```

1107      {\bbl@if@primes*^{\pr@@@t\egroup}}
1108 \endgroup

```

Usually the `~` is active and expands to `\penalty\@M\_\_`. When it is written to the `.aux` file it is written expanded. To prevent that and to be able to use the character `~` as a start character for a shorthand, it is redefined here as a one character shorthand on system level. The system declaration is in most cases redundant (when `~` is still a non-break space), and in some cases is inconvenient (if `~` has been redefined); however, for backward compatibility it is maintained (some existing documents may rely on the `babel` value).

```

1109 \initiate@active@char{~}
1110 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
1111 \bbl@activate{~}

```

`\OT1dqpos` The position of the double quote character is different for the OT1 and T1 encodings. It will  
`\T1dqpos` later be selected using the `\f@encoding` macro. Therefore we define two macros here to store the position of the character in these encodings.

```

1112 \expandafter\def\csname OT1dqpos\endcsname{127}
1113 \expandafter\def\csname T1dqpos\endcsname{4}

```

When the macro `\f@encoding` is undefined (as it is in plain  $\TeX$ ) we define it here to expand to OT1

```

1114 \ifx\f@encoding\@undefined
1115   \def\f@encoding{OT1}
1116 \fi

```

## 9.5 Language attributes

Language attributes provide a means to give the user control over which features of the language definition files he wants to enable.

`\languageattribute` The macro `\languageattribute` checks whether its arguments are valid and then activates the selected language attribute. First check whether the language is known, and then process each attribute in the list.

```

1117 \bbl@trace{Language attributes}
1118 \newcommand\languageattribute[2]{%
1119   \def\bbl@tempc{#1}%
1120   \bbl@fixname\bbl@tempc
1121   \bbl@iflanguage\bbl@tempc{%
1122     \bbl@vforeach{#2}{%

```

We want to make sure that each attribute is selected only once; therefore we store the already selected attributes in `\bbl@known@attribs`. When that control sequence is not yet defined this attribute is certainly not selected before.

```

1123     \ifx\bbl@known@attribs\@undefined
1124       \in@false
1125     \else

```

Now we need to see if the attribute occurs in the list of already selected attributes.

```

1126       \bbl@xin@{\bbl@tempc-##1,}{\bbl@known@attribs,}%
1127     \fi

```

When the attribute was in the list we issue a warning; this might not be the users intention.

```

1128     \ifin@
1129       \bbl@warning{%
1130         You have more than once selected the attribute '##1'\%
1131         for language #1. Reported}%
1132     \else

```

When we end up here the attribute is not selected before. So, we add it to the list of selected attributes and execute the associated  $\TeX$ -code.

```

1133     \bbl@exp{%
1134         \\bbl@add@list\\bbl@known@attribs{\bbl@tempc-##1}}%
1135     \edef\bbl@tempa{\bbl@tempc-##1}%
1136     \expandafter\bbl@ifknown@ttrib\expandafter{\bbl@tempa}\bbl@attributes%
1137     {\csname\bbl@tempc @attr@##1\endcsname}%
1138     {\@attrerr{\bbl@tempc}{##1}}%
1139     \fi}}

```

This command should only be used in the preamble of a document.

```

1140 \@onlypreamble\languageattribute

```

The error text to be issued when an unknown attribute is selected.

```

1141 \newcommand*{\@attrerr}[2]{%
1142     \bbl@error
1143     {The attribute #2 is unknown for language #1.}%
1144     {Your command will be ignored, type <return> to proceed}}

```

`\bbl@declare@ttribute` This command adds the new language/attribute combination to the list of known attributes.

Then it defines a control sequence to be executed when the attribute is used in a document. The result of this should be that the macro `\extras...` for the current language is extended, otherwise the attribute will not work as its code is removed from memory at `\begin{document}`.

```

1145 \def\bbl@declare@ttribute#1#2#3{%
1146     \bbl@xin@{,#2,}{,\BabelModifiers,}%
1147     \ifin@
1148         \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
1149     \fi
1150     \bbl@add@list\bbl@attributes{#1-#2}%
1151     \expandafter\def\csname#1@attr@#2\endcsname{#3}}

```

`\bbl@ifattributeset` This internal macro has 4 arguments. It can be used to interpret  $\TeX$  code based on whether a certain attribute was set. This command should appear inside the argument to `\AtBeginDocument` because the attributes are set in the document preamble, *after* babel is loaded.

The first argument is the language, the second argument the attribute being checked, and the third and fourth arguments are the true and false clauses.

```

1152 \def\bbl@ifattributeset#1#2#3#4{%

```

First we need to find out if any attributes were set; if not we're done.

```

1153     \ifx\bbl@known@attribs\@undefined
1154         \in@false
1155     \else

```

The we need to check the list of known attributes.

```

1156     \bbl@xin@{,#1-#2,}{,\bbl@known@attribs,}%
1157     \fi

```

When we're this far `\ifin@` has a value indicating if the attribute in question was set or not. Just to be safe the code to be executed is 'thrown over the `\fi`'.

```

1158     \ifin@
1159         \bbl@afterelse#3%
1160     \else
1161         \bbl@afterfi#4%
1162     \fi
1163 }

```

`\bbl@ifknown@ttrib` An internal macro to check whether a given language/attribute is known. The macro takes 4 arguments, the language/attribute, the attribute list, the  $\TeX$ -code to be executed when the attribute is known and the  $\TeX$ -code to be executed otherwise.

```
1164 \def\bbl@ifknown@ttrib#1#2{%
```

We first assume the attribute is unknown.

```
1165 \let\bbl@tempa\@secondoftwo
```

Then we loop over the list of known attributes, trying to find a match.

```
1166 \bbl@loopx\bbl@tempb{#2}{%
```

```
1167 \expandafter\in@\expandafter{\expandafter,\bbl@tempb,}{, #1,}%
```

```
1168 \ifin@
```

When a match is found the definition of `\bbl@tempa` is changed.

```
1169 \let\bbl@tempa\@firstoftwo
```

```
1170 \else
```

```
1171 \fi}%
```

Finally we execute `\bbl@tempa`.

```
1172 \bbl@tempa
```

```
1173 }
```

`\bbl@clear@ttribs` This macro removes all the attribute code from  $\TeX$ 's memory at `\begin{document}` time (if any is present).

```
1174 \def\bbl@clear@ttribs{%
```

```
1175 \ifx\bbl@attributes\undefined\else
```

```
1176 \bbl@loopx\bbl@tempa{\bbl@attributes}{%
```

```
1177 \expandafter\bbl@clear@ttrib\bbl@tempa.
```

```
1178 }%
```

```
1179 \let\bbl@attributes\undefined
```

```
1180 \fi}
```

```
1181 \def\bbl@clear@ttrib#1-#2.{%
```

```
1182 \expandafter\let\csname#1@attr#2\endcsname\undefined}
```

```
1183 \AtBeginDocument{\bbl@clear@ttribs}
```

## 9.6 Support for saving macro definitions

To save the meaning of control sequences using `\babel@save`, we use temporary control sequences. To save hash table entries for these control sequences, we don't use the name of the control sequence to be saved to construct the temporary name. Instead we simply use the value of a counter, which is reset to zero each time we begin to save new values. This works well because we release the saved meanings before we begin to save a new set of control sequence meanings (see `\selectlanguage` and `\originalTeX`). Note undefined macros are not undefined any more when saved – they are `\relax`'ed.

`\babel@savecnt` The initialization of a new save cycle: reset the counter to zero.

```
\babel@beginsave 1184 \bbl@trace{Macros for saving definitions}
```

```
1185 \def\babel@beginsave{\babel@savecnt\z@}
```

Before it's forgotten, allocate the counter and initialize all.

```
1186 \newcount\babel@savecnt
```

```
1187 \babel@beginsave
```

`\babel@save` The macro `\babel@save<csname>` saves the current meaning of the control sequence `<csname>` to `\originalTeX`<sup>32</sup>. To do this, we let the current meaning to a temporary control

<sup>32</sup>`\originalTeX` has to be expandable, i.e. you shouldn't let it to `\relax`.

sequence, the restore commands are appended to `\originalTeX` and the counter is incremented.

```
1188 \def\babel@save#1{%
1189   \expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax
1190   \toks@\expandafter{\originalTeX\let#1=}%
1191   \bbl@exp{%
1192     \def\originalTeX{\the\toks@<\babel@number\babel@savecnt>\relax}}%
1193   \advance\babel@savecnt\@ne}
```

`\babel@savevariable` The macro `\babel@savevariable<variable>` saves the value of the variable. `<variable>` can be anything allowed after the `\the` primitive.

```
1194 \def\babel@savevariable#1{%
1195   \toks@\expandafter{\originalTeX #1=}%
1196   \bbl@exp{\def\originalTeX{\the\toks@the#1\relax}}}
```

`\bbl@frenchspacing` Some languages need to have `\frenchspacing` in effect. Others don't want that. The  
`\bbl@nonfrenchspacing` command `\bbl@frenchspacing` switches it on when it isn't already in effect and  
`\bbl@nonfrenchspacing` switches it off if necessary.

```
1197 \def\bbl@frenchspacing{%
1198   \ifnum\the\sfcodes\@m
1199     \let\bbl@nonfrenchspacing\relax
1200   \else
1201     \frenchspacing
1202     \let\bbl@nonfrenchspacing\nonfrenchspacing
1203   \fi}
1204 \let\bbl@nonfrenchspacing\nonfrenchspacing
```

## 9.7 Short tags

`\babeltags` This macro is straightforward. After zapping spaces, we loop over the list and define the macros `\text<tag>` and `\<tag>`. Definitions are first expanded so that they don't contain `\csname` but the actual macro.

```
1205 \bbl@trace{Short tags}
1206 \def\babeltags#1{%
1207   \edef\bbl@tempa{\zap@space#1 \@empty}%
1208   \def\bbl@tempb##1=##2\@{#}%
1209   \edef\bbl@tempc{%
1210     \noexpand\newcommand
1211     \expandafter\noexpand\csname ##1\endcsname{%
1212       \noexpand\protect
1213       \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
1214     \noexpand\newcommand
1215     \expandafter\noexpand\csname text##1\endcsname{%
1216       \noexpand\foreignlanguage{##2}}
1217   \bbl@tempc}%
1218   \bbl@for\bbl@tempa\bbl@tempa{%
1219     \expandafter\bbl@tempb\bbl@tempa\@{#}}
```

## 9.8 Hyphens

`\babelhyphenation` This macro saves hyphenation exceptions. Two macros are used to store them:  
`\bbl@hyphenation@` for the global ones and `\bbl@hyphenation<lang>` for language ones.  
 See `\bbl@patterns` above for further details. We make sure there is a space between  
 words when multiple commands are used.

```
1220 \bbl@trace{Hyphens}
```

```

1221 \@onlypreamble\babelhyphenation
1222 \AtEndOfPackage{%
1223   \newcommand\babelhyphenation[2][\@empty]{%
1224     \ifx\bbbl@hyphenation@relax
1225       \let\bbbl@hyphenation@\@empty
1226     \fi
1227     \ifx\bbbl@hyphlist\@empty\else
1228       \bbbl@warning{%
1229         You must not intermingle \string\selectlanguage\space and\%
1230         \string\babelhyphenation\space or some exceptions will not\%
1231         be taken into account. Reported}%
1232     \fi
1233     \ifx\@empty#1%
1234       \protected@edef\bbbl@hyphenation@{\bbbl@hyphenation@\space#2}%
1235     \else
1236       \bbbl@vforeach{#1}{%
1237         \def\bbbl@tempa{##1}%
1238         \bbbl@fixname\bbbl@tempa
1239         \bbbl@iflanguage\bbbl@tempa{%
1240           \bbbl@csarg\protected@edef{hyphenation@\bbbl@tempa}{%
1241             \bbbl@ifunset{bbbl@hyphenation@\bbbl@tempa}%
1242             \@empty
1243             {\csname bbl@hyphenation@\bbbl@tempa\endcsname\space}%
1244             #2}}}%
1245       \fi}}

```

`\bbbl@allowhyphens` This macro makes hyphenation possible. Basically its definition is nothing more than `\nobreak \hskip 0pt plus 0pt`<sup>33</sup>.

```

1246 \def\bbbl@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}
1247 \def\bbbl@t@one{T1}
1248 \def\allowhyphens{\ifx\cf@encoding\bbbl@t@one\else\bbbl@allowhyphens\fi}

```

`\babelhyphen` Macros to insert common hyphens. Note the space before `@` in `\babelhyphen`. Instead of protecting it with `\DeclareRobustCommand`, which could insert a `\relax`, we use the same procedure as shorthands, with `\active@prefix`.

```

1249 \newcommand\babelnullhyphen{\char\hyphenchar\font}
1250 \def\babelhyphen{\active@prefix\babelhyphen\bbbl@hyphen}
1251 \def\bbbl@hyphen{%
1252   \@ifstar{\bbbl@hyphen@i @}{\bbbl@hyphen@i\@empty}}
1253 \def\bbbl@hyphen@i#1#2{%
1254   \bbbl@ifunset{bbbl@hy#1#2\@empty}%
1255   {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{-}{#2}}}%
1256   {\csname bbl@hy#1#2\@empty\endcsname}}

```

The following two commands are used to wrap the “hyphen” and set the behavior of the rest of the word – the version with a single `@` is used when further hyphenation is allowed, while that with `@@` if no more hyphens are allowed. In both cases, if the hyphen is preceded by a positive space, breaking after the hyphen is disallowed.

There should not be a discretionary after a hyphen at the beginning of a word, so it is prevented if preceded by a skip. Unfortunately, this does handle cases like “(-suffix)”. `\nobreak` is always preceded by `\leavevmode`, in case the shorthand starts a paragraph.

```

1257 \def\bbbl@usehyphen#1{%
1258   \leavevmode
1259   \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
1260   \nobreak\hskip\z@skip}
1261 \def\bbbl@@usehyphen#1{%
1262   \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

<sup>33</sup> $\mathrm{T}_{\mathrm{E}}\mathrm{X}$  begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.



The following macro inserts the hyphen char.

```
1263 \def\bbl@hyphenchar{%
1264   \ifnum\hyphenchar\font=\m@ne
1265     \babeinullhyphen
1266   \else
1267     \char\hyphenchar\font
1268   \fi}
```

Finally, we define the hyphen “types”. Their names will not change, so you may use them in ldf’s. After a space, the \mbox in \bbl@hy@nbreak is redundant.

```
1269 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1270 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1271 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
1272 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
1273 \def\bbl@hy@nbreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}
1274 \def\bbl@hy@nbreak{\mbox{\bbl@hyphenchar}}
1275 \def\bbl@hy@repeat{%
1276   \bbl@usehyphen{%
1277     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
1278 \def\bbl@hy@repeat{%
1279   \bbl@usehyphen{%
1280     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
1281 \def\bbl@hy@empty{\hskip\z@skip}
1282 \def\bbl@hy@empty{\discretionary{}{}{}}
```

\bbl@disc For some languages the macro \bbl@disc is used to ease the insertion of discretionaries for letters that behave ‘abnormally’ at a breakpoint.

```
1283 \def\bbl@disc#1#2{\nbreak\discretionary{#2-}{#1}\bbl@allowhyphens}
```

## 9.9 Multiencoding strings

The aim following commands is to provide a common interface for strings in several encodings. They also contains several hooks which can be used by luatex and xetex. The code is organized here with pseudo-guards, so we start with the basic commands.

**Tools** But first, a couple of tools. The first one makes global a local variable. This is not the best solution, but it works.

```
1284 \bbl@trace{Multiencoding strings}
1285 \def\bbl@tglobal#1{\global\let#1#1}
1286 \def\bbl@recatcode#1{%
1287   \@tempcnta="7F
1288   \def\bbl@tempa{%
1289     \ifnum\@tempcnta>"FF\else
1290       \catcode\@tempcnta=#1\relax
1291       \advance\@tempcnta\@ne
1292       \expandafter\bbl@tempa
1293     \fi}%
1294   \bbl@tempa}
```

The second one. We need to patch \@uclclist, but it is done once and only if \SetCase is used or if strings are encoded. The code is far from satisfactory for several reasons, including the fact \@uclclist is not a list any more. Therefore a package option is added to ignore it. Instead of gobbling the macro getting the next two elements (usually \reserved@a), we pass it as argument to \bbl@uclc. The parser is restarted inside \(\lang)\bbl@uclc because we do not know how many expansions are necessary (depends on whether strings are encoded). The last part is tricky – when uppercasing, we have:

```
\let\bbl@tolower\@empty\bbl@toupper\@empty
```

and starts over (and similarly when lowercasing).

```
1295 \@ifpackagewith{babel}{nocase}%
1296 {\let\bbl@patchucl\relax}%
1297 {\def\bbl@patchucl{%
1298   \global\let\bbl@patchucl\relax
1299   \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@ucl}}%
1300   \gdef\bbl@ucl##1{%
1301     \let\bbl@encoded\bbl@encoded@ucl
1302     \bbl@ifunset{\language @bbl@ucl}% and resumes it
1303     {##1}%
1304     {\let\bbl@tempa##1\relax % Used by LANG@bbl@ucl
1305       \csname\language @bbl@ucl\endcsname}%
1306     {\bbl@tolower\@empty}{\bbl@toupper\@empty}}}%
1307   \gdef\bbl@tolower{\csname\language @bbl@lc\endcsname}%
1308   \gdef\bbl@toupper{\csname\language @bbl@uc\endcsname}}%
1309 <<(*More package options)>> \equiv
1310 \DeclareOption{nocase}{}
1311 <</More package options>>
```

The following package options control the behavior of `\SetString`.

```
1312 <<(*More package options)>> \equiv
1313 \let\bbl@opt@strings\@nnil % accept strings=value
1314 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1315 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1316 \def\BabelStringsDefault{generic}
1317 <</More package options>>
```

**Main command** This is the main command. With the first use it is redefined to omit the basic setup in subsequent blocks. We make sure strings contain actual letters in the range 128-255, not active characters.

```
1318 \@onlypreamble\StartBabelCommands
1319 \def\StartBabelCommands{%
1320   \begingroup
1321   \bbl@recatcode{11}%
1322   <<Macros local to BabelCommands>>
1323   \def\bbl@provstring##1##2{%
1324     \providecommand##1{##2}%
1325     \bbl@tglobal##1}%
1326   \global\let\bbl@scafter\@empty
1327   \let\StartBabelCommands\bbl@startcmds
1328   \ifx\BabelLanguages\relax
1329     \let\BabelLanguages\CurrentOption
1330   \fi
1331   \begingroup
1332   \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
1333   \StartBabelCommands}
1334 \def\bbl@startcmds{%
1335   \ifx\bbl@screset\@nnil\else
1336     \bbl@usehooks{stopcommands}{}%
1337   \fi
1338   \endgroup
1339   \begingroup
1340   \@ifstar
1341   {\ifx\bbl@opt@strings\@nnil
```

```

1342 \let\bbl@opt@strings\BabelStringsDefault
1343 \fi
1344 \bbl@startcmds@i}%
1345 \bbl@startcmds@i}
1346 \def\bbl@startcmds@i#1#2{%
1347 \edef\bbl@L{\zap@space#1 \@empty}%
1348 \edef\bbl@G{\zap@space#2 \@empty}%
1349 \bbl@startcmds@ii}

```

Parse the encoding info to get the label, input, and font parts.

Select the behavior of \SetString. There are two main cases, depending of if there is an optional argument: without it and strings=encoded, strings are defined always; otherwise, they are set only if they are still undefined (ie, fallback values). With labelled blocks and strings=encoded, define the strings, but with another value, define strings only if the current label or font encoding is the value of strings; otherwise (ie, no strings or a block whose label is not in strings=) do nothing.

We presume the current block is not loaded, and therefore set (above) a couple of default values to gobble the arguments. Then, these macros are redefined if necessary according to several parameters.

```

1350 \newcommand\bbl@startcmds@ii[1][\@empty]{%
1351 \let\SetString@gobbletwo
1352 \let\bbl@stringdef@gobbletwo
1353 \let\AfterBabelCommands@gobble
1354 \ifx\@empty#1%
1355 \def\bbl@sc@label{generic}%
1356 \def\bbl@encstring##1##2{%
1357 \ProvideTextCommandDefault##1{##2}%
1358 \bbl@tglobal##1%
1359 \expandafter\bbl@tglobal\csname\string?\string##1\endcsname}%
1360 \let\bbl@sctest\in@true
1361 \else
1362 \let\bbl@sc@charset\space % <- zapped below
1363 \let\bbl@sc@fontenc\space % <- " "
1364 \def\bbl@tempa##1=##2\@nil{%
1365 \bbl@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
1366 \bbl@vforeach{label=#1}{\bbl@tempa##1\@nil}%
1367 \def\bbl@tempa##1 ##2{% space -> comma
1368 ##1%
1369 \ifx\@empty##2\else\ifx,##1,\else,\fi\bbl@afterfi\bbl@tempa##2\fi}%
1370 \edef\bbl@sc@fontenc{\expandafter\bbl@tempa\bbl@sc@fontenc\@empty}%
1371 \edef\bbl@sc@label{\expandafter\zap@space\bbl@sc@label\@empty}%
1372 \edef\bbl@sc@charset{\expandafter\zap@space\bbl@sc@charset\@empty}%
1373 \def\bbl@encstring##1##2{%
1374 \bbl@foreach\bbl@sc@fontenc{%
1375 \bbl@ifunset{T####1}%
1376 {}%
1377 {\ProvideTextCommand##1{####1}{##2}%
1378 \bbl@tglobal##1%
1379 \expandafter
1380 \bbl@tglobal\csname####1\string##1\endcsname}}}%
1381 \def\bbl@sctest{%
1382 \bbl@xin@{\, \bbl@opt@strings,}{, \bbl@sc@label, \bbl@sc@fontenc,}}%
1383 \fi
1384 \ifx\bbl@opt@strings\@nnil % ie, no strings key -> defaults
1385 \else\ifx\bbl@opt@strings\relax % ie, strings=encoded
1386 \let\AfterBabelCommands\bbl@aftercmds
1387 \let\SetString\bbl@setstring
1388 \let\bbl@stringdef\bbl@encstring

```

```

1389 \else % ie, strings=value
1390 \bbl@sctest
1391 \ifin@
1392 \let\AfterBabelCommands\bbl@aftercmds
1393 \let\SetString\bbl@setstring
1394 \let\bbl@stringdef\bbl@provstring
1395 \fi\fi\fi
1396 \bbl@scswitch
1397 \ifx\bbl@G\@empty
1398 \def\SetString##1##2{%
1399 \bbl@error{Missing group for string \string##1}%
1400 {You must assign strings to some category, typically\\%
1401 captions or extras, but you set none}}%
1402 \fi
1403 \ifx\@empty#1%
1404 \bbl@usehooks{defaultcommands}{}%
1405 \else
1406 \@expandtwoargs
1407 \bbl@usehooks{encodedcommands}{\bbl@sc@charset}{\bbl@sc@fontenc}}%
1408 \fi}

```

There are two versions of `\bbl@scswitch`. The first version is used when `ldfs` are read, and it makes sure `\langle group \rangle \langle language \rangle` is reset, but only once (`\bbl@screset` is used to keep track of this). The second version is used in the preamble and packages loaded after `babel` and does nothing. The macro `\bbl@forlang` loops `\bbl@L` but its body is executed only if the value is in `\BabelLanguages` (inside `babel`) or `\date \langle language \rangle` is defined (after `babel` has been loaded). There are also two version of `\bbl@forlang`. The first one skips the current iteration if the language is not in `\BabelLanguages` (used in `ldfs`), and the second one skips undefined languages (after `babel` has been loaded).

```

1409 \def\bbl@forlang#1#2{%
1410 \bbl@for#1\bbl@L{%
1411 \bbl@xin@{, #1,}{, \BabelLanguages,}%
1412 \ifin@#2\relax\fi}}
1413 \def\bbl@scswitch{%
1414 \bbl@forlang\bbl@tempa{%
1415 \ifx\bbl@G\@empty\else
1416 \ifx\SetString\@gobbletwo\else
1417 \edef\bbl@GL{\bbl@G\bbl@tempa}%
1418 \bbl@xin@{, \bbl@GL,}{, \bbl@screset,}%
1419 \ifin@\else
1420 \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
1421 \xdef\bbl@screset{\bbl@screset, \bbl@GL}%
1422 \fi
1423 \fi
1424 \fi}}
1425 \AtEndOfPackage{%
1426 \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{}{#2}}}%
1427 \let\bbl@scswitch\relax}
1428 \@onlypreamble\EndBabelCommands
1429 \def\EndBabelCommands{%
1430 \bbl@usehooks{stopcommands}{}%
1431 \endgroup
1432 \endgroup
1433 \bbl@scafter}

```

Now we define commands to be used inside `\StartBabelCommands`.

**Strings** The following macro is the actual definition of `\SetString` when it is “active”

First save the “switcher”. Create it if undefined. Strings are defined only if undefined (ie, like `\providescommand`). With the event stringprocess you can preprocess the string by manipulating the value of `\BabelString`. If there are several hooks assigned to this event, preprocessing is done in the same order as defined. Finally, the string is set.

```

1434 \def\bbl@setstring#1#2{%
1435   \bbl@forlang\bbl@tempa{%
1436     \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
1437     \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
1438     {\global\expandafter % TODO - con \bbl@exp ?
1439      \bbl@add\csname\bbl@G\bbl@tempa\expandafter\endcsname\expandafter
1440       {\expandafter\bbl@scset\expandafter#1\csname\bbl@LC\endcsname}}}%
1441     }%
1442   \def\BabelString{#2}%
1443   \bbl@usehooks{stringprocess}{}%
1444   \expandafter\bbl@stringdef
1445     \csname\bbl@LC\expandafter\endcsname\expandafter{\BabelString}}

```

Now, some additional stuff to be used when encoded strings are used. Captions then include `\bbl@encoded` for string to be expanded in case transformations. It is `\relax` by default, but in `\MakeUppercase` and `\MakeLowercase` its value is a modified expandable `\@changed@cmd`.

```

1446 \ifx\bbl@opt@strings\relax
1447   \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
1448   \bbl@patchuclc
1449   \let\bbl@encoded\relax
1450   \def\bbl@encoded@uclc#1{%
1451     \@inmathwarn#1%
1452     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
1453       \expandafter\ifx\csname ?\string#1\endcsname\relax
1454         \TextSymbolUnavailable#1%
1455       \else
1456         \csname ?\string#1\endcsname
1457       \fi
1458     \else
1459       \csname\cf@encoding\string#1\endcsname
1460     \fi}
1461 \else
1462   \def\bbl@scset#1#2{\def#1{#2}}
1463 \fi

```

Define `\SetStringLoop`, which is actually set inside `\StartBabelCommands`. The current definition is somewhat complicated because we need a count, but `\count@` is not under our control (remember `\SetString` may call hooks). Instead of defining a dedicated count, we just “pre-expand” its value.

```

1464 <<*Macros local to BabelCommands>> ≡
1465 \def\SetStringLoop##1##2{%
1466   \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
1467   \count@\z@
1468   \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
1469     \advance\count@\@ne
1470     \toks@\expandafter{\bbl@tempa}%
1471     \bbl@exp{%
1472       \\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
1473       \count@=\the\count@\relax}}}%
1474 <</Macros local to BabelCommands>>

```

**Delaying code** Now the definition of `\AfterBabelCommands` when it is activated.

```

1475 \def\bbl@aftercmds#1{%
1476   \toks@\expandafter{\bbl@scafter#1}%
1477   \xdef\bbl@scafter{\the\toks@}}

```

**Case mapping** The command `\SetCase` provides a way to change the behavior of `\MakeUppercase` and `\MakeLowercase`. `\bbl@tempa` is set by the patched `\@uclclist` to the parsing command.

```

1478 <<(*Macros local to BabelCommands)>> ≡
1479   \newcommand\SetCase[3][]{%
1480     \bbl@patchuclc
1481     \bbl@forlang\bbl@tempa{%
1482       \expandafter\bbl@encstring
1483       \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
1484       \expandafter\bbl@encstring
1485       \csname\bbl@tempa @bbl@uc\endcsname{##2}%
1486       \expandafter\bbl@encstring
1487       \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
1488 <</Macros local to BabelCommands>>

```

Macros to deal with case mapping for hyphenation. To decide if the document is monolingual or multilingual, we make a rough guess – just see if there is a comma in the languages list, built in the first pass of the package options.

```

1489 <<(*Macros local to BabelCommands)>> ≡
1490   \newcommand\SetHyphenMap[1]{%
1491     \bbl@forlang\bbl@tempa{%
1492       \expandafter\bbl@stringdef
1493       \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
1494 <</Macros local to BabelCommands>>

```

There are 3 helper macros which do most of the work for you.

```

1495 \newcommand\BabelLower[2]{% one to one.
1496   \ifnum\lccode#1=#2\else
1497     \babel@savevariable{\lccode#1}%
1498     \lccode#1=#2\relax
1499   \fi}
1500 \newcommand\BabelLowerMM[4]{% many-to-many
1501   \@tempcnta=#1\relax
1502   \@tempcntb=#4\relax
1503   \def\bbl@tempa{%
1504     \ifnum\@tempcnta>#2\else
1505       \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
1506       \advance\@tempcnta#3\relax
1507       \advance\@tempcntb#3\relax
1508       \expandafter\bbl@tempa
1509     \fi}%
1510   \bbl@tempa}
1511 \newcommand\BabelLowerMO[4]{% many-to-one
1512   \@tempcnta=#1\relax
1513   \def\bbl@tempa{%
1514     \ifnum\@tempcnta>#2\else
1515       \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
1516       \advance\@tempcnta#3
1517       \expandafter\bbl@tempa
1518     \fi}%
1519   \bbl@tempa}

```

The following package options control the behavior of hyphenation mapping.

```

1520 <<(*More package options)>> ≡

```

```

1521 \DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
1522 \DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\ne@}
1523 \DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
1524 \DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@@}
1525 \DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}
1526 <</More package options>>

```

Initial setup to provide a default behavior if hyphenmap is not set.

```

1527 \AtEndOfPackage{%
1528   \ifx\bbl@opt@hyphenmap\undefined
1529     \bbl@xin@{,}{\bbl@language@opts}%
1530     \chardef\bbl@opt@hyphenmap\ifin4\else\ne\fi
1531   \fi}

```

## 9.10 Macros common to a number of languages

`\set@low@box` The following macro is used to lower quotes to the same level as the comma. It prepares its argument in box register 0.

```

1532 \bbl@trace{Macros related to glyphs}
1533 \def\set@low@box#1{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}%
1534   \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@%
1535   \setbox\z@\hbox{\lower\dimen\z@ \box\z@}\ht\z@\ht\tw@ \dp\z@\dp\tw@}

```

`\save@sf@q` The macro `\save@sf@q` is used to save and reset the current space factor.

```

1536 \def\save@sf@q#1{\leavevmode
1537   \begingroup
1538   \edef\@SF{\spacefactor\the\spacefactor}\@SF
1539   \endgroup}

```

## 9.11 Making glyphs available

This section makes a number of glyphs available that either do not exist in the OT1 encoding and have to be ‘faked’, or that are not accessible through `T1enc.def`.

### 9.11.1 Quotation marks

`\quotedblbase` In the T1 encoding the opening double quote at the baseline is available as a separate character, accessible via `\quotedblbase`. In the OT1 encoding it is not available, therefore we make it available by lowering the normal open quote character to the baseline.

```

1540 \ProvideTextCommand{\quotedblbase}{OT1}{%
1541   \save@sf@q{\set@low@box{\textquotedblright\}}%
1542   \box\z@\kern-.04em\bbl@allowhyphens}}

```

Make sure that when an encoding other than OT1 or T1 is used this glyph can still be typeset.

```

1543 \ProvideTextCommandDefault{\quotedblbase}{%
1544   \UseTextSymbol{OT1}{\quotedblbase}}

```

`\quotesinglbase` We also need the single quote character at the baseline.

```

1545 \ProvideTextCommand{\quotesinglbase}{OT1}{%
1546   \save@sf@q{\set@low@box{\textquoteright\}}%
1547   \box\z@\kern-.04em\bbl@allowhyphens}}

```

Make sure that when an encoding other than OT1 or T1 is used this glyph can still be typeset.

```

1548 \ProvideTextCommandDefault{\quotesinglbase}{%
1549   \UseTextSymbol{OT1}{\quotesinglbase}}

```

`\guillemotleft` The guillemet characters are not available in OT1 encoding. They are faked.

```
\guillemotright 1550 \ProvideTextCommand{\guillemotleft}{OT1}{%
1551   \ifmmode
1552     \ll
1553   \else
1554     \save@sf@q{\nobreak
1555       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bbl@allowhyphens}%
1556   \fi}
1557 \ProvideTextCommand{\guillemotright}{OT1}{%
1558   \ifmmode
1559     \gg
1560   \else
1561     \save@sf@q{\nobreak
1562       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bbl@allowhyphens}%
1563   \fi}
```

Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.

```
1564 \ProvideTextCommandDefault{\guillemotleft}{%
1565   \UseTextSymbol{OT1}{\guillemotleft}}
1566 \ProvideTextCommandDefault{\guillemotright}{%
1567   \UseTextSymbol{OT1}{\guillemotright}}
```

`\guilsinglleft` The single guillemets are not available in OT1 encoding. They are faked.

```
\guilsinglright 1568 \ProvideTextCommand{\guilsinglleft}{OT1}{%
1569   \ifmmode
1570     <%
1571   \else
1572     \save@sf@q{\nobreak
1573       \raise.2ex\hbox{$\scriptscriptstyle<$}\bbl@allowhyphens}%
1574   \fi}
1575 \ProvideTextCommand{\guilsinglright}{OT1}{%
1576   \ifmmode
1577     >%
1578   \else
1579     \save@sf@q{\nobreak
1580       \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
1581   \fi}
```

Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.

```
1582 \ProvideTextCommandDefault{\guilsinglleft}{%
1583   \UseTextSymbol{OT1}{\guilsinglleft}}
1584 \ProvideTextCommandDefault{\guilsinglright}{%
1585   \UseTextSymbol{OT1}{\guilsinglright}}
```

### 9.11.2 Letters

`\ij` The dutch language uses the letter ‘ij’. It is available in T1 encoded fonts, but not in the OT1 encoded fonts. Therefore we fake it for the OT1 encoding.

```
1586 \DeclareTextCommand{\ij}{OT1}{%
1587   i\kern-0.02em\bbl@allowhyphens j}
1588 \DeclareTextCommand{\IJ}{OT1}{%
1589   I\kern-0.02em\bbl@allowhyphens J}
1590 \DeclareTextCommand{\ij}{T1}{\char188}
1591 \DeclareTextCommand{\IJ}{T1}{\char156}
```



Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.

```
1592 \ProvideTextCommandDefault{\ij}{%
1593   \UseTextSymbol{OT1}{\ij}}
1594 \ProvideTextCommandDefault{\IJ}{%
1595   \UseTextSymbol{OT1}{\IJ}}
```

`\dj` The croatian language needs the letters `\dj` and `\DJ`; they are available in the T1 encoding, `\DJ` but not in the OT1 encoding by default.

Some code to construct these glyphs for the OT1 encoding was made available to me by Stipčević Mario, (stipcevic@olimp.irb.hr).

```
1596 \def\crrtic@{\hrule height0.1ex width0.3em}
1597 \def\crttic@{\hrule height0.1ex width0.33em}
1598 \def\ddj@{%
1599   \setbox0\hbox{d}\dimen@=\ht0
1600   \advance\dimen@1ex
1601   \dimen@.45\dimen@
1602   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1603   \advance\dimen@ii.5ex
1604   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
1605 \def\DDJ@{%
1606   \setbox0\hbox{D}\dimen@=.55\ht0
1607   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1608   \advance\dimen@ii.15ex % correction for the dash position
1609   \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
1610   \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
1611   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
1612 %
1613 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
1614 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}
```

Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.

```
1615 \ProvideTextCommandDefault{\dj}{%
1616   \UseTextSymbol{OT1}{\dj}}
1617 \ProvideTextCommandDefault{\DJ}{%
1618   \UseTextSymbol{OT1}{\DJ}}
```

`\SS` For the T1 encoding `\SS` is defined and selects a specific glyph from the font, but for other encodings it is not available. Therefore we make it available here.

```
1619 \DeclareTextCommand{\SS}{OT1}{SS}
1620 \ProvideTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}
```

### 9.11.3 Shorthands for quotation marks

Shorthands are provided for a number of different quotation marks, which make them usable both outside and inside mathmode. They are defined with `\ProvideTextCommandDefault`, but this is very likely not required because their definitions are based on encoding-dependent macros.

`\glq` The ‘german’ single quotes.

```
\grq 1621 \ProvideTextCommandDefault{\glq}{%
1622   \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}
```

The definition of `\grq` depends on the fontencoding. With T1 encoding no extra kerning is needed.

```

1623 \ProvideTextCommand{\grq}{T1}{%
1624   \textormath{\kern\z@\textquoteleft}{\mbox{\textquoteleft}}}
1625 \ProvideTextCommand{\grq}{TU}{%
1626   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
1627 \ProvideTextCommand{\grq}{OT1}{%
1628   \save@sf@q{\kern-.0125em
1629     \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
1630     \kern.07em\relax}}
1631 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}

```

`\glqq` The ‘german’ double quotes.

```

\grqq 1632 \ProvideTextCommandDefault{\glqq}{%
1633   \textormath{\quotedblbase}{\mbox{\quotedblbase}}}

```

The definition of `\grqq` depends on the fontencoding. With T1 encoding no extra kerning is needed.

```

1634 \ProvideTextCommand{\grqq}{T1}{%
1635   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1636 \ProvideTextCommand{\grqq}{TU}{%
1637   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1638 \ProvideTextCommand{\grqq}{OT1}{%
1639   \save@sf@q{\kern-.07em
1640     \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
1641     \kern.07em\relax}}
1642 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}

```

`\flq` The ‘french’ single guillemets.

```

\frq 1643 \ProvideTextCommandDefault{\flq}{%
1644   \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
1645 \ProvideTextCommandDefault{\frq}{%
1646   \textormath{\guilsinglright}{\mbox{\guilsinglright}}}

```

`\flqq` The ‘french’ double guillemets.

```

\frqq 1647 \ProvideTextCommandDefault{\flqq}{%
1648   \textormath{\guillemotleft}{\mbox{\guillemotleft}}}
1649 \ProvideTextCommandDefault{\frqq}{%
1650   \textormath{\guillemotright}{\mbox{\guillemotright}}}

```

#### 9.11.4 Umlauts and tremas

The command `\"` needs to have a different effect for different languages. For German for instance, the ‘umlaut’ should be positioned lower than the default position for placing it over the letters a, o, u, A, O and U. When placed over an e, i, E or I it can retain its normal position. For Dutch the same glyph is always placed in the lower position.

`\umlauthigh` To be able to provide both positions of `\"` we provide two commands to switch the  
`\umlautlow` positioning, the default will be `\umlauthigh` (the normal positioning).

```

1651 \def\umlauthigh{%
1652   \def\bb1@umlauta##1{\leavevmode\bggroup%
1653     \expandafter\accent\csname\fontencoding dqpos\endcsname
1654     ##1\bb1@allowhyphens\egroup}%
1655   \let\bb1@umlaute\bb1@umlauta}
1656 \def\umlautlow{%
1657   \def\bb1@umlauta{\protect\lower@umlaut}}
1658 \def\umlautelow{%
1659   \def\bb1@umlaute{\protect\lower@umlaut}}
1660 \umlauthigh

```

`\lower@umlaut` The command `\lower@umlaut` is used to position the `\` closer to the letter. We want the umlaut character lowered, nearer to the letter. To do this we need an extra *<dimen>* register.

```
1661 \expandafter\ifx\csname U@D\endcsname\relax
1662   \csname newdimen\endcsname\U@D
1663 \fi
```

The following code fools T<sub>E</sub>X's `make_accent` procedure about the current x-height of the font to force another placement of the umlaut character. First we have to save the current x-height of the font, because we'll change this font dimension and this is always done globally.

Then we compute the new x-height in such a way that the umlaut character is lowered to the base character. The value of `.45ex` depends on the METAFONT parameters with which the fonts were built. (Just try out, which value will look best.) If the new x-height is too low, it is not changed. Finally we call the `\accent` primitive, reset the old x-height and insert the base character in the argument.

```
1664 \def\lower@umlaut#1{%
1665   \leavevmode\bggroup
1666     \U@D 1ex%
1667   {\setbox\z@\hbox{%
1668     \expandafter\char\csname\fontencoding dqpos\endcsname}%
1669     \dimen@ -.45ex\advance\dimen@\ht\z@
1670     \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
1671   \expandafter\accent\csname\fontencoding dqpos\endcsname
1672   \fontdimen5\font\U@D #1%
1673 \egroup}
```

For all vowels we declare `\` to be a composite command which uses `\bbl@umlauta` or `\bbl@umlaute` to position the umlaut character. We need to be sure that these definitions override the ones that are provided when the package `fontenc` with option `OT1` is used. Therefore these declarations are postponed until the beginning of the document. Note these definitions only apply to some languages, but `babel` sets them for *all* languages – you may want to redefine `\bbl@umlauta` and/or `\bbl@umlaute` for a language in the corresponding `ldf` (using the `babel` switching mechanism, of course).

```
1674 \AtBeginDocument{%
1675   \DeclareTextCompositeCommand{\}{OT1}{a}{\bbl@umlauta{a}}%
1676   \DeclareTextCompositeCommand{\}{OT1}{e}{\bbl@umlaute{e}}%
1677   \DeclareTextCompositeCommand{\}{OT1}{i}{\bbl@umlaute{~i}}%
1678   \DeclareTextCompositeCommand{\}{OT1}{~i}{\bbl@umlaute{~i}}%
1679   \DeclareTextCompositeCommand{\}{OT1}{o}{\bbl@umlauta{o}}%
1680   \DeclareTextCompositeCommand{\}{OT1}{u}{\bbl@umlauta{u}}%
1681   \DeclareTextCompositeCommand{\}{OT1}{A}{\bbl@umlauta{A}}%
1682   \DeclareTextCompositeCommand{\}{OT1}{E}{\bbl@umlaute{E}}%
1683   \DeclareTextCompositeCommand{\}{OT1}{I}{\bbl@umlaute{I}}%
1684   \DeclareTextCompositeCommand{\}{OT1}{O}{\bbl@umlauta{O}}%
1685   \DeclareTextCompositeCommand{\}{OT1}{U}{\bbl@umlauta{U}}%
1686 }
```

Finally, the default is to use English as the main language.

```
1687 \ifx\l@english\undefined
1688   \chardef\l@english\z@
1689 \fi
1690 \main@language{english}
```

## 9.12 Layout

**Work in progress.**

Layout is mainly intended to set bidi documents, but there is at least a tool useful in general.

```

1691 \bbl@trace{Bidi layout}
1692 \providecommand\IfBabelLayout[3]{#3}%
1693 \newcommand\BabelPatchSection[1]{%
1694   \@ifundefined{#1}{}{%
1695     \bbl@exp{\let\<bbl@ss@#1>\<#1>}%
1696     \@namedef{#1}{%
1697       \ifstar{\bbl@presec@s{#1}}%
1698       {\@dblarg{\bbl@presec@x{#1}}}}}%
1699 \def\bbl@presec@x#1[#2]#3{%
1700   \bbl@exp{%
1701     \\\select@language@x{\bbl@main@language}%
1702     \\\@nameuse{\bbl@sspre@#1}%
1703     \\\@nameuse{\bbl@ss@#1}%
1704     [\\foreignlanguage{\language}{\unexpanded{#2}}}%
1705     {\\foreignlanguage{\language}{\unexpanded{#3}}}%
1706     \\\select@language@x{\language}}}%
1707 \def\bbl@presec@s#1#2{%
1708   \bbl@exp{%
1709     \\\select@language@x{\bbl@main@language}%
1710     \\\@nameuse{\bbl@sspre@#1}%
1711     \\\@nameuse{\bbl@ss@#1}*%
1712     {\\foreignlanguage{\language}{\unexpanded{#2}}}%
1713     \\\select@language@x{\language}}}%
1714 \IfBabelLayout{sectioning}%
1715   {\BabelPatchSection{part}%
1716   \BabelPatchSection{chapter}%
1717   \BabelPatchSection{section}%
1718   \BabelPatchSection{subsection}%
1719   \BabelPatchSection{subsubsection}%
1720   \BabelPatchSection{paragraph}%
1721   \BabelPatchSection{subparagraph}}%
1722   \def\babel@toc#1{%
1723     \select@language@x{\bbl@main@language}}}%
1724 \IfBabelLayout{captions}%
1725   {\BabelPatchSection{caption}}}%

```

### 9.13 Load engine specific macros

```

1726 \bbl@trace{Input engine specific macros}
1727 \ifcase\bbl@engine
1728   \input txtbabel.def
1729 \or
1730   \input luababel.def
1731 \or
1732   \input xebabel.def
1733 \fi

```

### 9.14 Creating languages

\babelprovide is a general purpose tool for creating and modifying languages. It creates the language infrastructure, and loads, if requested, an ini file. It may be used in conjunction to previously loaded ldf files.

```

1734 \bbl@trace{Creating languages and reading ini files}
1735 \newcommand\babelprovide[2][{}]{%
1736   \let\bbl@savelangname\language
1737   \edef\bbl@savlocaleid{\the\localeid}%

```

```

1738 % Set name and locale id
1739 \def\languagename{#2}%
1740 \bbl@id@assign
1741 \let\bbl@KVP@captions\@nil
1742 \let\bbl@KVP@import\@nil
1743 \let\bbl@KVP@main\@nil
1744 \let\bbl@KVP@script\@nil
1745 \let\bbl@KVP@language\@nil
1746 \let\bbl@KVP@hyphenrules\@nil % only for provide@new
1747 \let\bbl@KVP@mapfont\@nil
1748 \let\bbl@KVP@maparabic\@nil
1749 \let\bbl@KVP@mapdigits\@nil
1750 \let\bbl@KVP@intraspace\@nil
1751 \let\bbl@KVP@intrapenalty\@nil
1752 \bbl@forkv{#1}{% TODO - error handling
1753   \in@{.}{##1}%
1754   \ifin@
1755     \bbl@renewinikey##1\@{##2}%
1756   \else
1757     \bbl@csarg\def{KVP@##1}{##2}%
1758   \fi}%
1759 % == import, captions ==
1760 \ifx\bbl@KVP@import\@nil\else
1761   \bbl@exp{\bbl@ifblank{\bbl@KVP@import}}%
1762   {\begingroup
1763     \def\BabelBeforeIni##1##2{\gdef\bbl@KVP@import{##1}\endinput}%
1764     \InputIfFileExists{babel-#2.tex}{}}%
1765   \endgroup}%
1766   {}%
1767 \fi
1768 \ifx\bbl@KVP@captions\@nil
1769   \let\bbl@KVP@captions\bbl@KVP@import
1770 \fi
1771 % Load ini
1772 \bbl@ifunset{date#2}%
1773   {\bbl@provide@new{#2}}%
1774   {\bbl@ifblank{#1}%
1775     {\bbl@error
1776       {If you want to modify `#2' you must tell how in\\
1777         the optional argument. See the manual for the\\
1778         available options.}%
1779       {Use this macro as documented}}%
1780     {\bbl@provide@renew{#2}}}%
1781 % Post tasks
1782 \bbl@exp{\bbl@babelensure[exclude=\\today]{#2}}%
1783 \bbl@ifunset{\bbl@ensure@\languagename}%
1784   {\bbl@exp{%
1785     \\DeclareRobustCommand\<bbl@ensure@\languagename>[1]{%
1786       \\foreignlanguage{\languagename}%
1787       {###1}}}%
1788   {}%
1789 % At this point all parameters are defined if 'import'. Now we
1790 % execute some code depending on them. But what about if nothing was
1791 % imported? We just load the very basic parameters: ids and a few
1792 % more.
1793 \bbl@ifunset{\bbl@lname@#2}%
1794   {\def\BabelBeforeIni##1##2{%
1795     \begingroup
1796     \catcode`\[12 \catcode`\]=12 \catcode`\==12 %

```

```

1797 \let\bbl@ini@captions@aux\@gobbletwo
1798 \def\bbl@inidate ###1.###2.###3.###4\relax ###5###6{%
1799 \bbl@read@ini{##1}{basic data}%
1800 \bbl@exportkey{chrng}{characters.ranges}{}%
1801 \bbl@exportkey{dgnat}{numbers.digits.native}{}%
1802 \bbl@exportkey{hyphr}{typography.hyphenrules}{}%
1803 \bbl@exportkey{intsp}{typography.intraspaces}{}%
1804 \endgroup}% boxed, to avoid extra spaces:
1805 {\setbox\z@\hbox{\InputIfFileExists{babel-#2.tex}{}}}%
1806 }%
1807 % -
1808 % == script, language ==
1809 % Override the values from ini or defines them
1810 \ifx\bbl@KVP@script\@nil\else
1811 \bbl@csarg\edef{sname@#2}{\bbl@KVP@script}%
1812 \fi
1813 \ifx\bbl@KVP@language\@nil\else
1814 \bbl@csarg\edef{lname@#2}{\bbl@KVP@language}%
1815 \fi
1816 % == mapfont ==
1817 % For bidi texts, to switch the font based on direction
1818 \ifx\bbl@KVP@mapfont\@nil\else
1819 \bbl@ifsamestring{\bbl@KVP@mapfont}{direction}{}%
1820 {\bbl@error{Option '\bbl@KVP@mapfont' unknown for\%
1821 mapfont. Use 'direction'.%
1822 {See the manual for details.}}}%
1823 \bbl@ifunset{\bbl@lsys@language}{\bbl@provide@lsys{language}}}%
1824 \bbl@ifunset{\bbl@wdir@language}{\bbl@provide@dirs{language}}}%
1825 \ifx\bbl@mapselect\@undefined
1826 \AtBeginDocument{%
1827 \expandafter\bbl@add\csname selectfont \endcsname{\bbl@mapselect}%
1828 {\selectfont}}%
1829 \def\bbl@mapselect{%
1830 \let\bbl@mapselect\relax
1831 \edef\bbl@prefontid{\fontid\font}%
1832 \def\bbl@mapdir##1{%
1833 {\def\language{##1}%
1834 \let\bbl@ifrestoring\@firstoftwo % avoid font warning
1835 \bbl@switchfont
1836 \directlua{Babel.fontmap
1837 [\the\csname bbl@wdir##1\endcsname]%
1838 [\bbl@prefontid]=\fontid\font}}}%
1839 \fi
1840 \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{language}}}%
1841 \fi
1842 % == intraspaces, intrapenalty ==
1843 % For CJK, East Asian, Southeast Asian, if interspace in ini
1844 \ifx\bbl@KVP@intraspaces\@nil\else % We can override the ini or set
1845 \bbl@csarg\edef{intsp@#2}{\bbl@KVP@intraspaces}%
1846 \fi
1847 \bbl@provide@intraspaces
1848 % == maparabic ==
1849 % Native digits, if provided in ini (TeX level, xe and lua)
1850 \ifcase\bbl@engine\else
1851 \bbl@ifunset{\bbl@dgnat@language}{}%
1852 {\expandafter\ifx\csname bbl@dgnat@language\endcsname\@empty\else
1853 \expandafter\expandafter\expandafter
1854 \bbl@setdigits\csname bbl@dgnat@language\endcsname
1855 \ifx\bbl@KVP@maparabic\@nil\else

```

```

1856         \ifx\bbl@latinarabic\@undefined
1857         \expandafter\let\expandafter\@arabic
1858             \csname bbl@counter@\languagename\endcsname
1859         \else % ie, if layout=counters, which redefines \@arabic
1860             \expandafter\let\expandafter\bbl@latinarabic
1861                 \csname bbl@counter@\languagename\endcsname
1862         \fi
1863     \fi
1864 \fi}%
1865 \fi
1866 % == mapdigits ==
1867 % Native digits (lua level).
1868 \ifodd\bbl@engine
1869     \ifx\bbl@KVP@mapdigits\@nil\else
1870         \bbl@ifunset{bbl@dgnat@\languagename}{}%
1871         {\RequirePackage{luatexbase}%
1872          \bbl@activate@preotf
1873          \directlua{
1874              Babel = Babel or {} %% -> presets in luababel
1875              Babel.digits_mapped = true
1876              Babel.digits = Babel.digits or {}
1877              Babel.digits[\the\localeid] =
1878                  table.pack(string.utfvalue('\bbl@cs{dgnat@\languagename}'))
1879              if not Babel.numbers then
1880                  function Babel.numbers(head)
1881                      local LOCALE = luatexbase.registernumber'bbl@attr@locale'
1882                      local GLYPH = node.id'glyph'
1883                      local inmath = false
1884                      for item in node.traverse(head) do
1885                          if not inmath and item.id == GLYPH then
1886                              local temp = node.get_attribute(item, LOCALE)
1887                              if Babel.digits[temp] then
1888                                  local chr = item.char
1889                                  if chr > 47 and chr < 58 then
1890                                      item.char = Babel.digits[temp][chr-47]
1891                                  end
1892                              end
1893                          elseif item.id == node.id'math' then
1894                              inmath = (item.subtype == 0)
1895                          end
1896                      end
1897                      return head
1898                  end
1899              end
1900          }}
1901     \fi
1902 \fi
1903 % == require.babel in ini ==
1904 % To load or reload the babel-*.tex, if require.babel in ini
1905 \bbl@ifunset{bbl@rqtex@\languagename}{}%
1906     {\expandafter\ifx\csname bbl@rqtex@\languagename\endcsname\@empty\else
1907         \let\BabelBeforeIni\@gobbletwo
1908         \chardef\atcatcode=\catcode`\@
1909         \catcode`\@=11\relax
1910         \InputIfFileExists{babel-\bbl@cs{rqtex@\languagename}.tex}{}%
1911         \catcode`\@=\atcatcode
1912         \let\atcatcode\relax
1913     \fi}%
1914 % == main ==

```

A tool to define the macros for native digits from the list provided in the ini file. Somewhat convoluted because there are 10 digits, but only 9 arguments in T<sub>E</sub>X.

Depending on whether or not the language exists, we define two macros.

103



```

1968 \StartBabelCommands*{#1}{date}%
1969 \ifx\bbbl@KVP@import\@nil
1970 \bbbl@exp{%
1971 \\\SetString\\today{\\bbbl@nocaption{today}{#1today}}}%
1972 \else
1973 \bbbl@savetoday
1974 \bbbl@savestate
1975 \fi
1976 \EndBabelCommands
1977 \bbbl@exp{%
1978 \def\<#1hyphenmins>{%
1979 {\bbbl@ifunset{bbbl@lfthm@#1}{2}{\@nameuse{bbbl@lfthm@#1}}}%
1980 {\bbbl@ifunset{bbbl@rgthm@#1}{3}{\@nameuse{bbbl@rgthm@#1}}}}}%
1981 \bbbl@provide@hyphens{#1}%
1982 \ifx\bbbl@KVP@main\@nil\else
1983 \expandafter\main@language\expandafter{#1}%
1984 \fi}
1985 \def\bbbl@provide@renew#1{%
1986 \ifx\bbbl@KVP@captions\@nil\else
1987 \StartBabelCommands*{#1}{captions}%
1988 \bbbl@read@ini{\bbbl@KVP@captions}{data}% Here all letters cat = 11
1989 \bbbl@after@ini
1990 \bbbl@savestrings
1991 \EndBabelCommands
1992 \fi
1993 \ifx\bbbl@KVP@import\@nil\else
1994 \StartBabelCommands*{#1}{date}%
1995 \bbbl@savetoday
1996 \bbbl@savestate
1997 \EndBabelCommands
1998 \fi
1999 % == hyphenrules ==
2000 \bbbl@provide@hyphens{#1}}

```

The hyphenrules option is handled with an auxiliary macro.

```

2001 \def\bbbl@provide@hyphens#1{%
2002 \let\bbbl@tempa\relax
2003 \ifx\bbbl@KVP@hyphenrules\@nil\else
2004 \bbbl@replace\bbbl@KVP@hyphenrules{ }{,}%
2005 \bbbl@foreach\bbbl@KVP@hyphenrules{%
2006 \ifx\bbbl@tempa\relax % if not yet found
2007 \bbbl@ifsamestring{##1}{+}%
2008 {\bbbl@exp{\\addlanguage\<l@##1>}}}%
2009 }%
2010 \bbbl@ifunset{l@##1}%
2011 }%
2012 {\bbbl@exp{\let\bbbl@tempa\<l@##1>}}}%
2013 \fi}%
2014 \fi
2015 \ifx\bbbl@tempa\relax % if no opt or no language in opt found
2016 \ifx\bbbl@KVP@import\@nil\else % if importing
2017 \bbbl@exp{%
2018 \\\bbbl@ifblank{\@nameuse{bbbl@hyphr@#1}}}%
2019 }%
2020 {\let\\bbbl@tempa\<l@\@nameuse{bbbl@hyphr@language}>}}}%
2021 \fi
2022 \fi
2023 \bbbl@ifunset{bbbl@tempa}% ie, relax or undefined
2024 {\bbbl@ifunset{l@##1}% no hyphenrules found - fallback

```

```

2025     {\bbl@exp{\adddialect\<l@#1>\language}}%
2026     {}}%                               so, l@<lang> is ok - nothing to do
2027     {\bbl@exp{\adddialect\<l@#1>\bbl@tempa}}% found in opt list or ini
2028     \bbl@ifunset{\bbl@prehc\language}%
2029     {}}% TODO - XeTeX, based on \babelfont and HyphenChar?
2030     {\ifodd\bbl@engine\bbl@exp{%
2031         \bbl@ifblank{\@nameuse{\bbl@prehc@#1}}%
2032         {}}%
2033         {\AddBabelHook[\language]{babel-prehc-\language}{patterns}%
2034         {\prehyphenchar=\@nameuse{\bbl@prehc\language}\relax}}}%
2035     \fi}}

```

The reader of ini files. There are 3 possible cases: a section name (in the form [ . . . ]), a comment (starting with ;) and a key/value pair. *TODO - Work in progress.*

```

2036 \def\bbl@read@ini#1#2{%
2037     \openin1=babel-#1.ini           % FIXME - number must not be hardcoded
2038     \ifeof1
2039         \bbl@error
2040         {There is no ini file for the requested language\%
2041         (#1). Perhaps you misspelled it or your installation\%
2042         is not complete.}%
2043         {Fix the name or reinstall babel.}%
2044     \else
2045         \let\bbl@section\@empty
2046         \let\bbl@savestrings\@empty
2047         \let\bbl@savetoday\@empty
2048         \let\bbl@savestate\@empty
2049         \def\bbl@inipreread##1=##2\@{%
2050             \bbl@trim\def\bbl@tempa{##1}% Redundant below !!
2051             % Move trims here ??
2052             \bbl@ifunset{\bbl@KVP@\bbl@section..\bbl@tempa}%
2053             {\expandafter\bbl@inireader\bbl@tempa=##2\@}%
2054             {}}%
2055         \let\bbl@inireader\bbl@iniskip
2056         \bbl@info{Importing #2 for \language\%
2057             from babel-#1.ini. Reported}%
2058         \loop
2059         \if T\ifeof1F\fi T\relax % Trick, because inside \loop
2060             \endlinechar\m@ne
2061             \read1 to \bbl@line
2062             \endlinechar\^^M
2063             \ifx\bbl@line\@empty\else
2064                 \expandafter\bbl@iniline\bbl@line\bbl@iniline
2065             \fi
2066         \repeat
2067         \bbl@foreach\bbl@renewlist{%
2068             \bbl@ifunset{\bbl@renew@##1}{\bbl@inisec[##1]\@}%
2069         \global\let\bbl@renewlist\@empty
2070         % Ends last section. See \bbl@inisec
2071         \def\bbl@elt##1##2{\bbl@inireader##1=##2\@}%
2072         \@nameuse{\bbl@renew@\bbl@section}%
2073         \global\bbl@csarg\let{\renew@\bbl@section}\relax
2074         \@nameuse{\bbl@secpost@\bbl@section}%
2075     \fi}
2076 \def\bbl@iniline#1\bbl@iniline{%
2077     \@ifnextchar[\bbl@inisec{\@ifnextchar;\bbl@iniskip\bbl@inipreread}#1\@}% ]

```

The special cases for comment lines and sections are handled by the two following commands. In sections, we provide the possibility to take extra actions at the end or at the

start (TODO - but note the last section is not ended). By default, key=val pairs are ignored. The secpost “hook” is used only by ‘identification’, while secpre only by date.gregorian.licr.

```

2078 \def\bbl@iniskip#1\@{%      if starts with ;
2079 \def\bbl@inisec[#1]#2\@{%    if starts with opening bracket
2080 \def\bbl@elt##1##2{\bbl@inireader##1=##2\@}%
2081 \@nameuse{\bbl@renew@\bbl@section}%
2082 \global\bbl@csarg\let{\renew@\bbl@section}\relax
2083 \@nameuse{\bbl@secpost@\bbl@section}% ends previous section
2084 \def\bbl@section{#1}%        starts current section
2085 \def\bbl@elt##1##2{%
2086   \@namedef{\bbl@KVP@#1..##1}{}}%
2087 \@nameuse{\bbl@renew@#1}%
2088 \@nameuse{\bbl@secpre@#1}% pre-section ‘hook’
2089 \bbl@ifunset{\bbl@inikv@#1}%
2090   {\let\bbl@inireader\bbl@iniskip}%
2091   {\bbl@exp{\let\\bbl@inireader<\bbl@inikv@#1>}}
2092 \let\bbl@renewlist\@empty
2093 \def\bbl@renewinkey#1..#2\@#3{%
2094   \bbl@ifunset{\bbl@renew@#1}%
2095   {\bbl@add@list\bbl@renewlist{#1}}%
2096   {}}%
2097   \bbl@csarg\bbl@add{\renew@#1}{\bbl@elt{#2}{#3}}

```

Reads a key=val line and stores the trimmed val in \bbl@kv@<section>.<key>.

```

2098 \def\bbl@inikv#1=#2\@{%      key=value
2099   \bbl@trim\def\bbl@tempa{#1}%
2100   \bbl@trim\toks@{#2}%
2101   \bbl@csarg\edef{\kv@\bbl@section.\bbl@tempa}{\the\toks@}}

```

The previous assignments are local, so we need to export them. If the value is empty, we can provide a default value.

```

2102 \def\bbl@exportkey#1#2#3{%
2103   \bbl@ifunset{\bbl@kv@#2}%
2104   {\bbl@csarg\gdef{#1@\language}\{#3}}%
2105   {\expandafter\ifx\csname\bbl@kv@#2\endcsname\@empty
2106     \bbl@csarg\gdef{#1@\language}\{#3}}%
2107   \else
2108     \bbl@exp{\global\let<\bbl@#1@\language>\<\bbl@kv@#2>}}%
2109   \fi}}

```

Key-value pairs are treated differently depending on the section in the ini file. The following macros are the readers for identification and typography. Note \bbl@secpost@identification is called always (via \bbl@inisec), while \bbl@after@ini must be called explicitly after \bbl@read@ini if necessary.

```

2110 \let\bbl@inikv@identification\bbl@inikv
2111 \def\bbl@secpost@identification{%
2112   \bbl@ifunset{\bbl@kv@identification.name.opentype}%
2113   {\bbl@exportkey{lname}{identification.name.english}{}}%
2114   {\bbl@exportkey{lname}{identification.name.opentype}{}}%
2115   \bbl@exportkey{lbcpr}{identification.tag.bcp47}{}}%
2116   \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
2117   \bbl@ifunset{\bbl@kv@identification.script.name.opentype}%
2118   {\bbl@exportkey{sname}{identification.script.name}{}}%
2119   {\bbl@exportkey{sname}{identification.script.name.opentype}{}}%
2120   \bbl@exportkey{sbcpr}{identification.script.tag.bcp47}{}}%
2121   \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}}
2122 \let\bbl@inikv@typography\bbl@inikv

```

```

2123 \let\bbl@inikv@characters\bbl@inikv
2124 \let\bbl@inikv@numbers\bbl@inikv
2125 \def\bbl@after@ini{%
2126   \bbl@exportkey{lfthm}{typography.lefthyphenmin}{2}%
2127   \bbl@exportkey{rgthm}{typography.righthyphenmin}{3}%
2128   \bbl@exportkey{prehc}{typography.prehyphenchar}{}%
2129   \bbl@exportkey{hyphr}{typography.hyphenrules}{}%
2130   \bbl@exportkey{intsp}{typography.intraspace}{}%
2131   \bbl@exportkey{jstfy}{typography.justify}{w}%
2132   \bbl@exportkey{chrng}{characters.ranges}{}%
2133   \bbl@exportkey{dgnat}{numbers.digits.native}{}%
2134   \bbl@exportkey{rqtex}{identification.require.babel}{}%
2135   \bbl@xin@{0.5}{\@nameuse{bbl@kv@identification.version}}%
2136   \ifin@
2137     \bbl@warning{%
2138       There are neither captions nor date in '\language' name.\%
2139       It may not be suitable for proper typesetting, and it\%
2140       could change. Reported}%
2141   \fi
2142   \bbl@xin@{0.9}{\@nameuse{bbl@kv@identification.version}}%
2143   \ifin@
2144     \bbl@warning{%
2145       The '\language' date format may not be suitable\%
2146       for proper typesetting, and therefore it very likely will\%
2147       change in a future release. Reported}%
2148   \fi
2149   \bbl@toggle\bbl@savetoday
2150   \bbl@toggle\bbl@savestate

```

Now captions and captions.licr, depending on the engine. And below also for dates. They rely on a few auxiliary macros. It is expected the ini file provides the complete set in Unicode and LICR, in that order.

```

2151 \ifcase\bbl@engine
2152   \bbl@csarg\def{inikv@captions.licr}#1=#2\@@{%
2153     \bbl@ini@captions@aux{#1}{#2}}
2154 \else
2155   \def\bbl@inikv@captions#1=#2\@@{%
2156     \bbl@ini@captions@aux{#1}{#2}}
2157 \fi

```

The auxiliary macro for captions define \<caption>name.

```

2158 \def\bbl@ini@captions@aux#1#2{%
2159   \bbl@trim\def\bbl@tempa{#1}%
2160   \bbl@ifblank{#2}%
2161   {\bbl@exp{%
2162     \toks@{\bbl@nocaption{\bbl@tempa}{\language\bbl@tempa name}}}%
2163   {\bbl@trim\toks@{#2}}%
2164   \bbl@exp{%
2165     \bbl@add\bbl@savestrings{%
2166       \SetString\<\bbl@tempa name>{\the\toks@}}%

```

But dates are more complex. The full date format is stores in date.gregorian, so we must read it in non-Unicode engines, too (saved months are just discarded when the LICR section is reached).

TODO. Remove copypaste pattern.

```

2167 \bbl@csarg\def{inikv@date.gregorian}#1=#2\@@%           for defaults
2168   \bbl@inidate#1...\relax{#2}{}}
2169 \bbl@csarg\def{inikv@date.islamic}#1=#2\@@{%
2170   \bbl@inidate#1...\relax{#2}{islamic}}

```

```

2171 \bbl@csarg\def{inikv@date.hebrew}#1=#2\@@{%
2172 \bbl@inidate#1...\relax{#2}{hebrew}}
2173 \bbl@csarg\def{inikv@date.persian}#1=#2\@@{%
2174 \bbl@inidate#1...\relax{#2}{persian}}
2175 \bbl@csarg\def{inikv@date.indian}#1=#2\@@{%
2176 \bbl@inidate#1...\relax{#2}{indian}}
2177 \ifcase\bbl@engine
2178 \bbl@csarg\def{inikv@date.gregorian.licr}#1=#2\@@{% override
2179 \bbl@inidate#1...\relax{#2}{}}
2180 \bbl@csarg\def{secpre@date.gregorian.licr}{% discard uni
2181 \ifcase\bbl@engine\let\bbl@savestate\empty\fi}
2182 \fi
2183 % eg: 1=months, 2=wide, 3=1, 4=dummy
2184 \def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
2185 \bbl@trim@def\bbl@tempa{#1.#2}%
2186 \bbl@ifsamestring{\bbl@tempa}{months.wide}% to savedate
2187 {\bbl@trim@def\bbl@tempa{#3}%
2188 \bbl@trim\toks@{#5}%
2189 \bbl@exp{%
2190 \\\bbl@add\\bbl@savestate{%
2191 \\\SetString<\month\romannumeral\bbl@tempa#6name>{\the\toks@}}}%
2192 {\bbl@ifsamestring{\bbl@tempa}{date.long}% defined now
2193 {\bbl@trim@def\bbl@toreplace{#5}%
2194 \bbl@TG@date
2195 \global\bbl@csarg\let{date@\language name}\bbl@toreplace
2196 \bbl@exp{%
2197 \gdef<\language name date>{\\\protect<\language name date >}%
2198 \gdef<\language name date >####1####2####3{%
2199 \\\bbl@usedategroupttrue
2200 \<bbl@ensure@\language name>{%
2201 \<bbl@date@\language name>{####1}{####2}{####3}}}%
2202 \\\bbl@add\\bbl@savetoday{%
2203 \\\SetString\\today{%
2204 \<\language name date>{\\\the\year}{\\the\month}{\\the\day}}}}}%
2205 {}

```

Dates will require some macros for the basic formatting. They may be redefined by language, so “semi-public” names (camel case) are used. Oddly enough, the CLDR places particles like “de” inconsistently in either in the date or in the month name.

```

2206 \let\bbl@calendar\empty
2207 \newcommand\BabelDateSpace{\nobreakspace}
2208 \newcommand\BabelDateDot{.\@}
2209 \newcommand\BabelDated[1]{\number#1}
2210 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
2211 \newcommand\BabelDateM[1]{\number#1}
2212 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}
2213 \newcommand\BabelDateMMM[1]{\number#1}
2214 \csname month\romannumeral#1\bbl@calendar name\endcsname}%
2215 \newcommand\BabelDatey[1]{\number#1}%
2216 \newcommand\BabelDateyy[1]{\number#1}%
2217 \ifnum#1<10 0\number#1 %
2218 \else\ifnum#1<100 \number#1 %
2219 \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
2220 \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
2221 \else
2222 \bbl@error
2223 {Currently two-digit years are restricted to the\
2224 range 0-9999.}%
2225 {There is little you can do. Sorry.}%

```

```

2226 \fi\fi\fi\fi}}
2227 \newcommand\BabelDateyyy[1]{\number#1} % FIXME - add leading 0
2228 \def\bbl@replace@finish@iii#1{%
2229 \bbl@exp{\def\#1###1###2###3{\the\toks@}}
2230 \def\bbl@TG@date{%
2231 \bbl@replace\bbl@toreplace{[ ]}{\BabelDateSpace{}}%
2232 \bbl@replace\bbl@toreplace{[.]}{\BabelDateDot{}}%
2233 \bbl@replace\bbl@toreplace{[d]}{\BabelDated{###3}}%
2234 \bbl@replace\bbl@toreplace{[dd]}{\BabelDatedd{###3}}%
2235 \bbl@replace\bbl@toreplace{[M]}{\BabelDateM{###2}}%
2236 \bbl@replace\bbl@toreplace{[MM]}{\BabelDateMM{###2}}%
2237 \bbl@replace\bbl@toreplace{[MMM]}{\BabelDateMMM{###2}}%
2238 \bbl@replace\bbl@toreplace{[y]}{\BabelDatey{###1}}%
2239 \bbl@replace\bbl@toreplace{[yy]}{\BabelDateyy{###1}}%
2240 \bbl@replace\bbl@toreplace{[yyyy]}{\BabelDateyyyy{###1}}%
2241 % Note after \bbl@replace \toks@ contains the resulting string.
2242 % TODO - Using this implicit behavior doesn't seem a good idea.
2243 \bbl@replace@finish@iii\bbl@toreplace}

```

Language and Script values to be used when defining a font or setting the direction are set with the following macros.

```

2244 \def\bbl@provide@lsys#1{%
2245 \bbl@ifunset{bbl@lname@#1}%
2246 {\bbl@ini@basic{#1}}%
2247 {}%
2248 \bbl@csarg\let{lsys@#1}\@empty
2249 \bbl@ifunset{bbl@sname@#1}{\bbl@csarg\gdef{sname@#1}{Default}}{}%
2250 \bbl@ifunset{bbl@sotf@#1}{\bbl@csarg\gdef{sotf@#1}{DFLT}}{}%
2251 \bbl@csarg\bbl@add@list{lsys@#1}{Script=\bbl@cs{sname@#1}}%
2252 \bbl@ifunset{bbl@lname@#1}{%
2253 {\bbl@csarg\bbl@add@list{lsys@#1}{Language=\bbl@cs{lname@#1}}}%
2254 \bbl@csarg\bbl@to@global{lsys@#1}}

```

The following ini reader ignores everything but the identification section. It is called when a font is defined (ie, when the language is first selected) to know which script/language must be enabled. This means we must make sure a few characters are not active. The ini is not read directly, but with a proxy tex file named as the language (which means any code in it must be skipped, too).

```

2255 \def\bbl@ini@basic#1{%
2256 \def\BabelBeforeIni##1##2{%
2257 \begingroup
2258 \bbl@add\bbl@secpost@identification{\closein1 }%
2259 \catcode`\[=12 \catcode`\]=12 \catcode`\==12 %
2260 \bbl@read@ini{##1}{font and identification data}%
2261 \endinput % babel- .tex may contain onlypreamble's
2262 \endgroup}% boxed, to avoid extra spaces:
2263 {\setbox\z@\hbox{\InputIfFileExists{babel-#1.tex}{}}}%
2264 % \section{Adjusting the Babel behavior}
2265 %
2266 % \changes{babel~3.36}{2019/10/30}{New macro \cs{babeladjust}}
2267 %
2268 % A generic high level inteface is provided to adjust some global
2269 % and general settings.
2270 %
2271 % \begin{macrocode}
2272 \newcommand\babeladjust[1]{% TODO. Error handling.
2273 \bbl@forkv{#1}{\@nameuse{bbl@ADJ@##1@##2}}
2274 %
2275 \def\bbl@adjust@lua#1#2{%

```

```

2276 \ifvmode
2277   \ifnum\currentgrouplevel=\z@
2278     \directlua{ Babel.#2 }%
2279     \expandafter\expandafter\expandafter\@gobble
2280   \fi
2281 \fi
2282 {\bbl@error % The error is gobbled if everything went ok.
2283   {Currently, #1 related features can be adjusted only\\%
2284     in the main vertical list.}%
2285   {Maybe things change in the future, but this is what it is.}}}
2286 \@namedef{bbl@ADJ@bidi.mirroring@on}{%
2287   \bbl@adjust@lua{bidi}{mirroring_enabled=true}}
2288 \@namedef{bbl@ADJ@bidi.mirroring@off}{%
2289   \bbl@adjust@lua{bidi}{mirroring_enabled=false}}
2290 \@namedef{bbl@ADJ@bidi.text@on}{%
2291   \bbl@adjust@lua{bidi}{bidi_enabled=true}}
2292 \@namedef{bbl@ADJ@bidi.text@off}{%
2293   \bbl@adjust@lua{bidi}{bidi_enabled=false}}
2294 \@namedef{bbl@ADJ@bidi.mapdigits@on}{%
2295   \bbl@adjust@lua{bidi}{digits_mapped=true}}
2296 \@namedef{bbl@ADJ@bidi.mapdigits@off}{%
2297   \bbl@adjust@lua{bidi}{digits_mapped=false}}
2298 %
2299 \@namedef{bbl@ADJ@linebreak.sea@on}{%
2300   \bbl@adjust@lua{linebreak}{sea_enabled=true}}
2301 \@namedef{bbl@ADJ@linebreak.sea@off}{%
2302   \bbl@adjust@lua{linebreak}{sea_enabled=false}}
2303 \@namedef{bbl@ADJ@linebreak.cjk@on}{%
2304   \bbl@adjust@lua{linebreak}{cjk_enabled=true}}
2305 \@namedef{bbl@ADJ@linebreak.cjk@off}{%
2306   \bbl@adjust@lua{linebreak}{cjk_enabled=false}}
2307 %
2308 \def\bbl@adjust@layout#1{%
2309   \ifvmode
2310     #1%
2311     \expandafter\@gobble
2312   \fi
2313   {\bbl@error % The error is gobbled if everything went ok.
2314     {Currently, layout related features can be adjusted only\\%
2315       in vertical mode.}%
2316     {Maybe things change in the future, but this is what it is.}}}
2317 \@namedef{bbl@ADJ@layout.tabular@on}{%
2318   \bbl@adjust@layout{\let\@tabular\bbl@NL@@tabular}}
2319 \@namedef{bbl@ADJ@layout.tabular@off}{%
2320   \bbl@adjust@layout{\let\@tabular\bbl@OL@@tabular}}
2321 \@namedef{bbl@ADJ@layout.lists@on}{%
2322   \bbl@adjust@layout{\let\list\bbl@NL@list}}
2323 \@namedef{bbl@ADJ@layout.lists@off}{%
2324   \bbl@adjust@layout{\let\list\bbl@OL@list}}

```

## 10 The kernel of Babel (babel.def for L<sup>A</sup>T<sub>E</sub>Xonly)

### 10.1 The redefinition of the style commands

The rest of the code in this file can only be processed by L<sup>A</sup>T<sub>E</sub>X, so we check the current format. If it is plain T<sub>E</sub>X, processing should stop here. But, because of the need to limit the scope of the definition of `\format`, a macro that is used locally in the following

`\if` statement, this comparison is done inside a group. To prevent  $\TeX$  from complaining about an unclosed group, the processing of the command `\endinput` is deferred until after the group is closed. This is accomplished by the command `\aftergroup`.

```
2325 {\def\format{plain}
2326 \ifx\fmtname\format
2327 \else
2328   \def\format{LaTeX2e}
2329   \ifx\fmtname\format
2330   \else
2331     \aftergroup\endinput
2332   \fi
2333 \fi}
```

## 10.2 Cross referencing macros

The  $\LaTeX$  book states:

The *key* argument is any sequence of letters, digits, and punctuation symbols; upper- and lowercase letters are regarded as different.

When the above quote should still be true when a document is typeset in a language that has active characters, special care has to be taken of the category codes of these characters when they appear in an argument of the cross referencing macros.

When a cross referencing command processes its argument, all tokens in this argument should be character tokens with category ‘letter’ or ‘other’.

The only way to accomplish this in most cases is to use the trick described in the  $\TeX$ book [2] (Appendix D, page 382). The primitive `\meaning` applied to a token expands to the current meaning of this token. For example, ‘`\meaning\A`’ with `\A` defined as ‘`\def\A#1{\B}`’ expands to the characters ‘`macro:#1->\B`’ with all category codes set to ‘other’ or ‘space’.

`\newlabel` The macro `\label` writes a line with a `\newlabel` command into the `.aux` file to define labels.

```
2334 %\bbl@redefine\newlabel#1#2{%
2335 % \@safe@activestrue\org@newlabel{#1}{#2}\@safe@activesfalse}
```

`\@newl@bel` We need to change the definition of the  $\LaTeX$ -internal macro `\@newl@bel`. This is needed because we need to make sure that shorthand characters expand to their non-active version.

The following package options control which macros are to be redefined.

```
2336 <<{*More package options}>> ≡
2337 \DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
2338 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
2339 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
2340 <</More package options>>
```

First we open a new group to keep the changed setting of `\protect` local and then we set the `@safe@actives` switch to true to make sure that any shorthand that appears in any of the arguments immediately expands to its non-active self.

```
2341 \bbl@trace{Cross referencing macros}
2342 \ifx\bbl@opt@safe\@empty\else
2343   \def\@newl@bel#1#2#3{%
2344     \@safe@activestrue
2345     \bbl@ifunset{#1@#2}%
2346     \relax
2347     {\gdef\@multiplelabels{%
```



```

2348      \@latex@warning@no@line{There were multiply-defined labels}}%
2349      \@latex@warning@no@line{Label `#2' multiply defined}}%
2350      \global\@namedef{#1@#2}{#3}}

```

`\@testdef` An internal  $\LaTeX$  macro used to test if the labels that have been written on the .aux file have changed. It is called by the `\enddocument` macro. This macro needs to be completely rewritten, using `\meaning`. The reason for this is that in some cases the expansion of `\#1@#2` contains the same characters as the `#3`; but the character codes differ. Therefore  $\LaTeX$  keeps reporting that the labels may have changed.

```

2351  \CheckCommand*\@testdef[3]{%
2352    \def\reserved@a{#3}%
2353    \expandafter\ifx\csname#1@#2\endcsname\reserved@a
2354    \else
2355      \@tempswatrue
2356    \fi}

```

Now that we made sure that `\@testdef` still has the same definition we can rewrite it. First we make the shorthands ‘safe’.

```

2357  \def\@testdef#1#2#3{%
2358    \@safe@activetrue

```

Then we use `\bbl@tempa` as an ‘alias’ for the macro that contains the label which is being checked.

```

2359    \expandafter\let\expandafter\bbl@tempa\csname #1@#2\endcsname

```

Then we define `\bbl@tempb` just as `\@newl@bel` does it.

```

2360    \def\bbl@tempb{#3}%
2361    \@safe@activetrue

```

When the label is defined we replace the definition of `\bbl@tempa` by its meaning.

```

2362    \ifx\bbl@tempa\relax
2363    \else
2364      \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
2365    \fi

```

We do the same for `\bbl@tempb`.

```

2366    \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%

```

If the label didn’t change, `\bbl@tempa` and `\bbl@tempb` should be identical macros.

```

2367    \ifx\bbl@tempa\bbl@tempb
2368    \else
2369      \@tempswatrue
2370    \fi}
2371 \fi

```

`\ref` `\pageref` The same holds for the macro `\ref` that references a label and `\pageref` to reference a page. So we redefine `\ref` and `\pageref`. While we change these macros, we make them robust as well (if they weren’t already) to prevent problems if they should become expanded at the wrong moment.

```

2372 \bbl@xin@{R}\bbl@opt@safe
2373 \ifin@
2374 \bbl@redefineroobust\ref#1{%
2375   \@safe@activetrue\org@ref{#1}\@safe@activetrue}
2376 \bbl@redefineroobust\pageref#1{%
2377   \@safe@activetrue\org@pageref{#1}\@safe@activetrue}
2378 \else
2379   \let\org@ref\ref
2380   \let\org@pageref\pageref
2381 \fi

```

`\@citex` The macro used to cite from a bibliography, `\cite`, uses an internal macro, `\@citex`. It is this internal macro that picks up the argument(s), so we redefine this internal macro and leave `\cite` alone. The first argument is used for typesetting, so the shorthands need only be deactivated in the second argument.

```
2382 \bbl@xin@{B}\bbl@opt@safe
2383 \ifin@
2384 \bbl@redefine\@citex[#1]#2{%
2385   \@safe@activetrue\edef\@tempa{#2}\@safe@activesfalse
2386   \org@@citex[#1]{\@tempa}}
```

Unfortunately, the packages `natbib` and `cite` need a different definition of `\@citex`... To begin with, `natbib` has a definition for `\@citex` with *three* arguments... We only know that a package is loaded when `\begin{document}` is executed, so we need to postpone the different redefinition.

```
2387 \AtBeginDocument{%
2388   \@ifpackageloaded{natbib}{%
```

Notice that we use `\def` here instead of `\bbl@redefine` because `\org@@citex` is already defined and we don't want to overwrite that definition (it would result in parameter stack overflow because of a circular definition).

(Recent versions of `natbib` change dynamically `\@citex`, so PR4087 doesn't seem fixable in a simple way. Just load `natbib` before.)

```
2389   \def\@citex[#1][#2]#3{%
2390     \@safe@activetrue\edef\@tempa{#3}\@safe@activesfalse
2391     \org@@citex[#1][#2]{\@tempa}}%
2392   }{}}
```

The package `cite` has a definition of `\@citex` where the shorthands need to be turned off in both arguments.

```
2393 \AtBeginDocument{%
2394   \@ifpackageloaded{cite}{%
2395     \def\@citex[#1]#2{%
2396       \@safe@activetrue\org@@citex[#1]{#2}\@safe@activesfalse}%
2397     }{}}
```

`\nocite` The macro `\nocite` which is used to instruct `BiBTeX` to extract uncited references from the database.

```
2398 \bbl@redefine\nocite#1{%
2399   \@safe@activetrue\org@nocite{#1}\@safe@activesfalse}
```

`\bibcite` The macro that is used in the `.aux` file to define citation labels. When packages such as `natbib` or `cite` are not loaded its second argument is used to typeset the citation label. In that case, this second argument can contain active characters but is used in an environment where `\@safe@activetrue` is in effect. This switch needs to be reset inside the `\hbox` which contains the citation label. In order to determine during `.aux` file processing which definition of `\bibcite` is needed we define `\bibcite` in such a way that it redefines itself with the proper definition. We call `\bbl@cite@choice` to select the proper definition for `\bibcite`. This new definition is then activated.

```
2400 \bbl@redefine\bibcite{%
2401   \bbl@cite@choice
2402   \bibcite}
```

`\bbl@bibcite` The macro `\bbl@bibcite` holds the definition of `\bibcite` needed when neither `natbib` nor `cite` is loaded.

```
2403 \def\bbl@bibcite#1#2{%
2404   \org@bibcite{#1}{\@safe@activesfalse#2}}
```

`\bbl@cite@choice` The macro `\bbl@cite@choice` determines which definition of `\bibcite` is needed. First we give `\bibcite` its default definition.

```
2405 \def\bbl@cite@choice{%
2406   \global\let\bibcite\bbl@bibcite
```

Then, when `natbib` is loaded we restore the original definition of `\bibcite`. For `cite` we do the same.

```
2407   \ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{}%
2408   \ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{}%
```

Make sure this only happens once.

```
2409   \global\let\bbl@cite@choice\relax}
```

When a document is run for the first time, no `.aux` file is available, and `\bibcite` will not yet be properly defined. In this case, this has to happen before the document starts.

```
2410 \AtBeginDocument{\bbl@cite@choice}
```

`\@bibitem` One of the two internal  $\TeX$  macros called by `\bibitem` that write the citation label on the `.aux` file.

```
2411 \bbl@redefine\@bibitem#1{%
2412   \@safe@activetrue\org@@bibitem{#1}\@safe@activetruefalse}
2413 \else
2414   \let\org@nocite\nocite
2415   \let\org@@citex\@citex
2416   \let\org@bibcite\bibcite
2417   \let\org@@bibitem\@bibitem
2418 \fi
```

### 10.3 Marks

`\markright` Because the output routine is asynchronous, we must pass the current language attribute to the head lines, together with the text that is put into them. To achieve this we need to adapt the definition of `\markright` and `\markboth` somewhat.

We check whether the argument is empty; if it is, we just make sure the scratch token register is empty. Next, we store the argument to `\markright` in the scratch token register. This way these commands will not be expanded later, and we make sure that the text is typeset using the correct language settings. While doing so, we make sure that active characters that may end up in the mark are not disabled by the output routine kicking in while `\@safe@activetrue` is in effect.

```
2419 \bbl@trace{Marks}
2420 \IfBabelLayout{sectioning}
2421   {\ifx\bbl@opt@headfoot\@nnil
2422     \g@addto@macro\@resetactivechars{%
2423       \set@typeset@protect
2424       \expandafter\select@language@x\expandafter{\bbl@main@language}%
2425       \let\protect\noexpand
2426       \edef\thepage{%
2427         \noexpand\babelsublr{\unexpanded\expandafter{\thepage}}}%
2428     \fi}
2429   {\ifbbl@single\else
2430     \bbl@ifunset{markright} \bbl@redefine\bbl@redefineroobust
2431     \markright#1{%
2432       \bbl@ifblank{#1}%
2433       {\org@markright}}}%
2434     {\toks@{#1}%
2435     \bbl@exp{%
```

```

2436      \org@markright{\protect\foreignlanguage{\language}%
2437      {\protect\bbl@restore@actives\the\toks@}}}%

```

`\markboth` The definition of `\markboth` is equivalent to that of `\markright`, except that we need two token registers. The documentclasses `report` and `book` define and set the headings for the page. While doing so they also store a copy of `\markboth` in `\@mkboth`. Therefore we need to check whether `\@mkboth` has already been set. If so we need to do that again with the new definition of `\markboth`. (As of Oct 2019,  $\LaTeX$  stores the definition in an intermediate macros, so it's not necessary anymore, but it's preserved for older versions.)

```

2438      \ifx\@mkboth\markboth
2439      \def\bbl@tempc{\let\@mkboth\markboth}
2440      \else
2441      \def\bbl@tempc{}
2442      \fi
2443      \bbl@ifunset{markboth }{\bbl@redefine\bbl@redefineroobust
2444      \markboth#1#2{%
2445      \protected@edef\bbl@tempb##1{%
2446      \protect\foreignlanguage
2447      {\language}{\protect\bbl@restore@actives##1}}%
2448      \bbl@ifblank{#1}%
2449      {\toks@{}}%
2450      {\toks@\expandafter{\bbl@tempb{#1}}}%
2451      \bbl@ifblank{#2}%
2452      {\@temptokena{}}%
2453      {\@temptokena\expandafter{\bbl@tempb{#2}}}%
2454      \bbl@exp{\org@markboth{\the\toks@}{\the\@temptokena}}}
2455      \bbl@tempc
2456      \fi} % end ifbbl@single, end \IfBabelLayout

```

## 10.4 Preventing clashes with other packages

### 10.4.1 `ifthen`

`\ifthenelse` Sometimes a document writer wants to create a special effect depending on the page a certain fragment of text appears on. This can be achieved by the following piece of code:

```

\ifthenelse{\isodd{\pageref{some:label}}}{
\code_for_odd_pages}
\code_for_even_pages}

```

In order for this to work the argument of `\isodd` needs to be fully expandable. With the above redefinition of `\pageref` it is not in the case of this example. To overcome that, we add some code to the definition of `\ifthenelse` to make things work.

The first thing we need to do is check if the package `ifthen` is loaded. This should be done at `\begin{document}` time.

```

2457 \bbl@trace{Preventing clashes with other packages}
2458 \bbl@xin@R\bbl@opt@safe
2459 \ifin@
2460 \AtBeginDocument{%
2461 \ifpackageloaded{ifthen}{%

```

Then we can redefine `\ifthenelse`:

```

2462 \bbl@redefine@long\ifthenelse#1#2#3{%

```

We want to revert the definition of `\pageref` and `\ref` to their original definition for the first argument of `\ifthenelse`, so we first need to store their current meanings.

```

2463      \let\bbl@temp@pref\pageref
2464      \let\pageref\org@pageref
2465      \let\bbl@temp@ref\ref
2466      \let\ref\org@ref

```

Then we can set the `\@safe@actives` switch and call the original `\ifthenelse`. In order to be able to use shorthands in the second and third arguments of `\ifthenelse` the resetting of the switch *and* the definition of `\pageref` happens inside those arguments. When the package wasn't loaded we do nothing.

```

2467      \@safe@activetrue
2468      \org@ifthenelse{#1}%
2469      {\let\pageref\bbl@temp@pref
2470       \let\ref\bbl@temp@ref
2471       \@safe@activesfalse
2472       #2}%
2473      {\let\pageref\bbl@temp@pref
2474       \let\ref\bbl@temp@ref
2475       \@safe@activesfalse
2476       #3}%
2477      }%
2478      }{}%
2479      }

```

#### 10.4.2 varioref

`\@@vpageref` When the package `varioref` is in use we need to modify its internal command `\@@vpageref`  
`\vrefpagenum` in order to prevent problems when an active character ends up in the argument of `\vref`.  
`\Ref` The same needs to happen for `\vrefpagenum`.

```

2480 \AtBeginDocument{%
2481   \ifpackageloaded{varioref}{%
2482     \bbl@redefine\@@vpageref#1[#2]#3{%
2483       \@safe@activetrue
2484       \org@@@vpageref{#1}[#2]{#3}%
2485       \@safe@activesfalse}%
2486     \bbl@redefine\vrefpagenum#1#2{%
2487       \@safe@activetrue
2488       \org@vrefpagenum{#1}{#2}%
2489       \@safe@activesfalse}%

```

The package `varioref` defines `\Ref` to be a robust command which uppercases the first character of the reference text. In order to be able to do that it needs to access the expandable form of `\ref`. So we employ a little trick here. We redefine the (internal) command `\Ref` to call `\org@ref` instead of `\ref`. The disadvantage of this solution is that whenever the definition of `\Ref` changes, this definition needs to be updated as well.

```

2490   \expandafter\def\csname Ref\endcsname#1{%
2491     \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
2492   }{}%
2493   }
2494 \fi

```

#### 10.4.3 hhline

`\hhline` Delaying the activation of the shorthand characters has introduced a problem with the `hhline` package. The reason is that it uses the `‘` character which is made active by the french support in `babel`. Therefore we need to *reload* the package when the `‘` is an active character.

So at `\begin{document}` we check whether `hhline` is loaded.

```
2495 \AtEndOfPackage{%
2496   \AtBeginDocument{%
2497     \ifpackageloaded{hhline}%
```

Then we check whether the expansion of `\normal@char:` is not equal to `\relax`.

```
2498     {\expandafter\ifx\csname normal@char\string\endcsname\relax
2499       \else
```

In that case we simply reload the package. Note that this happens *after* the category code of the `@`-sign has been changed to other, so we need to temporarily change it to letter again.

```
2500         \makeatletter
2501         \def\@currname{hhline}\input{hhline.sty}\makeatother
2502         \fi}%
2503     {}}}
```

#### 10.4.4 hyperref

`\pdfstringdefDisableCommands` A number of interworking problems between `babel` and `hyperref` are tackled by `hyperref` itself. The following code was introduced to prevent some annoying warnings but it broke bookmarks. This was quickly fixed in `hyperref`, which essentially made it no-op. However, it will not be removed for the moment because `hyperref` is expecting it.

```
2504 \AtBeginDocument{%
2505   \ifx\pdfstringdefDisableCommands\undefined\else
2506     \pdfstringdefDisableCommands{\languageshortands{system}}%
2507   \fi}
```

#### 10.4.5 fancyhdr

`\FOREIGNLANGUAGE` The package `fancyhdr` treats the running head and foot lines somewhat differently as the standard classes. A symptom of this is that the command `\foreignlanguage` which `babel` adds to the marks can end up inside the argument of `\MakeUppercase`. To prevent unexpected results we need to define `\FOREIGNLANGUAGE` here.

```
2508 \DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
2509   \lowercase{\foreignlanguage{#1}}}
```

`\substitutefontfamily` The command `\substitutefontfamily` creates an `.fd` file on the fly. The first argument is an encoding mnemonic, the second and third arguments are font family names.

```
2510 \def\substitutefontfamily#1#2#3{%
2511   \lowercase{\immediate\openout15=#1#2.fd\relax}%
2512   \immediate\write15{%
2513     \string\ProvidesFile{#1#2.fd}%
2514     [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
2515     \space generated font description file]^^J
2516     \string\DeclareFontFamily{#1}{#2}{^^J
2517     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{^^J
2518     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{^^J
2519     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{^^J
2520     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{^^J
2521     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{^^J
2522     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{^^J
2523     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{^^J
2524     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{^^J
2525     }%
2526   \closeout15
2527 }
```

This command should only be used in the preamble of a document.

```
2528 \@onlypreamble\substitutefontfamily
```

## 10.5 Encoding and fonts

Because documents may use non-ASCII font encodings, we make sure that the logos of  $\TeX$  and  $\LaTeX$  always come out in the right encoding. There is a list of non-ASCII encodings. Unfortunately, fontenc deletes its package options, so we must guess which encodings has been loaded by traversing `\@filelist` to search for `\enc.def`. If a non-ASCII has been loaded, we define versions of  $\TeX$  and  $\LaTeX$  for them using `\ensureascii`. The default ASCII encoding is set, too (in reverse order): the “main” encoding (when the document begins), the last loaded, or OT1.

`\ensureascii`

```
2529 \bbl@trace{Encoding and fonts}
2530 \newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU,PU,PD1}
2531 \newcommand\BabelNonText{TS1,T3,TS3}
2532 \let\org@TeX\TeX
2533 \let\org@LaTeX\LaTeX
2534 \let\ensureascii\@firstofone
2535 \AtBeginDocument{%
2536   \in@false
2537   \bbl@foreach\BabelNonASCII{% is there a text non-ascii enc?
2538     \ifin@
2539       \lowercase{\bbl@xin@{,#1enc.def,}{,\@filelist,}}%
2540     \fi}%
2541   \ifin@ % if a text non-ascii has been loaded
2542     \def\ensureascii#1{{\fontencoding{OT1}\selectfont#1}}%
2543     \DeclareTextCommandDefault{\TeX}{\org@TeX}%
2544     \DeclareTextCommandDefault{\LaTeX}{\org@LaTeX}%
2545     \def\bbl@tempb#1\@@{\uppercase{\bbl@tempc#1}ENC.DEF\@empty\@@}%
2546     \def\bbl@tempc#1ENC.DEF#2\@@{%
2547       \ifx\@empty#2\else
2548         \bbl@ifunset{T#1}%
2549         {}%
2550         {\bbl@xin@{,#1,}{,\BabelNonASCII,\BabelNonText,}}%
2551       \ifin@
2552         \DeclareTextCommand{\TeX}{#1}{\ensureascii{\org@TeX}}%
2553         \DeclareTextCommand{\LaTeX}{#1}{\ensureascii{\org@LaTeX}}%
2554       \else
2555         \def\ensureascii##1{{\fontencoding{#1}\selectfont##1}}%
2556       \fi}%
2557   \fi}%
2558   \bbl@foreach\@filelist{\bbl@tempb#1\@@}% TODO - \@@ de mas??
2559   \bbl@xin@{\cf@encoding,}{,\BabelNonASCII,\BabelNonText,}%
2560   \ifin@
2561     \edef\ensureascii#1{%
2562       \noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}%
2563   \fi
2564 \fi}
```

Now comes the old deprecated stuff (with a little change in 3.9l, for fontspec). The first thing we need to do is to determine, at `\begin{document}`, which latin fontencoding to use.

`\latinencoding`

When text is being typeset in an encoding other than ‘latin’ (OT1 or T1), it would be nice to still have Roman numerals come out in the Latin encoding. So we first assume that the current encoding at the end of processing the package is the Latin encoding.

```
2565 \AtEndOfPackage{\edef\latinencoding{\cf@encoding}}
```

But this might be overruled with a later loading of the package fontenc. Therefore we check at the execution of `\begin{document}` whether it was loaded with the T1 option. The normal way to do this (using `\ifpackageloaded`) is disabled for this package. Now we have to revert to parsing the internal macro `\@filelist` which contains all the filenames loaded.

```

2566 \AtBeginDocument{%
2567   \ifpackageloaded{fontspec}%
2568     {\xdef\latinencoding{%
2569       \ifx\UTFencname\@undefined
2570         EU\ifcase\bbl@engine\or2\or1\fi
2571       \else
2572         \UTFencname
2573       \fi}}%
2574   {\gdef\latinencoding{OT1}%
2575     \ifx\cf@encoding\bbl@t@one
2576       \xdef\latinencoding{\bbl@t@one}%
2577     \else
2578       \@ifl@aded{def}{t1enc}{\xdef\latinencoding{\bbl@t@one}}}%
2579     \fi}}

```

`\latintext` Then we can define the command `\latintext` which is a declarative switch to a latin font-encoding. Usage of this macro is deprecated.

```

2580 \DeclareRobustCommand{\latintext}{%
2581   \fontencoding{\latinencoding}\selectfont
2582   \def\encodingdefault{\latinencoding}}

```

`\textlatin` This command takes an argument which is then typeset using the requested font encoding. In order to avoid many encoding switches it operates in a local scope.

```

2583 \ifx\@undefined\DeclareTextFontCommand
2584   \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
2585 \else
2586   \DeclareTextFontCommand{\textlatin}{\latintext}
2587 \fi

```

## 10.6 Basic bidi support

**Work in progress.** This code is currently placed here for practical reasons.

It is loosely based on `rlbabel.def`, but most of it has been developed from scratch. This babel module (by Johannes Braams and Boris Lavva) has served the purpose of typesetting R documents for two decades, and despite its flaws I think it is still a good starting point (some parts have been copied here almost verbatim), partly thanks to its simplicity. I’ve also looked at ARABI (by Youssef Jabri), which is compatible with babel.

There are two ways of modifying macros to make them “bidi”, namely, by patching the internal low-level macros (which is what I have done with lists, columns, counters, tocs, much like `rlbabel` did), and by introducing a “middle layer” just below the user interface (sectioning, footnotes).

- `pdftex` provides a minimal support for bidi text, and it must be done by hand. Vertical typesetting is not possible.
- `xetex` is somewhat better, thanks to its font engine (even if not always reliable) and a few additional tools. However, very little is done at the paragraph level. Another challenging problem is text direction does not honour  $\TeX$  grouping.
- `luatex` can provide the most complete solution, as we can manipulate almost freely the node list, the generated lines, and so on, but bidi text does not work out of the box and



some development is necessary. It also provides tools to properly set left-to-right and right-to-left page layouts. As Lua<sub>T</sub><sub>E</sub>X-j<sub>a</sub> shows, vertical typesetting is possible, too. Its main drawback is font handling is often considered to be less mature than xetex, mainly in Indic scripts (but there are steps to make HarfBuzz, the xetex font engine, available in luatex; see <<https://github.com/tatzetwerk/luatex-harfbuzz>>).

```

2588 \bbl@trace{Basic (internal) bidi support}
2589 \def\bbl@alscripts{,Arabic,Syriac,Thaana,}
2590 \def\bbl@rscripts{%
2591   ,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
2592   Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaeen,%
2593   Manichaeen,Meroitic Cursive,Meroitic,Old North Arabian,%
2594   Nabataean,N'Ko,Orkhon,Palmyrene,Inscriptional Pahlavi,%
2595   Psalter Pahlavi,Phoenician,Inscriptional Parthian,Samaritan,%
2596   Old South Arabian,}%
2597 \def\bbl@provide@dirs#1{%
2598   \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts\bbl@rscripts}%
2599   \ifin@
2600     \global\bbl@csarg\chardef{wdir@#1}\@ne
2601     \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts}%
2602     \ifin@
2603       \global\bbl@csarg\chardef{wdir@#1}\tw@ % useless in xetex
2604       \fi
2605     \else
2606       \global\bbl@csarg\chardef{wdir@#1}\z@
2607     \fi
2608     \ifodd\bbl@engine
2609       \bbl@csarg\ifcase{wdir@#1}%
2610         \directlua{ Babel.locale_props[\the\localeid].texmdir = 'l' }%
2611       \or
2612         \directlua{ Babel.locale_props[\the\localeid].texmdir = 'r' }%
2613       \or
2614         \directlua{ Babel.locale_props[\the\localeid].texmdir = 'al' }%
2615       \fi
2616     \fi}
2617 \def\bbl@switchdir{%
2618   \bbl@ifunset{bbl@lsys@\languagename}{\bbl@provide@lsys{\languagename}}}%
2619   \bbl@ifunset{bbl@wdir@\languagename}{\bbl@provide@dirs{\languagename}}}%
2620   \bbl@exp{\bbl@setdirs\bbl@cs{wdir@\languagename}}}%
2621 \def\bbl@setdirs#1{% TODO - math
2622   \ifcase\bbl@select@type % TODO - strictly, not the right test
2623     \bbl@bodydir{#1}%
2624     \bbl@pardir{#1}%
2625   \fi
2626   \bbl@texmdir{#1}}
2627 \ifodd\bbl@engine % luatex=1
2628   \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
2629   \DisableBabelHook{babel-bidi}
2630   \chardef\bbl@thetexmdir\z@
2631   \chardef\bbl@thepardir\z@
2632   \def\bbl@getluadir#1{%
2633     \directlua{
2634       if tex.#1dir == 'TLT' then
2635         tex.sprint('0')
2636       elseif tex.#1dir == 'TRT' then
2637         tex.sprint('1')
2638       end}}
2639   \def\bbl@setluadir#1#2#3{% 1=text/par.. 2=\texmdir.. 3=0 lr/1 rl
2640     \ifcase#3relax

```

```

2641 \ifcase\bb1@getluadir{#1}\relax\else
2642 #2 TLT\relax
2643 \fi
2644 \else
2645 \ifcase\bb1@getluadir{#1}\relax
2646 #2 TRT\relax
2647 \fi
2648 \fi}
2649 \def\bb1@textdir#1{%
2650 \bb1@setluadir{text}\textdir{#1}%
2651 \chardef\bb1@thetextdir#1\relax
2652 \setattribute\bb1@attr@dir{\numexpr\bb1@thepardir*3+#1}}
2653 \def\bb1@pardir#1{%
2654 \bb1@setluadir{par}\pardir{#1}%
2655 \chardef\bb1@thepardir#1\relax}
2656 \def\bb1@bodydir{\bb1@setluadir{body}\bodydir}
2657 \def\bb1@pagedir{\bb1@setluadir{page}\pagedir}
2658 \def\bb1@dirparastext{\pardir\the\textdir\relax}% %%%
2659 % Sadly, we have to deal with boxes in math with basic.
2660 % Activated every math with the package option bidi=:
2661 \def\bb1@mathboxdir{%
2662 \ifcase\bb1@thetextdir\relax
2663 \everyhbox{\textdir TLT\relax}%
2664 \else
2665 \everyhbox{\textdir TRT\relax}%
2666 \fi}
2667 \else % pdftex=0, xetex=2
2668 \AddBabelHook{babel-bidi}{afterextras}{\bb1@switchdir}
2669 \DisableBabelHook{babel-bidi}
2670 \newcount\bb1@dirlevel
2671 \chardef\bb1@thetextdir\z@
2672 \chardef\bb1@thepardir\z@
2673 \def\bb1@textdir#1{%
2674 \ifcase#1\relax
2675 \chardef\bb1@thetextdir\z@
2676 \bb1@textdir@i\beginL\endL
2677 \else
2678 \chardef\bb1@thetextdir@ne
2679 \bb1@textdir@i\beginR\endR
2680 \fi}
2681 \def\bb1@textdir@i#1#2{%
2682 \ifhmode
2683 \ifnum\currentgrouplevel>\z@
2684 \ifnum\currentgrouplevel=\bb1@dirlevel
2685 \bb1@error{Multiple bidi settings inside a group}%
2686 {I'll insert a new group, but expect wrong results.}%
2687 \bgroup\aftergroup#2\aftergroup\egroup
2688 \else
2689 \ifcase\currentgrouptype\or % 0 bottom
2690 \aftergroup#2% 1 simple {}
2691 \or
2692 \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
2693 \or
2694 \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
2695 \or\or\or % vbox vtop align
2696 \or
2697 \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
2698 \or\or\or\or\or\or % output math disc insert vcent mathchoice
2699 \or

```

```

2700         \aftergroup#2% 14 \begingroup
2701     \else
2702         \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
2703     \fi
2704 \fi
2705     \bbl@dirlevel\currentgrouplevel
2706 \fi
2707 #1%
2708 \fi}
2709 \def\bbl@pardir#1{\chardef\bbl@thepardir#1\relax}
2710 \let\bbl@bodydir\@gobble
2711 \let\bbl@pagedir\@gobble
2712 \def\bbl@dirparastext{\chardef\bbl@thepardir\bbl@thetextdir}

```

The following command is executed only if there is a right-to-left script (once). It activates the `\everypar` hack for xetex, to properly handle the par direction. Note text and par dirs are decoupled to some extent (although not completely).

```

2713 \def\bbl@xebidipar{%
2714     \let\bbl@xebidipar\relax
2715     \TeXeTstate\@ne
2716     \def\bbl@xeverypar{%
2717         \ifcase\bbl@thepardir
2718             \ifcase\bbl@thetextdir\else\beginR\fi
2719         \else
2720             {\setbox\z@\lastbox\beginR\box\z@}%
2721         \fi}%
2722     \let\bbl@severypar\everypar
2723     \newtoks\everypar
2724     \everypar=\bbl@severypar
2725     \bbl@severypar{\bbl@xeverypar\the\everypar}}
2726 \@ifpackagewith{babel}{bidi=bidi}%
2727 {\let\bbl@textdir@i\@gobbletwo
2728     \let\bbl@xebidipar\@empty
2729     \AddBabelHook{bidi}{foreign}{%
2730         \def\bbl@tempa{\def\BabelText####1}%
2731         \ifcase\bbl@thetextdir
2732             \expandafter\bbl@tempa\expandafter{\BabelText{\LR{##1}}}%
2733         \else
2734             \expandafter\bbl@tempa\expandafter{\BabelText{\RL{##1}}}%
2735         \fi}
2736     \def\bbl@pardir#1{\ifcase#1\relax\setLR\else\setRL\fi}}
2737 {}%
2738 \fi

```

A tool for weak L (mainly digits). We also disable warnings with hyperref.

```

2739 \DeclareRobustCommand\babelsublr[1]{\leavevmode{\bbl@textdir\z@#1}}
2740 \AtBeginDocument{%
2741     \ifx\pdfstringdefDisableCommands\@undefined\else
2742         \ifx\pdfstringdefDisableCommands\relax\else
2743             \pdfstringdefDisableCommands{\let\babelsublr\@firstofone}%
2744         \fi
2745     \fi}

```

## 10.7 Local Language Configuration

`\loadlocalcfg` At some sites it may be necessary to add site-specific actions to a language definition file. This can be done by creating a file with the same name as the language definition file, but

with the extension .cfg. For instance the file norsk.cfg will be loaded when the language definition file norsk.ldf is loaded.

For plain-based formats we don't want to override the definition of \loadlocalcfg from plain.def.

```

2746 \bbl@trace{Local Language Configuration}
2747 \ifx\loadlocalcfg\@undefined
2748   \@ifpackagewith{babel}{noconfigs}%
2749   {\let\loadlocalcfg@gobble}%
2750   {\def\loadlocalcfg#1{%
2751     \InputIfFileExists{#1.cfg}%
2752     {\typeout{*****^J%
2753               * Local config file #1.cfg used^^J%
2754               *}}%
2755     \@empty}}
2756 \fi

```

Just to be compatible with L<sup>A</sup>T<sub>E</sub>X 2.09 we add a few more lines of code:

```

2757 \ifx\@unexpandable@protect\@undefined
2758   \def\@unexpandable@protect{\noexpand\protect\noexpand}
2759   \long\def\protected@write#1#2#3{%
2760     \begingroup
2761       \let\thepage\relax
2762       #2%
2763       \let\protect\@unexpandable@protect
2764       \edef\reserved@a{\write#1{#3}}%
2765       \reserved@a
2766     \endgroup
2767     \if@nobreak\ifvmode\nobreak\fi\fi}
2768 \fi
2769 </core>
2770 <*kernel>

```

## 11 Multiple languages (switch.def)

Plain T<sub>E</sub>X version 3.0 provides the primitive \language that is used to store the current language. When used with a pre-3.0 version this function has to be implemented by allocating a counter.

```

2771 <<Make sure ProvidesFile is defined>>
2772 \ProvidesFile{switch.def}[\<date>] [\<version>] Babel switching mechanism]
2773 <<Load macros for plain if not LaTeX>>
2774 <<Define core switching macros>>

```

**\adddialect** The macro \adddialect can be used to add the name of a dialect or variant language, for which an already defined hyphenation table can be used.

```

2775 \def\bbl@version{\<version>}
2776 \def\bbl@date{\<date>}
2777 \def\adddialect#1#2{%
2778   \global\chardef#1#2\relax
2779   \bbl@usehooks{adddialect}{\#1}{\#2}}%
2780 \begingroup
2781   \count@#1\relax
2782   \def\bbl@elt##1##2##3##4{%
2783     \ifnum\count@=##2\relax
2784       \bbl@info{\string#1 = using hyphenrules for ##1\%
2785                 (\string\language\the\count@)}%
2786       \def\bbl@elt####1####2####3####4}%

```

```

2787     \fi}%
2788     \bbl@languages
2789 \endgroup}

```

`\bbl@iflanguage` executes code only if the language `l@` exists. Otherwise raises an error. The argument of `\bbl@fixname` has to be a macro name, as it may get “fixed” if casing (lc/uc) is wrong. It’s intended to fix a long-standing bug when `\foreignlanguage` and the like appear in a `\MakeXXXcase`. However, a lowercase form is not imposed to improve backward compatibility (perhaps you defined a language named MYLANG, but unfortunately mixed case names cannot be trapped). Note `l@` is encapsulated, so that its case does not change.

```

2790 \def\bbl@fixname#1{%
2791   \begingroup
2792   \def\bbl@tempe{#1}%
2793   \edef\bbl@tempd{\noexpand\@ifundefined{\noexpand\bbl@tempe#1}}%
2794   \bbl@tempd
2795   {\lowercase\expandafter{\bbl@tempd}}%
2796   {\uppercase\expandafter{\bbl@tempd}}%
2797   \@empty
2798   {\edef\bbl@tempd{\def\noexpand#1{#1}}}%
2799   {\uppercase\expandafter{\bbl@tempd}}}%
2800   {\edef\bbl@tempd{\def\noexpand#1{#1}}}%
2801   {\lowercase\expandafter{\bbl@tempd}}}%
2802   \@empty
2803   \edef\bbl@tempd{\endgroup\def\noexpand#1{#1}}%
2804   \bbl@tempd}
2805 \def\bbl@iflanguage#1{%
2806   \@ifundefined{l@#1}{\@nolanerr{#1}\@gobble}\@firstofone}

```

`\iflanguage` Users might want to test (in a private package for instance) which language is currently active. For this we provide a test macro, `\iflanguage`, that has three arguments. It checks whether the first argument is a known language. If so, it compares the first argument with the value of `\language`. Then, depending on the result of the comparison, it executes either the second or the third argument.

```

2807 \def\iflanguage#1{%
2808   \bbl@iflanguage{#1}{%
2809     \ifnum\csname l@#1\endcsname=\language
2810       \expandafter\@firstoftwo
2811     \else
2812       \expandafter\@secondoftwo
2813     \fi}}

```

## 11.1 Selecting the language

`\selectlanguage` The macro `\selectlanguage` checks whether the language is already defined before it performs its actual task, which is to update `\language` and activate language-specific definitions.

To allow the call of `\selectlanguage` either with a control sequence name or with a simple string as argument, we have to use a trick to delete the optional escape character. To convert a control sequence to a string, we use the `\string` primitive. Next we have to look at the first character of this string and compare it with the escape character. Because this escape character can be changed by setting the internal integer `\escapechar` to a character number, we have to compare this number with the character of the string. To do this we have to use T<sub>E</sub>X’s backquote notation to specify the character as a number. If the first character of the `\string`’ed argument is the current escape character, the comparison has stripped this character and the rest in the ‘then’ part consists of the rest of

the control sequence name. Otherwise we know that either the argument is not a control sequence or `\escapechar` is set to a value outside of the character range 0–255. If the user gives an empty argument, we provide a default argument for `\string`. This argument should expand to nothing.

```
2814 \let\bbl@select@type\z@
2815 \edef\selectlanguage{%
2816   \noexpand\protect
2817   \expandafter\noexpand\csname selectlanguage \endcsname}
```

Because the command `\selectlanguage` could be used in a moving argument it expands to `\protect\selectlanguageL`. Therefore, we have to make sure that a macro `\protect` exists. If it doesn't it is `\let` to `\relax`.

```
2818 \ifx\@undefined\protect\let\protect\relax\fi
```

As  $\text{\LaTeX}$  2.09 writes to files *expanded* whereas  $\text{\LaTeX}$  2<sub>ε</sub> takes care *not* to expand the arguments of `\write` statements we need to be a bit clever about the way we add information to .aux files. Therefore we introduce the macro `\xstring` which should expand to the right amount of `\string`'s.

```
2819 \ifx\documentclass\@undefined
2820   \def\xstring{\string\string\string}
2821 \else
2822   \let\xstring\string
2823 \fi
```

Since version 3.5 babel writes entries to the auxiliary files in order to typeset table of contents etc. in the correct language environment.

`\bbl@pop@language` But when the language change happens *inside* a group the end of the group doesn't write anything to the auxiliary files. Therefore we need  $\text{\TeX}$ 's `aftergroup` mechanism to help us. The command `\aftergroup` stores the token immediately following it to be executed when the current group is closed. So we define a temporary control sequence `\bbl@pop@language` to be executed at the end of the group. It calls `\bbl@set@language` with the name of the current language as its argument.

`\bbl@language@stack` The previous solution works for one level of nesting groups, but as soon as more levels are used it is no longer adequate. For that case we need to keep track of the nested languages using a stack mechanism. This stack is called `\bbl@language@stack` and initially empty.

```
2824 \def\bbl@language@stack{}
```

When using a stack we need a mechanism to push an element on the stack and to retrieve the information afterwards.

`\bbl@push@language` The stack is simply a list of languagenames, separated with a '+' sign; the push function can be simple:

`\bbl@pop@language`

```
2825 \def\bbl@push@language{%
2826   \xdef\bbl@language@stack{\language+\bbl@language@stack}}
```

Retrieving information from the stack is a little bit less simple, as we need to remove the element from the stack while storing it in the macro `\language`. For this we first define a helper function.

`\bbl@pop@lang` This macro stores its first element (which is delimited by the '+'-sign) in `\language` and stores the rest of the string (delimited by '-') in its third argument.

```
2827 \def\bbl@pop@lang#1+#2-#3{%
2828   \edef\language{#1}\xdef#3{#2}}
```

The reason for the somewhat weird arrangement of arguments to the helper function is the fact it is called in the following way. This means that before `\bbl@pop@lang` is executed TeX first *expands* the stack, stored in `\bbl@language@stack`. The result of that is that the argument string of `\bbl@pop@lang` contains one or more language names, each followed by a ‘+’-sign (zero language names won’t occur as this macro will only be called after something has been pushed on the stack) followed by the ‘-’-sign and finally the reference to the stack.

```
2829 \let\bbl@ifrestoring\@secondoftwo
2830 \def\bbl@pop@language{%
2831   \expandafter\bbl@pop@lang\bbl@language@stack-\bbl@language@stack
2832   \let\bbl@ifrestoring\@firstoftwo
2833   \expandafter\bbl@set@language\expandafter{\language}%
2834   \let\bbl@ifrestoring\@secondoftwo}
```

Once the name of the previous language is retrieved from the stack, it is fed to `\bbl@set@language` to do the actual work of switching everything that needs switching.

An alternative way to identify languages (in the babel sense) with a numerical value is introduced in 3.30. This is one of the first steps for a new interface based on the concept of locale, which explains the name of `\localeid`. This means `\l@...` will be reserved for hyphenation patterns.

```
2835 \chardef\localeid\z@
2836 \def\bbl@id@last{0} % No real need for a new counter
2837 \def\bbl@id@assign{%
2838   \bbl@ifunset{bbl@id@\language}%
2839   {\count@bbl@id@last\relax
2840    \advance\count@\@ne
2841    \bbl@csarg\chardef{id@\language}\count@
2842    \edef\bbl@id@last{\the\count@}%
2843    \ifcase\bbl@engine\or
2844      \directlua{
2845        Babel = Babel or {}
2846        Babel.locale_props = Babel.locale_props or {}
2847        Babel.locale_props[\bbl@id@last] = {}
2848        Babel.locale_ids = Babel.locale_ids or {}
2849        Babel.locale_ids['\language'] = \bbl@id@last
2850      }%
2851    \fi}%
2852  }%
2853  \chardef\localeid\@nameuse{bbl@id@\language}}
```

The unprotected part of `\selectlanguage`.

```
2854 \expandafter\def\csname selectlanguage \endcsname#1{%
2855   \ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel\tw@\fi
2856   \bbl@push@language
2857   \aftergroup\bbl@pop@language
2858   \bbl@set@language{#1}}
```

`\bbl@set@language` The macro `\bbl@set@language` takes care of switching the language environment *and* of writing entries on the auxiliary files. For historical reasons, language names can be either language of `\language`. To catch either form a trick is used, but unfortunately as a side effect the catcodes of letters in `\language` are messed up. This is a bug, but preserved for backwards compatibility. The list of auxiliary files can be extended by redefining `\BabelContentsFiles`, but make sure they are loaded inside a group (as `aux`, `toc`, `lof`, and `lot` do) or the last language of the document will remain active afterwards. We also write a command to change the current language in the auxiliary files.

```
2859 \def\BabelContentsFiles{toc,lof,lot}
```

```

2860 \def\bbl@set@language#1{% from selectlanguage, pop@
2861   \edef\language#1%
2862   \ifnum\escapechar=\expandafter`\string#1\@empty
2863     \else\string#1\@empty\fi}%
2864   \select@language{\language}%
2865   % write to auxs
2866   \expandafter\ifx\csname date\language\endcsname\relax\else
2867     \if@filesw
2868       \ifx\babel@aux\@gobbletwo\else % Set if single in the first, redundant
2869         \protected@write\@auxout{}\string\babel@aux{\language}{}}%
2870       \fi
2871       \bbl@usehooks{write}{}}%
2872     \fi
2873   \fi}
2874 \def\select@language#1{% from set@, babel@aux
2875   % set hmap
2876   \ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi
2877   % set name
2878   \edef\language{#1}%
2879   \bbl@fixname\language
2880   \bbl@iflanguage\language{%
2881     \expandafter\ifx\csname date\language\endcsname\relax
2882       \bbl@error
2883       {Unknown language `#1'. Either you have\\%
2884        misspelled its name, it has not been installed,\\%
2885        or you requested it in a previous run. Fix its name,\\%
2886        install it or just rerun the file, respectively. In\\%
2887        some cases, you may need to remove the aux file}%
2888       {You may proceed, but expect wrong results}%
2889     \else
2890       % set type
2891       \let\bbl@select@type\z@
2892       \expandafter\bbl@switch\expandafter{\language}%
2893     \fi}}
2894 \def\babel@aux#1#2{%
2895   \expandafter\ifx\csname date#1\endcsname\relax
2896     \expandafter\ifx\csname bbl@auxwarn@#1\endcsname\relax
2897       \@namedef{bbl@auxwarn@#1}{}%
2898       \bbl@warning
2899       {Unknown language `#1'. Very likely you\\%
2900        requested it in a previous run. Expect some\\%
2901        wrong results in this run, which should vanish\\%
2902        in the next one. Reported}%
2903     \fi
2904   \else
2905     \select@language{#1}%
2906     \bbl@foreach\BabelContentsFiles{%
2907       \@writefile{##1}{\babel@toc{#1}{#2}}}% % TODO - ok in plain?
2908   \fi}
2909 \def\babel@toc#1#2{%
2910   \select@language{#1}}

```

A bit of optimization. Select in heads/foots the language only if necessary. The real thing is in babel.def.

```
2911 \let\select@language\x\select@language
```

First, check if the user asks for a known language. If so, update the value of `\language` and call `\originalTeX` to bring  $\TeX$  in a certain pre-defined state.

The name of the language is stored in the control sequence `\language`.



Then we have to *redefine* `\originalTeX` to compensate for the things that have been activated. To save memory space for the macro definition of `\originalTeX`, we construct the control sequence name for the `\noextras<lang>` command at definition time by expanding the `\csname` primitive.

Now activate the language-specific definitions. This is done by constructing the names of three macros by concatenating three words with the argument of `\selectlanguage`, and calling these macros.

The switching of the values of `\lefthyphenmin` and `\righthyphenmin` is somewhat different. First we save their current values, then we check if `\<lang>hyphenmins` is defined. If it is not, we set default values (2 and 3), otherwise the values in `\<lang>hyphenmins` will be used.

```

2912 \newif\ifbbl@usedategroup
2913 \def\bbl@switch#1{% from select@, foreign@
2914   % restore
2915   \originalTeX
2916   \expandafter\def\expandafter\originalTeX\expandafter{%
2917     \csname noextras#1\endcsname
2918     \let\originalTeX\@empty
2919     \babel@beginsave}%
2920   \bbl@usehooks{afterreset}}}%
2921   \languageshorthands{none}%
2922   % set the locale id
2923   \bbl@id@assign
2924   % switch captions, date
2925   \ifcase\bbl@select@type
2926     \ifhmode
2927       \hskip\z@skip % trick to ignore spaces
2928       \csname captions#1\endcsname\relax
2929       \csname date#1\endcsname\relax
2930       \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
2931     \else
2932       \csname captions#1\endcsname\relax
2933       \csname date#1\endcsname\relax
2934     \fi
2935   \else
2936     \ifbbl@usedategroup % if \foreign... within \<lang>date
2937       \bbl@usedategroupfalse
2938       \ifhmode
2939         \hskip\z@skip % trick to ignore spaces
2940         \csname date#1\endcsname\relax
2941         \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
2942       \else
2943         \csname date#1\endcsname\relax
2944       \fi
2945     \fi
2946   \fi
2947   % switch extras
2948   \bbl@usehooks{beforeextras}}}%
2949   \csname extras#1\endcsname\relax
2950   \bbl@usehooks{afterextras}}}%
2951   % > babel-ensure
2952   % > babel-sh-<short>
2953   % > babel-bidi
2954   % > babel-fontspec
2955   % hyphenation - case mapping
2956   \ifcase\bbl@opt@hyphenmap\or
2957     \def\BabelLower##1##2{\lccode##1=##2\relax}%

```

```

2958 \ifnum\bb1@hymapsel>4\else
2959 \csname\language @bb1@hyphenmap\endcsname
2960 \fi
2961 \chardef\bb1@opt@hyphenmap\z@
2962 \else
2963 \ifnum\bb1@hymapsel>\bb1@opt@hyphenmap\else
2964 \csname\language @bb1@hyphenmap\endcsname
2965 \fi
2966 \fi
2967 \global\let\bb1@hymapsel\@cclv
2968 % hyphenation - patterns
2969 \bb1@patterns{#1}%
2970 % hyphenation - mins
2971 \babel@savevariable\lefthyphenmin
2972 \babel@savevariable\righthyphenmin
2973 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2974 \set@hyphenmins\tw@\thr@@\relax
2975 \else
2976 \expandafter\expandafter\expandafter\set@hyphenmins
2977 \csname #1hyphenmins\endcsname\relax
2978 \fi}

```

`otherlanguage` The `otherlanguage` environment can be used as an alternative to using the `\selectlanguage` declarative command. When you are typesetting a document which mixes left-to-right and right-to-left typesetting you have to use this environment in order to let things work as you expect them to. The `\ignorespaces` command is necessary to hide the environment when it is entered in horizontal mode.

```

2979 \long\def\otherlanguage#1{%
2980 \ifnum\bb1@hymapsel=\@cclv\let\bb1@hymapsel\thr@@\fi
2981 \csname selectlanguage \endcsname{#1}%
2982 \ignorespaces}

```

The `\endotherlanguage` part of the environment tries to hide itself when it is called in horizontal mode.

```

2983 \long\def\endotherlanguage{%
2984 \global\@ignoretrue\ignorespaces}

```

`otherlanguage*` The `otherlanguage` environment is meant to be used when a large part of text from a different language needs to be typeset, but without changing the translation of words such as ‘figure’. This environment makes use of `\foreign@language`.

```

2985 \expandafter\def\csname otherlanguage*\endcsname#1{%
2986 \ifnum\bb1@hymapsel=\@cclv\chardef\bb1@hymapsel4\relax\fi
2987 \foreign@language{#1}}

```

At the end of the environment we need to switch off the extra definitions. The grouping mechanism of the environment will take care of resetting the correct hyphenation rules and “extras”.

```

2988 \expandafter\let\csname endotherlanguage*\endcsname\relax

```

`\foreignlanguage` The `\foreignlanguage` command is another substitute for the `\selectlanguage` command. This command takes two arguments, the first argument is the name of the language to use for typesetting the text specified in the second argument. Unlike `\selectlanguage` this command doesn’t switch *everything*, it only switches the hyphenation rules and the extra definitions for the language specified. It does this within a group and assumes the `\extras⟨lang⟩` command doesn’t make any `\global` changes. The coding is very similar to part of `\selectlanguage`.

`\bbl@beforeforeign` is a trick to fix a bug in bidi texts. `\foreignlanguage` is supposed to be a ‘text’ command, and therefore it must emit a `\leavevmode`, but it does not, and therefore the indent is placed on the opposite margin. For backward compatibility, however, it is done only if a right-to-left script is requested; otherwise, it is no-op.

(3.11) `\foreignlanguage*` is a temporary, experimental macro for a few lines with a different script direction, while preserving the paragraph format (thank the braces around `\par`, things like `\hangindent` are not reset). Do not use it in production, because its semantics and its syntax may change (and very likely will, or even it could be removed altogether). Currently it enters in vmode and then selects the language (which in turn sets the paragraph direction).

(3.11) Also experimental are the hook `foreign` and `foreign*`. With them you can redefine `\BabelText` which by default does nothing. Its behavior is not well defined yet. So, use it in horizontal mode only if you do not want surprises.

In other words, at the beginning of a paragraph `\foreignlanguage` enters into hmode with the surrounding lang, and with `\foreignlanguage*` with the new lang.

```

2989 \providecommand\bbl@beforeforeign{}
2990 \edef\foreignlanguage{%
2991   \noexpand\protect
2992   \expandafter\noexpand\csname foreignlanguage \endcsname}
2993 \expandafter\def\csname foreignlanguage \endcsname{%
2994   \@ifstar\bbl@foreign@s\bbl@foreign@x}
2995 \def\bbl@foreign@x#1#2{%
2996   \begingroup
2997     \let\BabelText\@firstofone
2998     \bbl@beforeforeign
2999     \foreign@language{#1}%
3000     \bbl@usehooks{foreign}{}%
3001     \BabelText{#2}% Now in horizontal mode!
3002   \endgroup}
3003 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \setpar, ?\@par
3004   \begingroup
3005     {\par}%
3006     \let\BabelText\@firstofone
3007     \foreign@language{#1}%
3008     \bbl@usehooks{foreign*}{}%
3009     \bbl@dirparastext
3010     \BabelText{#2}% Still in vertical mode!
3011     {\par}%
3012   \endgroup}

```

`\foreign@language` This macro does the work for `\foreignlanguage` and the other `language*` environment. First we need to store the name of the language and check that it is a known language. Then it just calls `bbl@switch`.

```

3013 \def\foreign@language#1{%
3014   % set name
3015   \edef\language#1%
3016   \bbl@fixname\language
3017   \bbl@iflanguage\language%
3018     \expandafter\ifx\csname date\language\endcsname\relax
3019       \bbl@warning % TODO - why a warning, not an error?
3020       {Unknown language `#1'. Either you have\\%
3021        misspelled its name, it has not been installed,\\%
3022        or you requested it in a previous run. Fix its name,\\%
3023        install it or just rerun the file, respectively. In\\%
3024        some cases, you may need to remove the aux file.\\%
3025        I'll proceed, but expect wrong results.\\%
3026        Reported}%

```

```

3027 \fi
3028 % set type
3029 \let\bbl@select@type\@ne
3030 \expandafter\bbl@switch\expandafter{\language}}

```

**\bbl@patterns** This macro selects the hyphenation patterns by changing the `\language` register. If special hyphenation patterns are available specifically for the current font encoding, use them instead of the default.

It also sets hyphenation exceptions, but only once, because they are global (here language `\lccode`'s has been set, too). `\bbl@hyphenation@` is set to relax until the very first `\babelhyphenation`, so do nothing with this value. If the exceptions for a language (by its number, not its name, so that `:ENC` is taken into account) has been set, then use `\hyphenation` with both global and language exceptions and empty the latter to mark they must not be set again.

```

3031 \let\bbl@hyphlist\@empty
3032 \let\bbl@hyphenation@relax
3033 \let\bbl@pttnlist\@empty
3034 \let\bbl@patterns@relax
3035 \let\bbl@hymapsel=\@ccclv
3036 \def\bbl@patterns#1{%
3037   \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
3038     \csname l@#1\endcsname
3039     \edef\bbl@tempa{#1}%
3040   \else
3041     \csname l@#1:\f@encoding\endcsname
3042     \edef\bbl@tempa{#1:\f@encoding}%
3043   \fi
3044   \@expandtwoargs\bbl@usehooks{patterns}{#1}{\bbl@tempa}%
3045   % > luatex
3046   \@ifundefined{bbl@hyphenation@}{% Can be \relax!
3047     \begingroup
3048       \bbl@xin@{,\number\language,}{,\bbl@hyphlist}%
3049     \ifin@else
3050       \@expandtwoargs\bbl@usehooks{hyphenation}{#1}{\bbl@tempa}%
3051       \hyphenation{%
3052         \bbl@hyphenation@
3053         \@ifundefined{bbl@hyphenation@#1}%
3054         \@empty
3055         {\space\csname bbl@hyphenation@#1\endcsname}}%
3056       \xdef\bbl@hyphlist{\bbl@hyphlist\number\language,}%
3057     \fi
3058   \endgroup}}

```

**hyphenrules** The environment `hyphenrules` can be used to select *just* the hyphenation rules. This environment does *not* change `\language` and when the hyphenation rules specified were not loaded it has no effect. Note however, `\lccode`'s and font encodings are not set at all, so in most cases you should use other language\*.

```

3059 \def\hyphenrules#1{%
3060   \edef\bbl@tempf{#1}%
3061   \bbl@fixname\bbl@tempf
3062   \bbl@iflanguage\bbl@tempf{%
3063     \expandafter\bbl@patterns\expandafter{\bbl@tempf}%
3064     \languageshortands{none}%
3065     \expandafter\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
3066       \set@hyphenmins\tw@\thr@@\relax
3067     \else
3068       \expandafter\expandafter\expandafter\set@hyphenmins

```

```

3069     \csname\bb1@tempf hyphenmins\endcsname\relax
3070     \fi}}
3071 \let\endhyphenrules\@empty

```

`\providehyphenmins` The macro `\providehyphenmins` should be used in the language definition files to provide a *default* setting for the hyphenation parameters `\lefthyphenmin` and `\righthyphenmin`. If the macro `\(lang)hyphenmins` is already defined this command has no effect.

```

3072 \def\providehyphenmins#1#2{%
3073     \expandafter\ifx\csname #1hyphenmins\endcsname\relax
3074         \@namedef{#1hyphenmins}{#2}%
3075     \fi}

```

`\set@hyphenmins` This macro sets the values of `\lefthyphenmin` and `\righthyphenmin`. It expects two values as its argument.

```

3076 \def\set@hyphenmins#1#2{%
3077     \lefthyphenmin#1\relax
3078     \righthyphenmin#2\relax}

```

`\ProvidesLanguage` The identification code for each file is something that was introduced in  $\text{\LaTeX 2}_{\epsilon}$ . When the command `\ProvidesFile` does not exist, a dummy definition is provided temporarily. For use in the language definition file the command `\ProvidesLanguage` is defined by `babel`. Depending on the format, ie, on if the former is defined, we use a similar definition or not.

```

3079 \ifx\ProvidesFile\@undefined
3080     \def\ProvidesLanguage#1[#2 #3 #4]{%
3081         \wlog{Language: #1 #4 #3 <#2>}%
3082     }
3083 \else
3084     \def\ProvidesLanguage#1{%
3085         \begingroup
3086         \catcode`\ 10 %
3087         \@makeother\/%
3088         \@ifnextchar[%]
3089             {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}
3090     \def\@provideslanguage#1[#2]{%
3091         \wlog{Language: #1 #2}%
3092         \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
3093     \endgroup}
3094 \fi

```

`\LdfInit` This macro is defined in two versions. The first version is to be part of the ‘kernel’ of `babel`, ie. the part that is loaded in the format; the second version is defined in `babel.def`. The version in the format just checks the category code of the ampersand and then loads `babel.def`. The category code of the ampersand is restored and the macro calls itself again with the new definition from `babel.def`

```

3095 \def\LdfInit{%
3096     \chardef\atcatcode=\catcode`\@
3097     \catcode`\@=11\relax
3098     \input babel.def\relax
3099     \catcode`\@=\atcatcode \let\atcatcode\relax
3100     \LdfInit}

```

`\originalTeX` The macro `\originalTeX` should be known to  $\text{\TeX}$  at this moment. As it has to be expandable we `\let` it to `\@empty` instead of `\relax`.

```

3101 \ifx\originalTeX\@undefined\let\originalTeX\@empty\fi

```

Because this part of the code can be included in a format, we make sure that the macro which initialises the save mechanism, `\babel@beginsave`, is not considered to be undefined.

```
3102 \ifx\babel@beginsave\@undefined\let\babel@beginsave\relax\fi
```

A few macro names are reserved for future releases of babel, which will use the concept of ‘locale’:

```
3103 \providecommand\setlocale{%
3104   \bbl@error
3105   {Not yet available}%
3106   {Find an armchair, sit down and wait}}
3107 \let\uselocale\setlocale
3108 \let\locale\setlocale
3109 \let\selectlocale\setlocale
3110 \let\textlocale\setlocale
3111 \let\textlanguage\setlocale
3112 \let\language\setlocale
```

## 11.2 Errors

`\@nolanerr` The babel package will signal an error when a documents tries to select a language that hasn’t been defined earlier. When a user selects a language for which no hyphenation patterns were loaded into the format he will be given a warning about that fact. We revert to the patterns for `\language=0` in that case. In most formats that will be (US)english, but it might also be empty.

`\@noopterr` When the package was loaded without options not everything will work as expected. An error message is issued in that case.  
When the format knows about `\PackageError` it must be  $\text{\LaTeX 2}_{\epsilon}$ , so we can safely use its error handling interface. Otherwise we’ll have to ‘keep it simple’.

```
3113 \edef\bbl@nulllanguage{\string\language=0}
3114 \ifx\PackageError\@undefined
3115   \def\bbl@error#1#2{%
3116     \begingroup
3117     \newlinechar=`^^J
3118     \def\{^^J(babel) }%
3119     \errhelp{#2}\errmessage{\{#1}%
3120     \endgroup}
3121   \def\bbl@warning#1{%
3122     \begingroup
3123     \newlinechar=`^^J
3124     \def\{^^J(babel) }%
3125     \message{\{#1}%
3126     \endgroup}
3127   \let\bbl@infowarn\bbl@warning
3128   \def\bbl@info#1{%
3129     \begingroup
3130     \newlinechar=`^^J
3131     \def\{^^J}%
3132     \wlog{#1}%
3133     \endgroup}
3134 \else
3135   \def\bbl@error#1#2{%
3136     \begingroup
3137     \def\{\MessageBreak}%
3138     \PackageError{babel}{#1}{#2}%
3139     \endgroup}
```

```

3140 \def\bbl@warning#1{%
3141   \begingroup
3142     \def\{\MessageBreak}%
3143     \PackageWarning{babel}{#1}%
3144   \endgroup}
3145 \def\bbl@infowarn#1{%
3146   \begingroup
3147     \def\{\MessageBreak}%
3148     \GenericWarning
3149       {(babel) \spaces\spaces\spaces}%
3150       {Package babel Info: #1}%
3151   \endgroup}
3152 \def\bbl@info#1{%
3153   \begingroup
3154     \def\{\MessageBreak}%
3155     \PackageInfo{babel}{#1}%
3156   \endgroup}
3157 \fi
3158 \@ifpackagewith{babel}{silent}
3159 {\let\bbl@info@gobble
3160  \let\bbl@infowarn@gobble
3161  \let\bbl@warning@gobble}
3162 {}
3163 \def\bbl@nocaption{\protect\bbl@nocaption@i}
3164 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
3165   \global\@namedef{#2}{\textbf{?#1?}}%
3166   \@nameuse{#2}%
3167   \bbl@warning%
3168     \@backslashchar#2 not set. Please, define\\%
3169     it in the preamble with something like:\\%
3170     \string\renewcommand\@backslashchar#2{..}\\%
3171     Reported}}
3172 \def\bbl@tentative{\protect\bbl@tentative@i}
3173 \def\bbl@tentative@i#1{%
3174   \bbl@warning%
3175     Some functions for '#1' are tentative.\\%
3176     They might not work as expected and their behavior\\%
3177     could change in the future.\\%
3178     Reported}}
3179 \def\@nolanerr#1{%
3180   \bbl@error
3181     {You haven't defined the language #1\space yet}%
3182     {Your command will be ignored, type <return> to proceed}}
3183 \def\@nopatterns#1{%
3184   \bbl@warning
3185     {No hyphenation patterns were preloaded for\\%
3186     the language `#1' into the format.\\%
3187     Please, configure your TeX system to add them and\\%
3188     rebuild the format. Now I will use the patterns\\%
3189     preloaded for \bbl@nulllanguage\space instead}}
3190 \let\bbl@usehooks@gobbletwo
3191 </kernel>
3192 <*patterns>

```

## 12 Loading hyphenation patterns

The following code is meant to be read by `iniTEX` because it should instruct `TEX` to read hyphenation patterns. To this end the `docstrip` option `patterns` can be used to include this code in the file `hyphen.cfg`. Code is written with lower level macros.

We want to add a message to the message `LATEX 2.09` puts in the `\everyjob` register. This could be done by the following code:

```

\let\orgeveryjob\everyjob
\def\everyjob#1{%
\orgeveryjob{#1}%
\orgeveryjob\expandafter{\the\orgeveryjob\immediate\write16{%
\hyphenation_patterns_for_the_loaded_patterns_loaded.}}%
\let\everyjob\orgeveryjob\let\orgeveryjob\@undefined}

```

The code above redefines the control sequence `\everyjob` in order to be able to add something to the current contents of the register. This is necessary because the processing of hyphenation patterns happens long before `LATEX` fills the register.

There are some problems with this approach though.

- When someone wants to use several hyphenation patterns with `LATEX` the above scheme won't work. The reason is that `LATEX` overwrites the contents of the `\everyjob` register with its own message.
- Plain `TEX` does not use the `\everyjob` register so the message would not be displayed.

To circumvent this a 'dirty trick' can be used. As this code is only processed when creating a new format file there is one command that is sure to be used, `\dump`. Therefore the original `\dump` is saved in `\org@dump` and a new definition is supplied.

To make sure that `LATEX 2.09` executes the `\@begindocumenthook` we would want to alter `\begin{document}`, but as this done too often already, we add the new code at the front of `\@preamblecmds`. But we can only do that after it has been defined, so we add this piece of code to `\dump`.

This new definition starts by adding an instruction to write a message on the terminal and in the transcript file to inform the user of the preloaded hyphenation patterns.

Then everything is restored to the old situation and the format is dumped.

```

3193 <<Make sure ProvidesFile is defined>>
3194 \ProvidesFile{hyphen.cfg}[<<date>>] <<version>> Babel hyphens]
3195 \xdef\bb1@format{\jobname}
3196 \ifx\AtBeginDocument\@undefined
3197   \def\@empty{}
3198   \let\orig@dump\dump
3199   \def\dump{%
3200     \ifx\@ztryfc\@undefined
3201       \else
3202         \toks0=\expandafter{\@preamblecmds}%
3203         \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
3204         \def\@begindocumenthook{}%
3205       \fi
3206       \let\dump\orig@dump\let\orig@dump\@undefined\dump}
3207 \fi
3208 <<Define core switching macros>>

```

`\process@line` Each line in the file `language.dat` is processed by `\process@line` after it is read. The first thing this macro does is to check whether the line starts with `=`. When the first token of a



line is an =, the macro `\process@synonym` is called; otherwise the macro `\process@language` will continue.

```
3209 \def\process@line#1#2 #3 #4 {%
3210   \ifx=#1%
3211     \process@synonym{#2}%
3212   \else
3213     \process@language{#1#2}{#3}{#4}%
3214   \fi
3215   \ignorespaces}
```

`\process@synonym` This macro takes care of the lines which start with an =. It needs an empty token register to begin with. `\bbl@languages` is also set to empty.

```
3216 \toks@{}
3217 \def\bbl@languages{}
```

When no languages have been loaded yet, the name following the = will be a synonym for hyphenation register 0. So, it is stored in a token register and executed when the first pattern file has been processed. (The `\relax` just helps to the `\if` below catching synonyms without a language.)

Otherwise the name will be a synonym for the language loaded last.

We also need to copy the hyphenmin parameters for the synonym.

```
3218 \def\process@synonym#1{%
3219   \ifnum\last@language=\m@ne
3220     \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
3221   \else
3222     \expandafter\chardef\csname l@#1\endcsname\last@language
3223     \wlog{\string\l@#1=\string\language\the\last@language}%
3224     \expandafter\let\csname #1hyphenmins\endcsname\expandafter\endcsname
3225     \csname\language\endcsname hyphenmins\endcsname
3226     \let\bbl@elt\relax
3227     \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}}}%
3228   \fi}
```

`\process@language` The macro `\process@language` is used to process a non-empty line from the ‘configuration file’. It has three arguments, each delimited by white space. The first argument is the ‘name’ of a language; the second is the name of the file that contains the patterns. The optional third argument is the name of a file containing hyphenation exceptions. The first thing to do is call `\addlanguage` to allocate a pattern register and to make that register ‘active’. Then the pattern file is read.

For some hyphenation patterns it is needed to load them with a specific font encoding selected. This can be specified in the file `language.dat` by adding for instance ‘:T1’ to the name of the language. The macro `\bbl@get@enc` extracts the font encoding from the language name and stores it in `\bbl@hyph@enc`. The latter can be used in hyphenation files if you need to set a behavior depending on the given encoding (it is set to empty if no encoding is given).

Pattern files may contain assignments to `\lefthyphenmin` and `\righthyphenmin`.  $\TeX$  does not keep track of these assignments. Therefore we try to detect such assignments and store them in the `\<lang>hyphenmins` macro. When no assignments were made we provide a default setting.

Some pattern files contain changes to the `\lccode` and `\uccode` arrays. Such changes should remain local to the language; therefore we process the pattern file in a group; the `\patterns` command acts globally so its effect will be remembered.

Then we globally store the settings of `\lefthyphenmin` and `\righthyphenmin` and close the group.

When the hyphenation patterns have been processed we need to see if a file with hyphenation exceptions needs to be read. This is the case when the third argument is not

empty and when it does not contain a space token. (Note however there is no need to save hyphenation exceptions into the format.)

\bbl@languages saves a snapshot of the loaded languages in the form

\bbl@elt{<language-name>}{<number>}{<patterns-file>}{<exceptions-file>}. Note the last 2 arguments are empty in ‘dialects’ defined in language.dat with =. Note also the language name can have encoding info.

Finally, if the counter \language is equal to zero we execute the synonyms stored.

```

3229 \def\process@language#1#2#3{%
3230   \expandafter\addlanguage\csname l@#1\endcsname
3231   \expandafter\language\csname l@#1\endcsname
3232   \edef\language#1{%
3233     \bbl@hook@everylanguage{#1}%
3234     % > luatex
3235     \bbl@get@enc#1::\@@@
3236     \beginngroup
3237       \lefthyphenmin\m@ne
3238       \bbl@hook@loadpatterns{#2}%
3239       % > luatex
3240       \ifnum\lefthyphenmin=\m@ne
3241         \else
3242           \expandafter\xdef\csname #1hyphenmins\endcsname{%
3243             \the\lefthyphenmin\the\righthyphenmin}%
3244           \fi
3245         \endgroup
3246       \def\bbl@tempa{#3}%
3247       \ifx\bbl@tempa\@empty\else
3248         \bbl@hook@loadexceptions{#3}%
3249         % > luatex
3250       \fi
3251       \let\bbl@elt\relax
3252       \edef\bbl@languages{%
3253         \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
3254       \ifnum\the\language=\z@
3255         \expandafter\ifx\csname #1hyphenmins\endcsname\relax
3256           \set@hyphenmins\tw@\thr@@\relax
3257         \else
3258           \expandafter\expandafter\expandafter\set@hyphenmins
3259             \csname #1hyphenmins\endcsname
3260         \fi
3261         \the\toks@
3262         \toks@{}%
3263       \fi}

```

\bbl@get@enc The macro \bbl@get@enc extracts the font encoding from the language name and stores it in \bbl@hyph@enc. It uses delimited arguments to achieve this.

```

3264 \def\bbl@get@enc#1:#2:#3\@@{\def\bbl@hyph@enc{#2}}

```

Now, hooks are defined. For efficiency reasons, they are dealt here in a special way. Besides luatex, format-specific configuration files are taken into account.

```

3265 \def\bbl@hook@everylanguage#1{}
3266 \def\bbl@hook@loadpatterns#1{\input #1\relax}
3267 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
3268 \let\bbl@hook@loadkernel\bbl@hook@loadpatterns
3269 \beginngroup
3270   \def\AddBabelHook#1#2{%
3271     \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
3272       \def\next{\toks1}%

```

```

3273 \else
3274 \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname####1}%
3275 \fi
3276 \next}
3277 \ifx\directlua\@undefined
3278 \ifx\XeTeXinputencoding\@undefined\else
3279 \input xebabel.def
3280 \fi
3281 \else
3282 \input luababel.def
3283 \fi
3284 \openin1 = babel-\bbl@format.cfg
3285 \ifeof1
3286 \else
3287 \input babel-\bbl@format.cfg\relax
3288 \fi
3289 \closein1
3290 \endgroup
3291 \bbl@hook@loadkernel{switch.def}

```

\readconfigfile The configuration file can now be opened for reading.

```

3292 \openin1 = language.dat

```

See if the file exists, if not, use the default hyphenation file hyphen.tex. The user will be informed about this.

```

3293 \def\language{english}%
3294 \ifeof1
3295 \message{I couldn't find the file language.dat,\space
3296          I will try the file hyphen.tex}
3297 \input hyphen.tex\relax
3298 \chardef\l@english\z@
3299 \else

```

Pattern registers are allocated using count register \last@language. Its initial value is 0. The definition of the macro \newlanguage is such that it first increments the count register and then defines the language. In order to have the first patterns loaded in pattern register number 0 we initialize \last@language with the value -1.

```

3300 \last@language\m@ne

```

We now read lines from the file until the end is found

```

3301 \loop

```

While reading from the input, it is useful to switch off recognition of the end-of-line character. This saves us stripping off spaces from the contents of the control sequence.

```

3302 \endlinechar\m@ne
3303 \read1 to \bbl@line
3304 \endlinechar\^^M

```

If the file has reached its end, exit from the loop here. If not, empty lines are skipped. Add 3 space characters to the end of \bbl@line. This is needed to be able to recognize the arguments of \process@line later on. The default language should be the very first one.

```

3305 \if T\ifeof1F\fi T\relax
3306 \ifx\bbl@line\@empty\else
3307 \edef\bbl@line{\bbl@line\space\space\space}%
3308 \expandafter\process@line\bbl@line\relax
3309 \fi
3310 \repeat

```

Check for the end of the file. We must reverse the test for `\ifeof` without `\else`. Then reactivate the default patterns.

```

3311 \begingroup
3312   \def\bbl@elt#1#2#3#4{%
3313     \global\language=#2\relax
3314     \gdef\language{#1}%
3315     \def\bbl@elt##1##2##3##4{}}%
3316   \bbl@languages
3317 \endgroup
3318 \fi

```

and close the configuration file.

```

3319 \closein1

```

We add a message about the fact that babel is loaded in the format and with which language patterns to the `\everyjob` register.

```

3320 \if/\the\toks@/\else
3321   \errhelp{language.dat loads no language, only synonyms}
3322   \errmessage{Orphan language synonym}
3323 \fi

```

Also remove some macros from memory and raise an error if `\toks@` is not empty. Finally load `switch.def`, but the latter is not required and the line inputting it may be commented out.

```

3324 \let\bbl@line\@undefined
3325 \let\process@line\@undefined
3326 \let\process@synonym\@undefined
3327 \let\process@language\@undefined
3328 \let\bbl@get@enc\@undefined
3329 \let\bbl@hyph@enc\@undefined
3330 \let\bbl@tempa\@undefined
3331 \let\bbl@hook@loadkernel\@undefined
3332 \let\bbl@hook@everylanguage\@undefined
3333 \let\bbl@hook@loadpatterns\@undefined
3334 \let\bbl@hook@loadexceptions\@undefined
3335 \let\patterns\@undefined

```

Here the code for `iniTEX` ends.

## 13 Font handling with fontspec

Add the bidi handler just before `luaotfload`, which is loaded by default by LaTeX. Just in case, consider the possibility it has not been loaded. First, a couple of definitions related to bidi [misplaced].

```

3336 <<(*More package options)>> ≡
3337 \ifodd\bbl@engine
3338   \DeclareOption{bidi=basic-r}%
3339   {\ExecuteOptions{bidi=basic}}
3340   \DeclareOption{bidi=basic}%
3341   {\let\bbl@beforeforeign\leavevmode
3342     % TODO - to locale_props, not as separate attribute
3343     \newattribute\bbl@attr@dir
3344     % I don't like it, hackish:
3345     \frozen@everymath\expandafter{%
3346       \expandafter\bbl@mathboxdir\the\frozen@everymath}%
3347     \frozen@everydisplay\expandafter{%
3348       \expandafter\bbl@mathboxdir\the\frozen@everydisplay}%

```

```

3349     \bbl@exp{\output{\bodydir\pagedir\the\output}}%
3350     \AtEndOfPackage{\EnableBabelHook{babel-bidi}}%
3351 \else
3352   \DeclareOption{bidi=basic-r}%
3353   {\ExecuteOptions{bidi=basic}}
3354   \DeclareOption{bidi=basic}%
3355   {\bbl@error
3356     {The bidi method 'basic' is available only in\\%
3357       luatex. I'll continue with 'bidi=default', so\\%
3358       expect wrong results}%
3359     {See the manual for further details.}%
3360   \let\bbl@beforeforeign\leavevmode
3361   \AtEndOfPackage{%
3362     \EnableBabelHook{babel-bidi}%
3363     \bbl@xebidipar}}
3364   \def\bbl@loadxebidi#1{%
3365     \ifx\RTLfootnotetext\undefined
3366       \AtEndOfPackage{%
3367         \EnableBabelHook{babel-bidi}%
3368         \ifx\fontspec\undefined
3369           \usepackage{fontspec}% bidi needs fontspec
3370         \fi
3371         \usepackage#1{bidi}}%
3372     \fi}
3373   \DeclareOption{bidi=bidi}%
3374   {\bbl@tentative{bidi=bidi}%
3375     \bbl@loadxebidi{}}
3376   \DeclareOption{bidi=bidi-r}%
3377   {\bbl@tentative{bidi=bidi-r}%
3378     \bbl@loadxebidi{[rldocument]}}
3379   \DeclareOption{bidi=bidi-l}%
3380   {\bbl@tentative{bidi=bidi-l}%
3381     \bbl@loadxebidi{}}
3382 \fi
3383 \DeclareOption{bidi=default}%
3384 {\let\bbl@beforeforeign\leavevmode
3385   \ifodd\bbl@engine
3386     \newattribute\bbl@attr@dir
3387     \bbl@exp{\output{\bodydir\pagedir\the\output}}%
3388   \fi
3389   \AtEndOfPackage{%
3390     \EnableBabelHook{babel-bidi}%
3391     \ifodd\bbl@engine\else
3392       \bbl@xebidipar
3393     \fi}}
3394 <</More package options>>

```

With explicit languages, we could define the font at once, but we don't. Just wait and see if the language is actually activated. `bbl@font` replaces hardcoded font names inside `\. . family` by the corresponding macro `\. . default`.

```

3395 << *Font selection>> ≡
3396 \bbl@trace{Font handling with fontspec}
3397 \@onlypreamble\babelfont
3398 \newcommand\babelfont[2][]{% 1=langs/scripts 2=fam
3399   \edef\bbl@tempa{#1}%
3400   \def\bbl@tempb{#2}% Used by \bbl@bblfont
3401   \ifx\fontspec\undefined
3402     \usepackage{fontspec}%
3403   \fi

```

```

3404 \EnableBabelHook{babel-fontspec}% Just calls \bbl@switchfont
3405 \bbl@bblfont}
3406 \newcommand\bbl@bblfont[2][]{% 1=features 2=fontname, @font=rm|sf|tt
3407 \bbl@ifunset{\bbl@tempb family}%
3408 {\bbl@providedefam{\bbl@tempb}}%
3409 {\bbl@exp{%
3410 \\\bbl@sreplace\<\bbl@tempb family >%
3411 {\@nameuse{\bbl@tempb default}}{\<\bbl@tempb default>}}}%
3412 % For the default font, just in case:
3413 \bbl@ifunset{\bbl@lsys@\language name}{\bbl@provide@lsys{\language name}}}%
3414 \expandafter\bbl@ifblank\expandafter{\bbl@tempa}%
3415 {\bbl@csarg\edef{\bbl@tempb dflt@}{<{#1}{#2}}% save bbl@rmdflt@
3416 \bbl@exp{%
3417 \let\<bbl@\bbl@tempb dflt@\language name>\<bbl@\bbl@tempb dflt@>%
3418 \\\bbl@font@set\<bbl@\bbl@tempb dflt@\language name>%
3419 \<\bbl@tempb default>\<\bbl@tempb family>}}%
3420 {\bbl@foreach\bbl@tempa{% ie bbl@rmdflt@lang / *scrt
3421 \bbl@csarg\def{\bbl@tempb dflt@##1}{<{#1}{#2}}}}}%

```

If the family in the previous command does not exist, it must be defined. Here is how:

```

3422 \def\bbl@providedefam#1{%
3423 \bbl@exp{%
3424 \\\newcommand\<#1default>{}% Just define it
3425 \\\bbl@add@list\\bbl@font@fams{#1}%
3426 \\\DeclareRobustCommand\<#1family>%
3427 \\\not@math@alphabet\<#1family>\relax
3428 \\\fontfamily\<#1default>\selectfont}%
3429 \\\DeclareTextFontCommand{\<text#1>}{\<#1family>}}

```

The following macro is activated when the hook babel-fontspec is enabled. But before we define a macro for a warning, which sets a flag to avoid duplicate them.

```

3430 \def\bbl@nostdfont#1{%
3431 \bbl@ifunset{\bbl@WFF@\f@family}%
3432 {\bbl@csarg\gdef{\bbl@WFF@\f@family}}{% Flag, to avoid dupl warns
3433 \bbl@infowarn{The current font is not a babel standard family:\\%
3434 #1%
3435 \fontname\font\\%
3436 There is nothing intrinsically wrong with this warning, and\\%
3437 you can ignore it altogether if you do not need these\\%
3438 families. But if they are used in the document, you should be\\%
3439 aware 'babel' will no set Script and Language for them, so\\%
3440 you may consider defining a new family with \string\babelfont.\\%
3441 See the manual for further details about \string\babelfont.\\%
3442 Reported}}
3443 {}}%
3444 \gdef\bbl@switchfont{%
3445 \bbl@ifunset{\bbl@lsys@\language name}{\bbl@provide@lsys{\language name}}}%
3446 \bbl@exp{% eg Arabic -> arabic
3447 \lowercase{\edef\\bbl@tempa{\bbl@cs{sname@\language name}}}}%
3448 \bbl@foreach\bbl@font@fams{%
3449 \bbl@ifunset{\bbl@##1dflt@\language name}% (1) language?
3450 {\bbl@ifunset{\bbl@##1dflt@*\bbl@tempa}% (2) from script?
3451 {\bbl@ifunset{\bbl@##1dflt@}% 2=F - (3) from generic?
3452 {}% 123=F - nothing!
3453 {\bbl@exp{% 3=T - from generic
3454 \global\let\<bbl@##1dflt@\language name>%
3455 \<bbl@##1dflt@>}}}%
3456 {\bbl@exp{% 2=T - from script
3457 \global\let\<bbl@##1dflt@\language name>%

```

```

3458 \<bbl@##1dflt@*\bbl@tempa>}}}%
3459 {}}% 1=T - language, already defined
3460 \def\bbl@tempa{\bbl@nostdfont{}}}%
3461 \bbl@foreach\bbl@font@fams{% don't gather with prev for
3462 \bbl@ifunset{\bbl@##1dflt@\language}%
3463 {\bbl@cs{famrst@##1}%
3464 \global\bbl@csarg\let{famrst@##1}\relax}%
3465 {\bbl@exp{% order is relevant
3466 \\bbl@add\\originalTeX{%
3467 \\bbl@font@rst{\bbl@cs{##1dflt@\language}}}%
3468 \<##1default>\<##1family>{##1}}}%
3469 \\bbl@font@set\bbl@##1dflt@\language>% the main part!
3470 \<##1default>\<##1family>}}}%
3471 \bbl@ifrestoring{ }\bbl@tempa}}}%

```

The following is executed at the beginning of the aux file or the document to warn about fonts not defined with `\babelfont`.

```

3472 \ifx\f@family\undefined\else % if latex
3473 \ifcase\bbl@engine % if pdftex
3474 \let\bbl@ckeckstdfonts\relax
3475 \else
3476 \def\bbl@ckeckstdfonts{%
3477 \begingroup
3478 \global\let\bbl@ckeckstdfonts\relax
3479 \let\bbl@tempa\@empty
3480 \bbl@foreach\bbl@font@fams{%
3481 \bbl@ifunset{\bbl@##1dflt@}%
3482 {\@nameuse{##1family}%
3483 \bbl@csarg\gdef{WFF@f@family}}}% Flag
3484 \bbl@exp{\\bbl@add\\bbl@tempa{* \<##1family>= \f@family\\
3485 \space\space\fontname\font\\}}}%
3486 \bbl@csarg\xdef{##1dflt@}{\f@family}%
3487 \expandafter\xdef\csname ##1default\endcsname{\f@family}}}%
3488 {}}%
3489 \ifx\bbl@tempa\@empty\else
3490 \bbl@infowarn{The following fonts are not babel standard families:\\
3491 \bbl@tempa
3492 There is nothing intrinsically wrong with it, but\\
3493 'babel' will no set Script and Language. Consider\\
3494 defining a new family with \string\babelfont.\\
3495 Reported}%
3496 \fi
3497 \endgroup}
3498 \fi
3499 \fi

```

Now the macros defining the font with `fontspec`.

When there are repeated keys in `fontspec`, the last value wins. So, we just place the ini settings at the beginning, and user settings will take precedence. We must deactivate temporarily `\bbl@mapselect` because `\selectfont` is called internally when a font is defined.

```

3500 \def\bbl@font@set#1#2#3{% eg \bbl@rmdflt@lang \rmdefault \rmfamily
3501 \bbl@xin@{<>}{#1}%
3502 \ifin@
3503 \bbl@exp{\\bbl@fontspec@set\\#1\expandafter@gobbletwo#1\\#3}%
3504 \fi
3505 \bbl@exp{%
3506 \def\\#2{#1}% eg, \rmdefault{\bbl@rmdflt@lang}
3507 \\bbl@ifsamestring{#2}{\f@family}{\\#3\let\\bbl@tempa\relax}}}%

```

```

3508%      TODO - next should be global?, but even local does its job. I'm
3509%      still not sure -- must investigate:
3510\def\bb1@fontspec@set#1#2#3#4{% eg \bb1@rmdflt@lang fnt-opt fnt-nme \xxfamily
3511  \let\bb1@tempe\bb1@mapselect
3512  \let\bb1@mapselect\relax
3513  \let\bb1@temp@fam#4%      eg, '\rmfamily', to be restored below
3514  \let#4\relax      % So that can be used with \newfontfamily
3515  \bb1@exp{%
3516    \let\\bb1@temp@pfam\<\bb1@stripslash#4\space>% eg, '\rmfamily '
3517    \<keys_if_exist:nnF>{fontspec-opentype}%
3518      {Script/\bb1@cs{sname@\language}}%
3519      {\newfontscript{\bb1@cs{sname@\language}}}%
3520      {\bb1@cs{sotf@\language}}}%
3521    \<keys_if_exist:nnF>{fontspec-opentype}%
3522      {Language/\bb1@cs{lname@\language}}%
3523      {\newfontlanguage{\bb1@cs{lname@\language}}}%
3524      {\bb1@cs{lotf@\language}}}%
3525    \newfontfamily\\#4%
3526    [\bb1@cs{lsys@\language},#2]{#3}% ie \bb1@exp{.}{#3}
3527  \begingroup
3528    #4%
3529    \xdef#1{\f@family}%      eg, \bb1@rmdflt@lang{FreeSerif(0)}
3530  \endgroup
3531  \let#4\bb1@temp@fam
3532  \bb1@exp{\let\<\bb1@stripslash#4\space>}\bb1@temp@pfam
3533  \let\bb1@mapselect\bb1@tempe}%

```

font@rst and famrst are only used when there is no global settings, to save and restore de previous families. Not really necessary, but done for optimization.

```

3534\def\bb1@font@rst#1#2#3#4{%
3535  \bb1@csarg\def{famrst@#4}{\bb1@font@set{#1}#2#3}}

```

The default font families. They are eurocentric, but the list can be expanded easily with \babelfont.

```

3536\def\bb1@font@fams{rm,sf,tt}

```

The old tentative way. Short and preverved for compatibility, but deprecated. Note there is no direct alternative for \babelFSfeatures. The reason in explained in the user guide, but essentially – that was not the way to go :-).

```

3537\newcommand\babelFSstore[2][{%
3538  \bb1@ifblank{#1}%
3539    {\bb1@csarg\def{sname@#2}{Latin}}%
3540    {\bb1@csarg\def{sname@#2}{#1}}%
3541  \bb1@provide@dirs{#2}%
3542  \bb1@csarg\ifnum{wdir@#2}>\z@
3543    \let\bb1@beforeforeign\leavevmode
3544    \EnableBabelHook{babel-bidi}%
3545  \fi
3546  \bb1@foreach{#2}{%
3547    \bb1@FSstore{##1}{rm}\rmdefault\bb1@save@rmdefault
3548    \bb1@FSstore{##1}{sf}\sfdefault\bb1@save@sfdefault
3549    \bb1@FSstore{##1}{tt}\ttdefault\bb1@save@ttdefault}}
3550\def\bb1@FSstore#1#2#3#4{%
3551  \bb1@csarg\edef{#2default#1}{#3}%
3552  \expandafter\addto\csname extras#1\endcsname{%
3553    \let#4#3%
3554    \ifx#3\f@family
3555      \edef#3{\csname \bb1@#2default#1\endcsname}%
3556      \fontfamily{#3}\selectfont

```



```

3557 \else
3558 \edef#3{\csname bbl@#2default#1\endcsname}%
3559 \fi}%
3560 \expandafter\addto\csname noextras#1\endcsname{%
3561 \ifx#3\fontfamily
3562 \fontfamily{#4}\selectfont
3563 \fi
3564 \let#3#4}}
3565 \let\bbl@langfeatures\@empty
3566 \def\babelFSfeatures{% make sure \fontspec is redefined once
3567 \let\bbl@ori@fontspec\fontspec
3568 \renewcommand\fontspec[1][{}]{%
3569 \bbl@ori@fontspec[\bbl@langfeatures##1]}
3570 \let\babelFSfeatures\bbl@FSfeatures
3571 \babelFSfeatures}
3572 \def\bbl@FSfeatures#1#2{%
3573 \expandafter\addto\csname extras#1\endcsname{%
3574 \babel@save\bbl@langfeatures
3575 \edef\bbl@langfeatures{#2,}}
3576 <</Font selection>>

```

## 14 Hooks for XeTeX and LuaTeX

### 14.1 XeTeX

Unfortunately, the current encoding cannot be retrieved and therefore it is reset always to `utf8`, which seems a sensible default.

LaTeX sets many “codes” just before loading `hyphen.cfg`. That is not a problem in `luatex`, but in `xetex` they must be reset to the proper value. Most of the work is done in `xe(la)tex.ini`, so here we just “undo” some of the changes done by LaTeX. Anyway, for consistency LuaTeX also resets the catcodes.

```

3577 <<{*Restore Unicode catcodes before loading patterns}>> ≡
3578 \begingroup
3579 % Reset chars "80-"C0 to category "other", no case mapping:
3580 \catcode\@=11 \count@=128
3581 \loop\ifnum\count@<192
3582 \global\uccode\count@=0 \global\lccode\count@=0
3583 \global\catcode\count@=12 \global\sffcode\count@=1000
3584 \advance\count@ by 1 \repeat
3585 % Other:
3586 \def\O ##1 {%
3587 \global\uccode"##1=0 \global\lccode"##1=0
3588 \global\catcode"##1=12 \global\sffcode"##1=1000 }%
3589 % Letter:
3590 \def\L ##1 ##2 ##3 {\global\catcode"##1=11
3591 \global\uccode"##1="##2
3592 \global\lccode"##1="##3
3593 % Uppercase letters have sffcode=999:
3594 \ifnum"##1="##3 \else \global\sffcode"##1=999 \fi }%
3595 % Letter without case mappings:
3596 \def\l ##1 {\L ##1 ##1 ##1 }%
3597 \l 00AA
3598 \L 00B5 039C 00B5
3599 \l 00BA
3600 \O 00D7
3601 \l 00DF
3602 \O 00F7

```

```

3603 \L 00FF 0178 00FF
3604 \endgroup
3605 \input #1\relax
3606 <</Restore Unicode catcodes before loading patterns>>

```

Some more common code.

```

3607 <<(*Footnote changes)>> ≡
3608 \bbl@trace{Bidi footnotes}
3609 \ifx\bbl@beforeforeign\leavevmode
3610 \def\bbl@footnote#1#2#3{%
3611 \ifnextchar[%
3612 {\bbl@footnote@o{#1}{#2}{#3}}%
3613 {\bbl@footnote@x{#1}{#2}{#3}}}
3614 \def\bbl@footnote@x#1#2#3#4{%
3615 \bgroup
3616 \select@language@x{\bbl@main@language}%
3617 \bbl@fn@footnote{#2#1{\ignorespaces#4}#3}%
3618 \egroup}
3619 \def\bbl@footnote@o#1#2#3[#4]#5{%
3620 \bgroup
3621 \select@language@x{\bbl@main@language}%
3622 \bbl@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}%
3623 \egroup}
3624 \def\bbl@footnotetext#1#2#3{%
3625 \ifnextchar[%
3626 {\bbl@footnotetext@o{#1}{#2}{#3}}%
3627 {\bbl@footnotetext@x{#1}{#2}{#3}}}
3628 \def\bbl@footnotetext@x#1#2#3#4{%
3629 \bgroup
3630 \select@language@x{\bbl@main@language}%
3631 \bbl@fn@footnotetext{#2#1{\ignorespaces#4}#3}%
3632 \egroup}
3633 \def\bbl@footnotetext@o#1#2#3[#4]#5{%
3634 \bgroup
3635 \select@language@x{\bbl@main@language}%
3636 \bbl@fn@footnotetext[#4]{#2#1{\ignorespaces#5}#3}%
3637 \egroup}
3638 \def\BabelFootnote#1#2#3#4{%
3639 \ifx\bbl@fn@footnote\@undefined
3640 \let\bbl@fn@footnote\footnote
3641 \fi
3642 \ifx\bbl@fn@footnotetext\@undefined
3643 \let\bbl@fn@footnotetext\footnotetext
3644 \fi
3645 \bbl@ifblank{#2}%
3646 {\def#1{\bbl@footnote{\@firstofone}{#3}{#4}}
3647 \@namedef{\bbl@stripslash#1text}%
3648 {\bbl@footnotetext{\@firstofone}{#3}{#4}}}%
3649 {\def#1{\bbl@exp{\bbl@footnote{\bbl@foreignlanguage{#2}}}{#3}{#4}}%
3650 \@namedef{\bbl@stripslash#1text}%
3651 {\bbl@exp{\bbl@footnotetext{\bbl@foreignlanguage{#2}}}{#3}{#4}}}%
3652 \fi
3653 <</Footnote changes>>

```

Now, the code.

```

3654 (*xetex)
3655 \def\BabelStringsDefault{unicode}
3656 \let\xebbl@stop\relax
3657 \AddBabelHook{xetex}{encodedcommands}{%

```

```

3658 \def\bb1@tempa{#1}%
3659 \ifx\bb1@tempa\@empty
3660 \XeTeXinputencoding"bytes"%
3661 \else
3662 \XeTeXinputencoding"#1"%
3663 \fi
3664 \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
3665 \AddBabelHook{xetex}{stopcommands}{%
3666 \xebbl@stop
3667 \let\xebbl@stop\relax}
3668 \def\bb1@intraspace#1 #2 #3\@{
3669 \bb1@csarg\gdef{\xeisp@bb1@cs{sbc@}\languagename}}%
3670 {\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}}
3671 \def\bb1@intrapenalty#1\@{
3672 \bb1@csarg\gdef{\xeipn@bb1@cs{sbc@}\languagename}}%
3673 {\XeTeXlinebreakpenalty #1\relax}}
3674 \def\bb1@provide@intraspace{%
3675 \bb1@xin@{\bb1@cs{sbc@}\languagename}}{Thai,Lao,Khmr}%
3676 \ifin@ % sea (currently ckj not handled)
3677 \bb1@ifunset{\bb1@intsp@}\languagename}}%
3678 {\expandafter\ifx\csname \bb1@intsp@\languagename\endcsname\@empty\else
3679 \ifx\bb1@KVP@intraspace\@nil
3680 \bb1@exp{%
3681 \bb1@intraspace\bb1@cs{intsp@\languagename}\@}%
3682 \fi
3683 \ifx\bb1@KVP@intrapenalty\@nil
3684 \bb1@intrapenalty0\@
3685 \fi
3686 \fi
3687 \ifx\bb1@KVP@intraspace\@nil\else % We may override the ini
3688 \expandafter\bb1@intraspace\bb1@KVP@intraspace\@
3689 \fi
3690 \ifx\bb1@KVP@intrapenalty\@nil\else
3691 \expandafter\bb1@intrapenalty\bb1@KVP@intrapenalty\@
3692 \fi
3693 \ifx\bb1@ispace@size\@undefined
3694 \AtBeginDocument{%
3695 \expandafter\bb1@add
3696 \csname selectfont\endcsname{\bb1@ispace@size}}%
3697 \def\bb1@ispace@size{\bb1@cs{\xeisp@bb1@cs{sbc@}\languagename}}%
3698 \fi}%
3699 \fi}
3700 \AddBabelHook{xetex}{loadkernel}{%
3701 <<Restore Unicode catcodes before loading patterns>>}
3702 \ifx\DisableBabelHook\@undefined\endinput\fi
3703 \AddBabelHook{babel-fontspec}{afterextras}{\bb1@switchfont}
3704 \AddBabelHook{babel-fontspec}{beforestart}{\bb1@ckeckstdfonts}
3705 \DisableBabelHook{babel-fontspec}
3706 <<Font selection>>
3707 \input txtbabel.def
3708 </xetex>

```

## 14.2 Layout

*In progress.*

Note elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titlesp, and geometry.

\bb1@startskip and \bb1@endskip are available to package authors. Thanks to the T<sub>E</sub>X

expansion mechanism the following constructs are valid: `\adim\bbbl@startskip`,  
`\advance\bbbl@startskip\adim`, `\bbbl@startskip\adim`.  
Consider `txtbabel` as a shorthand for *tex-xet babel*, which is the bidi model in both `pdfTeX`  
and `xetex`.

```

3709 (*texxet)
3710 \providecommand\bbbl@provide@intraspace{}
3711 \bbbl@trace{Redefinitions for bidi layout}
3712 \def\bbbl@sspre@caption{%
3713   \bbbl@exp{\everyhbox{\bbbl@textdir\bbbl@cs{wdir@\bbbl@main@language}}}}
3714 \ifx\bbbl@opt@layout\@nnil\endinput\fi % No layout
3715 \def\bbbl@startskip{\ifcase\bbbl@thepardir\leftskip\else\rightskip\fi}
3716 \def\bbbl@endskip{\ifcase\bbbl@thepardir\rightskip\else\leftskip\fi}
3717 \ifx\bbbl@beforeforeign\leavevmode % A poor test for bidi=
3718   \def\hangfrom#1{%
3719     \setbox\@tempboxa\hbox{#1}%
3720     \hangindent\ifcase\bbbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
3721     \noindent\box\@tempboxa}
3722 \def\raggedright{%
3723   \let\@centercr
3724   \bbbl@startskip\z@skip
3725   \@rightskip\@flushglue
3726   \bbbl@endskip\@rightskip
3727   \parindent\z@
3728   \parfillskip\bbbl@startskip}
3729 \def\raggedleft{%
3730   \let\@centercr
3731   \bbbl@startskip\@flushglue
3732   \bbbl@endskip\z@skip
3733   \parindent\z@
3734   \parfillskip\bbbl@endskip}
3735 \fi
3736 \IfBabelLayout{lists}
3737   {\bbbl@sreplace\list
3738     {\@totalleftmargin\leftmargin}{\@totalleftmargin\bbbl@listleftmargin}%
3739     \def\bbbl@listleftmargin{%
3740       \ifcase\bbbl@thepardir\leftmargin\else\rightmargin\fi}%
3741     \ifcase\bbbl@engine
3742       \def\labelenumii{}\theenumii{}\pdfTeX doesn't reverse ()
3743       \def\p@enumiii{\p@enumii}\theenumii{}
3744     \fi
3745     \bbbl@sreplace\@verbatim
3746       {\leftskip\@totalleftmargin}%
3747       {\bbbl@startskip\textwidth
3748         \advance\bbbl@startskip-\linewidth}%
3749     \bbbl@sreplace\@verbatim
3750       {\rightskip\z@skip}%
3751       {\bbbl@endskip\z@skip}}%
3752   {}
3753 \IfBabelLayout{contents}
3754   {\bbbl@sreplace\@dottedtocline{\leftskip}{\bbbl@startskip}%
3755     \bbbl@sreplace\@dottedtocline{\rightskip}{\bbbl@endskip}}
3756   {}
3757 \IfBabelLayout{columns}
3758   {\bbbl@sreplace\@outputdblcol{\hb@xt@\textwidth}{\bbbl@outputbox}%
3759     \def\bbbl@outputbox#1{%
3760       \hb@xt@\textwidth{%
3761         \hskip\columnwidth
3762         \hfil

```

```

3763      {\normalcolor\vrule \@width\columnseprule}%
3764      \hfil
3765      \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
3766      \hskip-\textwidth
3767      \hb@xt@\columnwidth{\box\@outputbox \hss}%
3768      \hskip\columnsep
3769      \hskip\columnwidth}}}%
3770  {}
3771  <<Footnote changes>>
3772  \IfBabelLayout{footnotes}%
3773  {\BabelFootnote\footnote\language\language{}{}}%
3774  \BabelFootnote\localfootnote\language\language{}{}}%
3775  \BabelFootnote\mainfootnote{}{}}{}
3776  {}

```

Implicitly reverses sectioning labels in bidi=basic, because the full stop is not in contact with L numbers any more. I think there must be a better way.

```

3777 \IfBabelLayout{counters}%
3778  {\let\bbbl@latinarabic=@arabic
3779   \def\@arabic#1{\babelsublr{\bbbl@latinarabic#1}}}%
3780   \let\bbbl@asciroman=@roman
3781   \def\@roman#1{\babelsublr{\ensureascii{\bbbl@asciroman#1}}}%
3782   \let\bbbl@asciiRoman=@Roman
3783   \def\@Roman#1{\babelsublr{\ensureascii{\bbbl@asciiRoman#1}}}}{}
3784 </texxet>

```

## 14.3 LuaTeX

The new loader for luatex is based solely on `language.dat`, which is read on the fly. The code shouldn't be executed when the format is build, so we check if `\AddBabelHook` is defined. Then comes a modified version of the loader in `hyphen.cfg` (without the `hyphenmins` stuff, which is under the direct control of `babel`).

The names `\l@<language>` are defined and take some value from the beginning because all `ldf` files assume this for the corresponding language to be considered valid, but patterns are not loaded (except the first one). This is done later, when the language is first selected (which usually means when the `ldf` finishes). If a language has been loaded, `\bbbl@hyphendata@<num>` exists (with the names of the files read).

The default setup preloads the first language into the format. This is intended mainly for 'english', so that it's available without further intervention from the user. To avoid duplicating it, the following rule applies: if the "0th" language and the first language in `language.dat` have the same name then just ignore the latter. If there are new synonymous, they are added, but note if the language patterns have not been preloaded they won't at run time.

Other preloaded languages could be read twice, if they have been preloaded into the format. This is not optimal, but it shouldn't happen very often – with luatex patterns are best loaded when the document is typeset, and the "0th" language is preloaded just for backwards compatibility.

As of 1.1b, lua(e)tex is taken into account. Formerly, loading of patterns on the fly didn't work in this format, but with the new loader it does. Unfortunately, the format is not based on `babel`, and data could be duplicated, because languages are reassigned above those in the format (nothing serious, anyway). Note even with this format `language.dat` is used (under the principle of a single source), instead of `language.def`.

Of course, there is room for improvements, like tools to read and reassign languages, which would require modifying the language list, and better error handling.

We need catcode tables, but no format (targeted by `babel`) provide a command to allocate them (although there are packages like `ctablestack`). For the moment, a dangerous

approach is used – just allocate a high random number and cross the fingers. To complicate things, etex.sty changes the way languages are allocated.

```

3785 (*luatex)
3786 \ifx\AddBabelHook\@undefined
3787 \bbl@trace{Read language.dat}
3788 \begingroup
3789 \toks@{}
3790 \count@ \z@ % 0=start, 1=0th, 2=normal
3791 \def\bbl@process@line#1#2 #3 #4 {%
3792   \ifx=#1%
3793     \bbl@process@synonym{#2}%
3794   \else
3795     \bbl@process@language{#1#2}{#3}{#4}%
3796   \fi
3797   \ignorespaces}
3798 \def\bbl@manylang{%
3799   \ifnum\bbl@last>\@ne
3800     \bbl@info{Non-standard hyphenation setup}%
3801   \fi
3802   \let\bbl@manylang\relax}
3803 \def\bbl@process@language#1#2#3{%
3804   \ifcase\count@
3805     \@ifundefined{zth#1}{\count@\tw@}{\count@\@ne}%
3806   \or
3807     \count@\tw@
3808   \fi
3809   \ifnum\count@=\tw@
3810     \expandafter\addlanguage\csname l@#1\endcsname
3811     \language\allocationnumber
3812     \chardef\bbl@last\allocationnumber
3813     \bbl@manylang
3814     \let\bbl@elt\relax
3815     \xdef\bbl@languages{%
3816       \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
3817   \fi
3818   \the\toks@
3819   \toks@{}}
3820 \def\bbl@process@synonym@aux#1#2{%
3821   \global\expandafter\chardef\csname l@#1\endcsname#2\relax
3822   \let\bbl@elt\relax
3823   \xdef\bbl@languages{%
3824     \bbl@languages\bbl@elt{#1}{#2}{}}}%
3825 \def\bbl@process@synonym#1{%
3826   \ifcase\count@
3827     \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
3828   \or
3829     \@ifundefined{zth#1}{\bbl@process@synonym@aux{#1}{0}}{%
3830   \else
3831     \bbl@process@synonym@aux{#1}{\the\bbl@last}%
3832   \fi}
3833 \ifx\bbl@languages\@undefined % Just a (sensible?) guess
3834   \chardef\l@english\z@
3835   \chardef\l@USenglish\z@
3836   \chardef\bbl@last\z@
3837   \global\namedef\bbl@hyphendata@0{{hyphen.tex}}
3838   \gdef\bbl@languages{%
3839     \bbl@elt{english}{0}{hyphen.tex}}%
3840     \bbl@elt{USenglish}{0}{}}

```

```

3841 \else
3842 \global\let\bbl@languages@format\bbl@languages
3843 \def\bbl@elt#1#2#3#4{% Remove all except language 0
3844 \ifnum#2>\z@\else
3845 \noexpand\bbl@elt{#1}{#2}{#3}{#4}%
3846 \fi}%
3847 \xdef\bbl@languages{\bbl@languages}%
3848 \fi
3849 \def\bbl@elt#1#2#3#4{\@namedef{zth@#1}{}} % Define flags
3850 \bbl@languages
3851 \openin1=language.dat
3852 \ifeof1
3853 \bbl@warning{I couldn't find language.dat. No additional\\%
3854 patterns loaded. Reported}%
3855 \else
3856 \loop
3857 \endlinechar\m@ne
3858 \read1 to \bbl@line
3859 \endlinechar`\^^M
3860 \if T\ifeof1F\fi T\relax
3861 \ifx\bbl@line\@empty\else
3862 \edef\bbl@line{\bbl@line\space\space\space}%
3863 \expandafter\bbl@process@line\bbl@line\relax
3864 \fi
3865 \repeat
3866 \fi
3867 \endgroup
3868 \bbl@trace{Macros for reading patterns files}
3869 \def\bbl@get@enc#1:#2:#3\@@@{\def\bbl@hyph@enc{#2}}
3870 \ifx\babelcatcodetablenum\undefined
3871 \def\babelcatcodetablenum{5211}
3872 \fi
3873 \def\bbl@luapatterns#1#2{%
3874 \bbl@get@enc#1::\@@@
3875 \setbox\z@\hbox\bgroup
3876 \begingroup
3877 \ifx\catcodetable\undefined
3878 \let\savecatcodetable\luatexsavecatcodetable
3879 \let\initcatcodetable\luatexinitcatcodetable
3880 \let\catcodetable\luatexcatcodetable
3881 \fi
3882 \savecatcodetable\babelcatcodetablenum\relax
3883 \initcatcodetable\numexpr\babelcatcodetablenum+1\relax
3884 \catcodetable\numexpr\babelcatcodetablenum+1\relax
3885 \catcode`\#=6 \catcode`\$=3 \catcode`\&=4 \catcode`\^=7
3886 \catcode`\_ =8 \catcode`\{=1 \catcode`\}=2 \catcode`\~ =13
3887 \catcode`\@=11 \catcode`\^^I=10 \catcode`\^^J=12
3888 \catcode`\<=12 \catcode`\>=12 \catcode`\*=12 \catcode`\.=12
3889 \catcode`\-=12 \catcode`\/=12 \catcode`\[=12 \catcode`\]=12
3890 \catcode`\`=12 \catcode`\'=12 \catcode`\\"=12
3891 \input #1\relax
3892 \catcodetable\babelcatcodetablenum\relax
3893 \endgroup
3894 \def\bbl@tempa{#2}%
3895 \ifx\bbl@tempa\@empty\else
3896 \input #2\relax
3897 \fi
3898 \egroup}%
3899 \def\bbl@patterns@lua#1{%

```

```

3900 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
3901   \csname l@#1\endcsname
3902   \edef\bbl@tempa{#1}%
3903   \else
3904     \csname l@#1:\f@encoding\endcsname
3905     \edef\bbl@tempa{#1:\f@encoding}%
3906   \fi\relax
3907   \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
3908   \@ifundefined{bbl@hyphendata@the\language}%
3909     {\def\bbl@elt##1##2##3##4{%
3910       \ifnum##2=\csname l@bbl@tempa\endcsname % #2=spanish, dutch:OT1...
3911       \def\bbl@tempb{##3}%
3912       \ifx\bbl@tempb\@empty\else % if not a synonymous
3913         \def\bbl@tempc{##3}{##4}}%
3914       \fi
3915       \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
3916     \fi}%
3917   \bbl@languages
3918   \@ifundefined{bbl@hyphendata@the\language}%
3919     {\bbl@info{No hyphenation patterns were set for\%
3920       language '\bbl@tempa'. Reported}}%
3921     {\expandafter\expandafter\expandafter\bbl@luapatterns
3922       \csname bbl@hyphendata@the\language\endcsname}}}%
3923 \endinput\fi
3924 \begingroup
3925 \catcode`\%=12
3926 \catcode`\'=12
3927 \catcode`\%=12
3928 \catcode`\:=12
3929 \directlua{
3930   Babel = Babel or {}
3931   function Babel.bytes(line)
3932     return line:gsub("(.)",
3933       function (chr) return unicode.utf8.char(string.byte(chr)) end)
3934   end
3935   function Babel.begin_process_input()
3936     if luatexbase and luatexbase.add_to_callback then
3937       luatexbase.add_to_callback('process_input_buffer',
3938         Babel.bytes, 'Babel.bytes')
3939     else
3940       Babel.callback = callback.find('process_input_buffer')
3941       callback.register('process_input_buffer', Babel.bytes)
3942     end
3943   end
3944   function Babel.end_process_input ()
3945     if luatexbase and luatexbase.remove_from_callback then
3946       luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
3947     else
3948       callback.register('process_input_buffer', Babel.callback)
3949     end
3950   end
3951   function Babel.addpatterns(pp, lg)
3952     local lg = lang.new(lg)
3953     local pats = lang.patterns(lg) or ''
3954     lang.clear_patterns(lg)
3955     for p in pp:gmatch('[^%s]+') do
3956       ss = ''
3957       for i in string.utfcharacters(p:gsub('%d', '')) do
3958         ss = ss .. '%d?' .. i

```



```

3959     end
3960     ss = ss:gsub('^%%d%?%', '%%.') .. '%d?'
3961     ss = ss:gsub('%.%%d%?$', '%%.')
3962     pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
3963     if n == 0 then
3964         tex.sprint(
3965             [[\string\csname\space bbl@info\endcsname{New pattern: }]
3966             .. p .. [{}]])
3967         pats = pats .. ' ' .. p
3968     else
3969         tex.sprint(
3970             [[\string\csname\space bbl@info\endcsname{Renew pattern: }]
3971             .. p .. [{}]])
3972     end
3973 end
3974 lang.patterns(lg, pats)
3975 end
3976 }
3977 \endgroup
3978 \ifx\newattribute\undefined\else
3979   \newattribute\bbl@attr@locale
3980   \AddBabelHook{luatex}{beforeextras}{%
3981     \setattribute\bbl@attr@locale\localeid}
3982 \fi
3983 \def\BabelStringsDefault{unicode}
3984 \let\luabbl@stop\relax
3985 \AddBabelHook{luatex}{encodedcommands}{%
3986   \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
3987   \ifx\bbl@tempa\bbl@tempb\else
3988     \directlua{Babel.begin_process_input()}%
3989     \def\luabbl@stop{%
3990       \directlua{Babel.end_process_input()}}%
3991   \fi}%
3992 \AddBabelHook{luatex}{stopcommands}{%
3993   \luabbl@stop
3994   \let\luabbl@stop\relax}
3995 \AddBabelHook{luatex}{patterns}{%
3996   \@ifundefined{bbl@hyphendata@the\language}%
3997     {\def\bbl@elt##1##2##3##4{%
3998       \ifnum##2=\csname l@#2\endcsname % #2=spanish, dutch:OT1...
3999       \def\bbl@tempb{##3}%
4000       \ifx\bbl@tempb\empty\else % if not a synonymous
4001         \def\bbl@tempc{##3}{##4}}%
4002       \fi
4003       \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4004     \fi}%
4005   \bbl@languages
4006   \@ifundefined{bbl@hyphendata@the\language}%
4007     {\bbl@info{No hyphenation patterns were set for\%
4008       language '#2'. Reported}}%
4009     {\expandafter\expandafter\expandafter\bbl@luapatterns
4010       \csname bbl@hyphendata@the\language\endcsname}}}%
4011   \@ifundefined{bbl@patterns@}{}%
4012   \begingroup
4013     \bbl@xin@{,\number\language,}{,\bbl@pttnlist}%
4014     \ifin@else
4015       \ifx\bbl@patterns@\empty\else
4016         \directlua{ Babel.addpatterns(
4017           [[\bbl@patterns@]], \number\language) }%

```

```

4018     \fi
4019     \@ifundefined{bbl@patterns@#1}%
4020     \empty
4021     {\directlua{ Babel.addpatterns(
4022       [[\space\csname bbl@patterns@#1\endcsname]],
4023       \number\language) }}%
4024     \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
4025     \fi
4026   \endgroup}}
4027 \AddBabelHook{luatex}{everylanguage}{%
4028   \def\process@language##1##2##3{%
4029     \def\process@line####1####2 ####3 ####4 {}}}
4030 \AddBabelHook{luatex}{loadpatterns}{%
4031   \input #1\relax
4032   \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
4033     {#{1}}{}}
4034 \AddBabelHook{luatex}{loadexceptions}{%
4035   \input #1\relax
4036   \def\bbl@tempb##1##2{#{##1}-#{##2}}%
4037   \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
4038     {\expandafter\expandafter\expandafter\bbl@tempb
4039       \csname bbl@hyphendata@the\language\endcsname}}

```

`\babelpatterns` This macro adds patterns. Two macros are used to store them: `\bbl@patterns@` for the global ones and `\bbl@patterns@<lang>` for language ones. We make sure there is a space between words when multiple commands are used.

```

4040 \@onlypreamble\babelpatterns
4041 \AtEndOfPackage{%
4042   \newcommand\babelpatterns[2][\empty]{%
4043     \ifx\bbl@patterns@relax
4044       \let\bbl@patterns@empty
4045     \fi
4046     \ifx\bbl@pttnlistempty\else
4047       \bbl@warning{%
4048         You must not intermingle \string\selectlanguage\space and\%
4049         \string\babelpatterns\space or some patterns will not\%
4050         be taken into account. Reported}%
4051     \fi
4052     \ifxempty#1%
4053       \protected@edef\bbl@patterns@{\bbl@patterns@space#2}%
4054     \else
4055       \edef\bbl@tempb{\zap@space#1 \empty}%
4056       \bbl@for\bbl@tempa\bbl@tempb{%
4057         \bbl@fixname\bbl@tempa
4058         \bbl@iflanguage\bbl@tempa{%
4059           \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
4060             \@ifundefined{bbl@patterns@\bbl@tempa}%
4061             \empty
4062             {\csname bbl@patterns@\bbl@tempa\endcsname\space}%
4063             #2}}}%
4064     \fi}}

```

## 14.4 Southeast Asian scripts

*In progress.* Replace regular (ie, implicit) discretionaries by spaceskips, based on the previous glyph (which I think makes sense, because the hyphen and the previous char go always together). Other discretionaries are not touched.

For the moment, only 3 SA languages are activated by default (see Unicode UAX 14).

```

4065 \directlua{
4066   Babel = Babel or {}
4067   Babel.linebreaking = Babel.linebreaking or {}
4068   Babel.linebreaking.before = {}
4069   Babel.linebreaking.after = {}
4070   Babel.locale = {} % Free to use, indexed with \localeid
4071   function Babel.linebreaking.add_before(func)
4072     tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
4073     table.insert(Babel.linebreaking.before, func)
4074   end
4075   function Babel.linebreaking.add_after(func)
4076     tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
4077     table.insert(Babel.linebreaking.after, func)
4078   end
4079 }
4080 \def\bbl@intraspace#1 #2 #3\@{#%
4081   \directlua{
4082     Babel = Babel or {}
4083     Babel.intraspaces = Babel.intraspaces or {}
4084     Babel.intraspaces['\csname bbl@sbc@language\endcsname'] = %
4085       {b = #1, p = #2, m = #3}
4086     Babel.locale_props[\the\localeid].intraspace = %
4087       {b = #1, p = #2, m = #3}
4088   }}
4089 \def\bbl@intrapenalty#1\@{#%
4090   \directlua{
4091     Babel = Babel or {}
4092     Babel.intrapenalties = Babel.intrapenalties or {}
4093     Babel.intrapenalties['\csname bbl@sbc@language\endcsname'] = #1
4094     Babel.locale_props[\the\localeid].intrapenalty = #1
4095   }}
4096 \begingroup
4097 \catcode`\%=12
4098 \catcode`\^=14
4099 \catcode`\'=12
4100 \catcode`\~=12
4101 \gdef\bbl@seaintraspace{^
4102   \let\bbl@seaintraspace\relax
4103   \directlua{
4104     Babel = Babel or {}
4105     Babel.sea_enabled = true
4106     Babel.sea_ranges = Babel.sea_ranges or {}
4107     function Babel.set_chranges (script, chrng)
4108       local c = 0
4109       for s, e in string.gmatch(chrng..' ', '(.-%).%.(.-%)s') do
4110         Babel.sea_ranges[script..c]={tonumber(s,16), tonumber(e,16)}
4111         c = c + 1
4112       end
4113     end
4114     function Babel.sea_disc_to_space (head)
4115       local sea_ranges = Babel.sea_ranges
4116       local last_char = nil
4117       local quad = 655360 ^^ 10 pt = 655360 = 10 * 65536
4118       for item in node.traverse(head) do
4119         local i = item.id
4120         if i == node.id'glyph' then
4121           last_char = item

```

```

4122     elseif i == 7 and item.subtype == 3 and last_char
4123         and last_char.char > 0x0C99 then
4124         quad = font.getfont(last_char.font).size
4125         for lg, rg in pairs(sea_ranges) do
4126             if last_char.char > rg[1] and last_char.char < rg[2] then
4127                 lg = lg:sub(1, 4)
4128                 local intraspace = Babel.intraspaces[lg]
4129                 local intrapenalty = Babel.intrapenalties[lg]
4130                 local n
4131                 if intrapenalty ~= 0 then
4132                     n = node.new(14, 0)    ^^ penalty
4133                     n.penalty = intrapenalty
4134                     node.insert_before(head, item, n)
4135                 end
4136                 n = node.new(12, 13)    ^^ (glue, spaceskip)
4137                 node.setglue(n, intraspace.b * quad,
4138                     intraspace.p * quad,
4139                     intraspace.m * quad)
4140                 node.insert_before(head, item, n)
4141                 node.remove(head, item)
4142             end
4143         end
4144     end
4145 end
4146 end
4147 }^^
4148 \bbl@luahyphenate}
4149 \catcode`\%=14
4150 \gdef\bbl@cjkintraspaces{%
4151     \let\bbl@cjkintraspaces\relax
4152     \directlua{
4153         Babel = Babel or {}
4154         require'babel-data-cjk.lua'
4155         Babel.cjk_enabled = true
4156         function Babel.cjk_linebreak(head)
4157             local GLYPH = node.id'glyph'
4158             local last_char = nil
4159             local quad = 655360    % 10 pt = 655360 = 10 * 65536
4160             local last_class = nil
4161             local last_lang = nil
4162
4163             for item in node.traverse(head) do
4164                 if item.id == GLYPH then
4165
4166                     local lang = item.lang
4167
4168                     local LOCALE = node.get_attribute(item,
4169                         luatexbase.registernumber'bbl@attr@locale')
4170                     local props = Babel.locale_props[LOCALE]
4171
4172                     class = Babel.cjk_class[item.char].c
4173
4174                     if class == 'cp' then class = 'cl' end % )] as CL
4175                     if class == 'id' then class = 'I' end
4176
4177                     if class and last_class and Babel.cjk_breaks[last_class][class] then
4178                         br = Babel.cjk_breaks[last_class][class]
4179                     else
4180                         br = 0

```

```

4181         end
4182
4183         if br == 1 and props.linebreak == 'c' and
4184             lang ~= \the\l@nohyphenation\space and
4185             last_lang ~= \the\l@nohyphenation then
4186             local intrapenalty = props.intrapenalty
4187             if intrapenalty ~= 0 then
4188                 local n = node.new(14, 0)    % penalty
4189                 n.penalty = intrapenalty
4190                 node.insert_before(head, item, n)
4191             end
4192             local intraspace = props.intraspace
4193             local n = node.new(12, 13)    % (glue, spaceskip)
4194             node.setglue(n, intraspace.b * quad,
4195                 intraspace.p * quad,
4196                 intraspace.m * quad)
4197             node.insert_before(head, item, n)
4198         end
4199
4200         quad = font.getfont(item.font).size
4201         last_class = class
4202         last_lang = lang
4203     else % if penalty, glue or anything else
4204         last_class = nil
4205     end
4206 end
4207 lang.hyphenate(head)
4208 end
4209 }%
4210 \bbl@luahyphenate}
4211 \gdef\bbl@luahyphenate{%
4212     \let\bbl@luahyphenate\relax
4213     \directlua{
4214         luatexbase.add_to_callback('hyphenate',
4215             function (head, tail)
4216                 if Babel.cjk_enabled then
4217                     Babel.cjk_linebreak(head)
4218                 end
4219                 for k, func in ipairs(Babel.linebreaking.before) do
4220                     func(head)
4221                 end
4222                 lang.hyphenate(head)
4223                 for k, func in ipairs(Babel.linebreaking.after) do
4224                     func(head)
4225                 end
4226                 if Babel.sea_enabled then
4227                     Babel.sea_disc_to_space(head)
4228                 end
4229             end,
4230             'Babel.hyphenate')
4231     }
4232 }
4233 \endgroup
4234 \def\bbl@provide@intraspace{%
4235     \bbl@ifunset{\bbl@intsp@languagename}{}%
4236     {\expandafter\ifx\csname bbl@intsp@languagename\endcsname\@empty\else
4237         \bbl@xin@{\bbl@cs{sbc}p@languagename}}{Hant,Hans,Jpan,Kore,Kana}%
4238         \ifin@           % cjk
4239         \bbl@cjk@intraspace

```

```

4240 \directlua{
4241     Babel = Babel or {}
4242     Babel.locale_props = Babel.locale_props or {}
4243     Babel.locale_props[\the\localeid].linebreak = 'c'
4244 }%
4245 \bbl@exp{\bbl@intraspace\bbl@cs{intsp@\languagename}\@}%
4246 \ifx\bbl@KVP@intrapenalty\@nil
4247     \bbl@intrapenalty0\@
4248 \fi
4249 \else % sea
4250     \bbl@seaintraspace
4251     \bbl@exp{\bbl@intraspace\bbl@cs{intsp@\languagename}\@}%
4252 \directlua{
4253     Babel = Babel or {}
4254     Babel.sea_ranges = Babel.sea_ranges or {}
4255     Babel.set_chranges('\bbl@cs{sbcpr@\languagename}',
4256                       '\bbl@cs{chrng@\languagename}')
4257 }%
4258 \ifx\bbl@KVP@intrapenalty\@nil
4259     \bbl@intrapenalty0\@
4260 \fi
4261 \fi
4262 \fi
4263 \ifx\bbl@KVP@intrapenalty\@nil\else
4264     \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@
4265 \fi}}

```

## 14.5 CJK line breaking

Minimal line breaking for CJK scripts, mainly intended for simple documents and short texts as a secondary language. Only line breaking, with a little stretching for justification, without any attempt to adjust the spacing. It is based on (but does not strictly follow) the Unicode algorithm.

We first need a little table with the corresponding line breaking properties. A few characters have an additional key for the width (fullwidth vs. halfwidth), not yet used.

There is a separate file, defined below.

*Work in progress.*

Common stuff.

```

4266 \AddBabelHook{luatex}{loadkernel}{%
4267 <<Restore Unicode catcodes before loading patterns>>}
4268 \ifx\DisableBabelHook\undefined\endinput\fi
4269 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
4270 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ckeckstdfonts}
4271 \DisableBabelHook{babel-fontspec}
4272 <<Font selection>>

```

The code for `\babelcharproperty` is straightforward. Just note the modified lua table can be different.

```

4273 \newcommand\babelcharproperty[1]{%
4274     \count@=#1\relax
4275     \ifvmode
4276         \expandafter\bbl@chprop
4277     \else
4278         \bbl@error{\string\babelcharproperty\space can be used only in\%
4279                 vertical mode (preamble or between paragraphs)}%
4280         {See the manual for futher info}%
4281     \fi}

```

```

4282 \newcommand\bbl@chprop[3][\the\count@]{%
4283   \@tempcnta=#1\relax
4284   \bbl@ifunset{\bbl@chprop@#2}%
4285   {\bbl@error{No property named '#2'. Allowed values are\%
4286     direction (bc), mirror (bmg), and linebreak (lb)}}%
4287     {See the manual for futher info}}%
4288   }%
4289   \loop
4290     \@nameuse{\bbl@chprop@#2}{#3}%
4291     \ifnum\count@<\@tempcnta
4292       \advance\count@\@ne
4293     \repeat}
4294 \def\bbl@chprop@direction#1{%
4295   \directlua{
4296     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
4297     Babel.characters[\the\count@]['d'] = '#1'
4298   }}
4299 \let\bbl@chprop@bc\bbl@chprop@direction
4300 \def\bbl@chprop@mirror#1{%
4301   \directlua{
4302     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
4303     Babel.characters[\the\count@]['m'] = '\number#1'
4304   }}
4305 \let\bbl@chprop@bmg\bbl@chprop@mirror
4306 \def\bbl@chprop@linebreak#1{%
4307   \directlua{
4308     Babel.Babel.cjk_characters[\the\count@] = Babel.Babel.cjk_characters[\the\count@] or {}
4309     Babel.Babel.cjk_characters[\the\count@]['c'] = '#1'
4310   }}
4311 \let\bbl@chprop@lb\bbl@chprop@linebreak

```

Post-handling hyphenation patterns for non-standard rules, like ff to ff-f.

```

4312 \begingroup
4313 \catcode`#=11
4314 \directlua{
4315   function Babel.get_locale(n)
4316     return node.get_attribute(n, luatexbase.registernumber'bbl@attr@locale')
4317   end
4318
4319   function Babel.str_to_nodes(text, base)
4320     local n, head, last
4321     for s in string.utfvalues(text) do
4322       if base.id == 7 then
4323         base = base.replace
4324       end
4325       n = node.copy(base)
4326       n.char = s
4327       if not head then
4328         head = n
4329       else
4330         last.next = n
4331       end
4332       last = n
4333     end
4334     return head
4335   end
4336
4337   function Babel.fetch_word(head, funct)
4338     local word_string = ''

```

```

4339     local word_nodes = {}
4340     local locale, last_locale
4341
4342     for item in node.traverse(head) do
4343         locale = Babel.get_locale(item)
4344         last_locale = last_locale or locale
4345
4346         if item.id == 29 and not(item.char == 124) % ie, not |
4347             and locale == last_locale
4348             and not (item.lang == \the\l@nohyphenation) then
4349             word_string = word_string .. unicode.utf8.char(item.char)
4350             table.insert(word_nodes, item)
4351             last_locale = locale
4352
4353         elseif item.id == 7 and item.subtype == 2
4354             and locale == last_locale
4355             and not (item.lang == \the\l@nohyphenation) then
4356             word_string = word_string .. '-'
4357             table.insert(word_nodes, item)
4358             last_locale = locale
4359
4360         elseif item.id == 7 and item.subtype == 3
4361             and locale == last_locale
4362             and not (item.lang == \the\l@nohyphenation) then
4363             word_string = word_string .. '|'
4364             table.insert(word_nodes, item)
4365             last_locale = locale
4366
4367         elseif word_string == '' then
4368             % pass
4369
4370         else
4371             return word_string, word_nodes, item
4372         end
4373     end
4374 end
4375
4376 Babel.linebreaking.replacements = {}
4377
4378 function Babel.post_hyphenate_replace(head)
4379     local u = unicode.utf8
4380     local word_head = head
4381
4382     while true do
4383         local w, wn, nw = Babel.fetch_word(word_head)
4384         if not nw then return head end
4385
4386         if not Babel.linebreaking.replacements[Babel.get_locale(wn[1])] then
4387             break
4388         end
4389
4390         for _, reps in ipairs(Babel.linebreaking.replacements[Babel.get_locale(wn[1])]) do
4391             local r = reps.replace
4392             local p = reps.pattern
4393
4394             % This should be done when set:
4395             if not u.find(p, '()', nil, true) then
4396                 p = '()' .. p .. '()'
4397             end

```



```

4398     p = u.gsub(p, '{(.)}', % This should be done when set, too
4399         function (n)
4400             return '\@percentchar'
4401             .. (tonumber(n) and (tonumber(n)+1) or n)
4402         end)
4403
4404     while true do
4405         first, A, B, C, E, F, G = u.match(w, p)
4406         if not A then break end
4407
4408         local last = {A, B, C, E, F, G} % Ugly. Must be a better way
4409         last = last[#last]
4410
4411         % Fix offsets, from bytes to unicode
4412         first = u.len(w:sub(1, first-1)) + 1
4413         last = u.len(w:sub(1, last-1))
4414
4415         % This EXpands {n} in replacement fields pre, post, no
4416         local x = function (s)
4417             if s == nil then return '' end
4418             return u.gsub(s, '{([0-9])}',
4419                 function (n)
4420                     return u.sub(w, n+first-1, n+first-1)
4421                 end)
4422         end
4423
4424         local new % used when inserting and removing nodes
4425         local changed = 0
4426
4427         for q = first, last do
4428             local rep_i = r[q-first+1]
4429             local char_node = wn[q]
4430             local char_base = char_node
4431
4432             if rep_i and rep_i.data then
4433                 char_base = wn[rep_i.data]
4434             end
4435             if rep_i == nil then
4436                 rep_i = { string='' }
4437             end
4438
4439             if rep_i and (rep_i.pre or rep_i.no or rep_i.post) then
4440                 changed = changed + 1
4441                 d = node.new(7, 0) % (disc, discretionary)
4442                 d.pre = Babel.str_to_nodes(x(rep_i.pre), char_base)
4443                 d.post = Babel.str_to_nodes(x(rep_i.post), char_base)
4444                 d.replace = Babel.str_to_nodes(x(rep_i.no), char_base)
4445                 d.attr = char_base.attr
4446                 if rep_i.pre and not(rep_i.pre == '') then % TeXbook p96
4447                     d.penalty = rep_i.penalty or tex.hyphenpenalty
4448                 else
4449                     d.penalty = rep_i.penalty or tex.exhyphenpenalty
4450                 end
4451                 head, new = node.insert_before(head, char_node, d)
4452                 node.remove(head, char_node)
4453                 if q == 1 then
4454                     word_head = new
4455                 end
4456             elseif rep_i and rep_i.string then

```

```

4457         changed = changed + 1
4458         rep_i.string = x(rep_i.string, char_base)
4459         if rep_i.string == '' then
4460             if q == 1 then
4461                 word_head = char_node.next
4462             end
4463             head, new = node.remove(head, char_node)
4464         else
4465             local n
4466             for s in string.utfvalues(rep_i.string) do
4467                 if char_node.id == 7 then
4468                     texio.write_nl('Automatic hyphens cannot be replaced, just removed.')
4469                 else
4470                     n = node.copy(char_base)
4471                     end
4472                     n.char = s
4473                     if q == 1 then
4474                         head, new = node.insert_before(head, char_node, n)
4475                         word_head = new
4476                     else
4477                         node.insert_before(head, char_node, n)
4478                     end
4479                 end
4480             end
4481             node.remove(head, char_node)
4482         end
4483         end % if char and char.string
4484     end % for char in match
4485     if changed > 20 then
4486         texio.write('Too many changes. Ignoring the rest.')
4487     elseif changed > 0 then
4488         w, wn, nw = Babel.fetch_word(word_head)
4489     end
4490
4491     end % for match
4492 end % for patterns
4493 word_head = nw
4494 end % for words
4495 return head
4496 end
4497
4498 function Babel.linebreaking.add_replacement(lang, patt, repl)
4499     local lbk = Babel.linebreaking
4500     lbk.replacements[Babel.locale_ids[lang]] =
4501         lbk.replacements[Babel.locale_ids[lang]] or {}
4502     table.insert(lbk.replacements[Babel.locale_ids[lang]],
4503         { pattern = patt, replace = repl })
4504 end
4505
4506 }
4507 \endgroup
4508 %
4509 % \subsection{Layout}
4510 %
4511 % \textbf{Work in progress}.
4512 %
4513 % Unlike \xetex{}, \luatex{} requires only minimal changes for
4514 % right-to-left layouts, particularly in monolingual documents (the
4515 % engine itself reverses boxes -- including column order or headings

```

```

4516% --, margins, etc.) with |bidi=basic|, without having to patch
4517% almost any macro where text direction is relevant.
4518%
4519% |\@hangfrom| is useful in many contexts and it is redefined always
4520% with the |layout| option.
4521%
4522% There are, however, a number of issues when the text direction is not
4523% the same as the box direction (as set by |\bodydir|), and when
4524% |\parbox| and |\hangindent| are involved. Fortunately, latest
4525% releases of \luatex{} simplify a lot the solution with |\shapemode|.
4526%
4527% With the issue \#15 I realized commands are best patched, instead of
4528% redefined. With a few lines, a modification could be applied to
4529% several classes and packages. Now, |tabular| seems to work (at least
4530% in simple cases) with \textsf{array}, \textsf{tabularx},
4531% \textsf{hhline}, \textsf{colortbl}, \textsf{longtable},
4532% \textsf{booktabs}, etc. However, \textsf{dcolumn} still fails.
4533%
4534% \changes{babel-3.18}{2018/02/14}{\cs{bbl@nextfake}, similar to the old
4535% \cs{nextfakemath} in Omega}
4536% \changes{babel-3.19}{2018/04/23}{Patch \cs{@eqnnum}, somewhat ad hoc. To be
4537% improved.}
4538% \changes{babel-3.21}{2018/05/09}{The ad hoc \cs{@eqnnum} was buggy. Fixed.}
4539% \changes{babel-3.31}{2019/05/04}{\cs{@tabular} and \cs{list} are patched,
4540% instead of redefined.}
4541%
4542% \begin{macrocode}
4543 \bbl@trace{Redefinitions for bidi layout}
4544 \ifx\@eqnnum\undefined\else
4545 \ifx\bbl@attr@dir\undefined\else
4546 \edef\@eqnnum{%
4547 \unexpanded{\ifcase\bbl@attr@dir\else\bbl@textdir\@ne\fi}%
4548 \unexpanded\expandafter{\@eqnnum}}
4549 \fi
4550 \fi
4551 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
4552 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
4553 \def\bbl@nextfake#1{% non-local changes, use always inside a group!
4554 \bbl@exp{%
4555 \mathdir\the\bodydir
4556 #1% Once entered in math, set boxes to restore values
4557 \<ifmmode>%
4558 \everyvbox{%
4559 \the\everyvbox
4560 \bodydir\the\bodydir
4561 \mathdir\the\mathdir
4562 \everyhbox{\the\everyhbox}%
4563 \everyvbox{\the\everyvbox}}%
4564 \everyhbox{%
4565 \the\everyhbox
4566 \bodydir\the\bodydir
4567 \mathdir\the\mathdir
4568 \everyhbox{\the\everyhbox}%
4569 \everyvbox{\the\everyvbox}}%
4570 \<fi>}}%
4571 \def\@hangfrom#1{%
4572 \setbox\@tempboxa\hbox{#1}%
4573 \hangindent\wd\@tempboxa
4574 \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else

```

```

4575     \shapemode\@ne
4576     \fi
4577     \noindent\box\@tempboxa}
4578 \fi
4579 \IfBabelLayout{tabular}
4580   {\let\bbl@OL@tabular\@tabular
4581     \bbl@replace\@tabular{$}\bbl@nextfake$}%
4582     \let\bbl@NL@tabular\@tabular
4583     \AtBeginDocument{%
4584       \ifx\bbl@NL@tabular\@tabular\else
4585         \bbl@replace\@tabular{$}\bbl@nextfake$}%
4586         \let\bbl@NL@tabular\@tabular
4587       \fi}}
4588   {}
4589 \IfBabelLayout{lists}
4590   {\let\bbl@OL@list\list
4591     \bbl@sreplace\list{\parshape}\bbl@listparshape}%
4592     \let\bbl@NL@list\list
4593     \def\bbl@listparshape#1#2#3{%
4594       \parshape #1 #2 #3 %
4595       \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
4596         \shapemode\tw@
4597       \fi}}
4598   {}
4599 \IfBabelLayout{graphics}
4600   {\let\bbl@pictresetdir\relax
4601     \def\bbl@pictsetdir{%
4602       \ifcase\bbl@thetextdir
4603         \let\bbl@pictresetdir\relax
4604       \else
4605         \textdir TLT\relax
4606         \def\bbl@pictresetdir{\textdir TRT\relax}%
4607       \fi}%
4608     \let\bbl@OL@picture\@picture
4609     \let\bbl@OL@put\put
4610     \bbl@sreplace\@picture{\hskip-}\bbl@pictsetdir\hskip-}%
4611     \def\put(#1,#2)#3{% Not easy to patch. Better redefine.
4612       \@killglue
4613       \raise#2\unitlength
4614       \hb@xt\z@{\kern#1\unitlength\bbl@pictresetdir#3}\hss}}%
4615     \AtBeginDocument
4616       {\ifx\tikz@atbegin@node\@undefined\else
4617         \let\bbl@OL@pgfpicture\pgfpicture
4618         \bbl@sreplace\pgfpicture{\pgfpicturetrue}\bbl@pictsetdir\pgfpicturetrue}%
4619         \bbl@add\pgfsys@beginpicture{\bbl@pictsetdir}%
4620         \bbl@add\tikz@atbegin@node{\bbl@pictresetdir}%
4621       \fi}}
4622   {}

```

Implicitly reverses sectioning labels in bidi=basic-r, because the full stop is not in contact with L numbers any more. I think there must be a better way. Assumes bidi=basic, but there are some additional readjustments for bidi=default.

```

4623 \IfBabelLayout{counters}%
4624   {\let\bbl@OL@textsuperscript\textsuperscript
4625     \bbl@sreplace\textsuperscript{\m@th}\m@th\mathdir\pagedir}%
4626     \let\bbl@latinarabic=\@arabic
4627     \let\bbl@OL@arabic\@arabic
4628     \def\@arabic#1{\babelsublr\bbl@latinarabic#1}}%
4629     \ifpackagewith{babel}{bidi=default}%

```

```

4630 {\let\bbl@asciroman=\@roman
4631 \let\bbl@OL@@roman\@roman
4632 \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
4633 \let\bbl@asciiRoman=\@Roman
4634 \let\bbl@OL@@roman\@Roman
4635 \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}%
4636 \let\bbl@OL@labelenumii\labelenumii
4637 \def\labelenumii{}\theenumii}%
4638 \let\bbl@OL@p@enumiii\p@enumiii
4639 \def\p@enumiii{\p@enumii)\theenumii{}}{}{}
4640 <<Footnote changes>>
4641 \IfBabelLayout{footnotes}%
4642 {\let\bbl@OL@footnote\footnote
4643 \BabelFootnote\footnote\language{}{}}%
4644 \BabelFootnote\localfootnote\language{}{}}%
4645 \BabelFootnote\mainfootnote{}}{}{}
4646 {}

```

Some  $\TeX$  macros use internally the math mode for text formatting. They have very little in common and are grouped here, as a single option.

```

4647 \IfBabelLayout{extras}%
4648 {\let\bbl@OL@underline\underline
4649 \bbl@sreplace\underline{$\@@underline}{\bbl@nextfake$\@@underline}%
4650 \let\bbl@OL@LaTeX2e\LaTeX2e
4651 \DeclareRobustCommand{\LaTeXe}{\mbox{\m@th
4652 \if b\expandafter\@car\f@series\@nil\boldmath\fi
4653 \babelsublr}%
4654 \LaTeX\kern.15em2\bbl@nextfake$_{\textstyle\varepsilon}$}}}%
4655 {}
4656 </luatex>

```

## 14.6 Auto bidi with basic and basic-r

The file `babel-data-bidi.lua` currently only contains data. It is a large and boring file and it is not shown here (see the generated file), but here is a sample:

```

┐[0x25]={d='et'},
┐[0x26]={d='on'},
┐[0x27]={d='on'},
┐[0x28]={d='on',┐m=0x29},
┐[0x29]={d='on',┐m=0x28},
┐[0x2A]={d='on'},
┐[0x2B]={d='es'},
┐[0x2C]={d='cs'},

```

For the meaning of these codes, see the Unicode standard.

Now the `basic-r` bidi mode. One of the aims is to implement a fast and simple bidi algorithm, with a single loop. I managed to do it for R texts, with a second smaller loop for a special case. The code is still somewhat chaotic, but its behavior is essentially correct. I cannot resist copying the following text from `Emacs bidi.c` (which also attempts to implement the bidi algorithm with a single loop):

Arrrrgh!! The UAX#9 algorithm is too deeply entrenched in the assumption of batch-style processing [...]. May the fleas of a thousand camels infest the armpits of those who design supposedly general-purpose algorithms by looking at their own implementations, and fail to consider other possible implementations!

Well, it took me some time to guess what the batch rules in UAX#9 actually mean (in other word, *what* they do and *why*, and not only *how*), but I think (or I hope) I've managed to understand them.

In some sense, there are two bidi modes, one for numbers, and the other for text.

Furthermore, setting just the direction in R text is not enough, because there are actually *two* R modes (set explicitly in Unicode with RLM and ALM). In babel the dir is set by a higher protocol based on the language/script, which in turn sets the correct dir (<l>, <r> or <al>).

From UAX#9: "Where available, markup should be used instead of the explicit formatting characters". So, this simple version just ignores formatting characters. Actually, most of that annex is devoted to how to handle them.

BD14-BD16 are not implemented. Unicode (and the W3C) are making a great effort to deal with some special problematic cases in "streamed" plain text. I don't think this is the way to go – particular issues should be fixed by a high level interface taking into account the needs of the document. And here is where luatex excels, because everything related to bidi writing is under our control.

```

4657 (*basic-r)
4658 Babel = Babel or {}
4659
4660 Babel.bidi_enabled = true
4661
4662 require('babel-data-bidi.lua')
4663
4664 local characters = Babel.characters
4665 local ranges = Babel.ranges
4666
4667 local DIR = node.id("dir")
4668
4669 local function dir_mark(head, from, to, outer)
4670   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
4671   local d = node.new(DIR)
4672   d.dir = '+' .. dir
4673   node.insert_before(head, from, d)
4674   d = node.new(DIR)
4675   d.dir = '-' .. dir
4676   node.insert_after(head, to, d)
4677 end
4678
4679 function Babel.bidi(head, ispar)
4680   local first_n, last_n      -- first and last char with nums
4681   local last_es              -- an auxiliary 'last' used with nums
4682   local first_d, last_d      -- first and last char in L/R block
4683   local dir, dir_real

```

Next also depends on script/lang (<al>/<r>). To be set by babel. tex.pardir is dangerous, could be (re)set but it should be changed only in vmode. There are two strong's – strong = l/al/r and strong\_lr = l/r (there must be a better way):

```

4684   local strong = ('TRT' == tex.pardir) and 'r' or 'l'
4685   local strong_lr = (strong == 'l') and 'l' or 'r'
4686   local outer = strong
4687
4688   local new_dir = false
4689   local first_dir = false
4690   local inmath = false
4691
4692   local last_lr
4693

```

```

4694 local type_n = ''
4695
4696 for item in node.traverse(head) do
4697   -- three cases: glyph, dir, otherwise
4698   if item.id == node.id'glyph'
4699     or (item.id == 7 and item.subtype == 2) then
4700
4701     local itemchar
4702     if item.id == 7 and item.subtype == 2 then
4703       itemchar = item.replace.char
4704     else
4705       itemchar = item.char
4706     end
4707     local chardata = characters[itemchar]
4708     dir = chardata and chardata.d or nil
4709     if not dir then
4710       for nn, et in ipairs(ranges) do
4711         if itemchar < et[1] then
4712           break
4713         elseif itemchar <= et[2] then
4714           dir = et[3]
4715           break
4716         end
4717       end
4718     end
4719     dir = dir or 'l'
4720     if inmath then dir = ('TRT' == tex.mathdir) and 'r' or 'l' end
4721

```

Next is based on the assumption babel sets the language AND switches the script with its dir. We treat a language block as a separate Unicode sequence. The following piece of code is executed at the first glyph after a 'dir' node. We don't know the current language until then. This is not exactly true, as the math mode may insert explicit dirs in the node list, so, for the moment there is a hack by brute force (just above).

```

4722   if new_dir then
4723     attr_dir = 0
4724     for at in node.traverse(item.attr) do
4725       if at.number == luatexbase.registernumber'bbl@attr@dir' then
4726         attr_dir = at.value % 3
4727       end
4728     end
4729     if attr_dir == 1 then
4730       strong = 'r'
4731     elseif attr_dir == 2 then
4732       strong = 'al'
4733     else
4734       strong = 'l'
4735     end
4736     strong_lr = (strong == 'l') and 'l' or 'r'
4737     outer = strong_lr
4738     new_dir = false
4739   end
4740
4741   if dir == 'nsm' then dir = strong end -- W1

```

**Numbers.** The dual <al>/<r> system for R is somewhat cumbersome.

```

4742   dir_real = dir -- We need dir_real to set strong below
4743   if dir == 'al' then dir = 'r' end -- W3

```

By W2, there are no <en> <et> <es> if strong == <al>, only <an>. Therefore, there are not <et en> nor <en et>, W5 can be ignored, and W6 applied:

```

4744     if strong == 'al' then
4745         if dir == 'en' then dir = 'an' end           -- W2
4746         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
4747         strong_lr = 'r'                               -- W3
4748     end

```

Once finished the basic setup for glyphs, consider the two other cases: dir node and the rest.

```

4749     elseif item.id == node.id'dir' and not inmath then
4750         new_dir = true
4751         dir = nil
4752     elseif item.id == node.id'math' then
4753         inmath = (item.subtype == 0)
4754     else
4755         dir = nil           -- Not a char
4756     end

```

Numbers in R mode. A sequence of <en>, <et>, <an>, <es> and <cs> is typeset (with some rules) in L mode. We store the starting and ending points, and only when anything different is found (including nil, ie, a non-char), the textdir is set. This means you cannot insert, say, a whatsit, but this is what I would expect (with luacolor you may colorize some digits). Anyway, this behavior could be changed with a switch in the future. Note in the first branch only <an> is relevant if <al>.

```

4757     if dir == 'en' or dir == 'an' or dir == 'et' then
4758         if dir ~= 'et' then
4759             type_n = dir
4760         end
4761         first_n = first_n or item
4762         last_n = last_es or item
4763         last_es = nil
4764     elseif dir == 'es' and last_n then -- W3+W6
4765         last_es = item
4766     elseif dir == 'cs' then           -- it's right - do nothing
4767     elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
4768         if strong_lr == 'r' and type_n ~= '' then
4769             dir_mark(head, first_n, last_n, 'r')
4770         elseif strong_lr == 'l' and first_d and type_n == 'an' then
4771             dir_mark(head, first_n, last_n, 'r')
4772             dir_mark(head, first_d, last_d, outer)
4773             first_d, last_d = nil, nil
4774         elseif strong_lr == 'l' and type_n ~= '' then
4775             last_d = last_n
4776         end
4777         type_n = ''
4778         first_n, last_n = nil, nil
4779     end

```

R text in L, or L text in R. Order of dir\_ mark's are relevant: d goes outside n, and therefore it's emitted after. See dir\_mark to understand why (but is the nesting actually necessary or is a flat dir structure enough?). Only L, R (and AL) chars are taken into account – everything else, including spaces, whatsits, etc., are ignored:

```

4780     if dir == 'l' or dir == 'r' then
4781         if dir ~= outer then
4782             first_d = first_d or item
4783             last_d = item
4784         elseif first_d and dir ~= strong_lr then

```



```

4785         dir_mark(head, first_d, last_d, outer)
4786         first_d, last_d = nil, nil
4787     end
4788 end

```

**Mirroring.** Each chunk of text in a certain language is considered a “closed” sequence. If <r on r> and <l on l>, it’s clearly <r> and <l>, resp’tly, but with other combinations depends on outer. From all these, we select only those resolving <on> → <r>. At the beginning (when last\_lr is nil) of an R text, they are mirrored directly.

TODO - numbers in R mode are processed. It doesn’t hurt, but should not be done.

```

4789     if dir and not last_lr and dir ~= 'l' and outer == 'r' then
4790         item.char = characters[item.char] and
4791             characters[item.char].m or item.char
4792     elseif (dir or new_dir) and last_lr ~= item then
4793         local mir = outer .. strong_lr .. (dir or outer)
4794         if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
4795             for ch in node.traverse(node.next(last_lr)) do
4796                 if ch == item then break end
4797                 if ch.id == node.id'glyph' and characters[ch.char] then
4798                     ch.char = characters[ch.char].m or ch.char
4799                 end
4800             end
4801         end
4802     end

```

Save some values for the next iteration. If the current node is ‘dir’, open a new sequence. Since dir could be changed, strong is set with its real value (dir\_real).

```

4803     if dir == 'l' or dir == 'r' then
4804         last_lr = item
4805         strong = dir_real -- Don't search back - best save now
4806         strong_lr = (strong == 'l') and 'l' or 'r'
4807     elseif new_dir then
4808         last_lr = nil
4809     end
4810 end

```

Mirror the last chars if they are no directed. And make sure any open block is closed, too.

```

4811     if last_lr and outer == 'r' then
4812         for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
4813             if characters[ch.char] then
4814                 ch.char = characters[ch.char].m or ch.char
4815             end
4816         end
4817     end
4818     if first_n then
4819         dir_mark(head, first_n, last_n, outer)
4820     end
4821     if first_d then
4822         dir_mark(head, first_d, last_d, outer)
4823     end

```

In boxes, the dir node could be added before the original head, so the actual head is the previous node.

```

4824     return node.prev(head) or head
4825 end
4826 </basic-r>

```

And here the Lua code for bidi=basic:

```

4827 <*basic>

```

```

4828 Babel = Babel or {}
4829
4830 -- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
4831
4832 Babel.fontmap = Babel.fontmap or {}
4833 Babel.fontmap[0] = {}      -- l
4834 Babel.fontmap[1] = {}      -- r
4835 Babel.fontmap[2] = {}      -- al/an
4836
4837 Babel.bidi_enabled = true
4838 Babel.mirroring_enabled = true
4839
4840 -- Temporary:
4841
4842 if harf then
4843   Babel.mirroring_enabled = false
4844 end
4845
4846 require('babel-data-bidi.lua')
4847
4848 local characters = Babel.characters
4849 local ranges = Babel.ranges
4850
4851 local DIR = node.id('dir')
4852 local GLYPH = node.id('glyph')
4853
4854 local function insert_implicit(head, state, outer)
4855   local new_state = state
4856   if state.sim and state.eim and state.sim ~= state.eim then
4857     dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
4858     local d = node.new(DIR)
4859     d.dir = '+' .. dir
4860     node.insert_before(head, state.sim, d)
4861     local d = node.new(DIR)
4862     d.dir = '-' .. dir
4863     node.insert_after(head, state.eim, d)
4864   end
4865   new_state.sim, new_state.eim = nil, nil
4866   return head, new_state
4867 end
4868
4869 local function insert_numeric(head, state)
4870   local new
4871   local new_state = state
4872   if state.san and state.ean and state.san ~= state.ean then
4873     local d = node.new(DIR)
4874     d.dir = '+TLT'
4875     _, new = node.insert_before(head, state.san, d)
4876     if state.san == state.sim then state.sim = new end
4877     local d = node.new(DIR)
4878     d.dir = '-TLT'
4879     _, new = node.insert_after(head, state.ean, d)
4880     if state.ean == state.eim then state.eim = new end
4881   end
4882   new_state.san, new_state.ean = nil, nil
4883   return head, new_state
4884 end
4885
4886 -- TODO - \hbox with an explicit dir can lead to wrong results

```

```

4887 -- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>. A small attempt
4888 -- was s made to improve the situation, but the problem is the 3-dir
4889 -- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
4890 -- well.
4891
4892 function Babel.bidi(head, ispar, hdir)
4893   local d    -- d is used mainly for computations in a loop
4894   local prev_d = ''
4895   local new_d = false
4896
4897   local nodes = {}
4898   local outer_first = nil
4899   local inmath = false
4900
4901   local glue_d = nil
4902   local glue_i = nil
4903
4904   local has_en = false
4905   local first_et = nil
4906
4907   local ATDIR = luatexbase.registernumber'bbl@attr@dir'
4908
4909   local save_outer
4910   local temp = node.get_attribute(head, ATDIR)
4911   if temp then
4912     temp = temp % 3
4913     save_outer = (temp == 0 and 'l') or
4914                  (temp == 1 and 'r') or
4915                  (temp == 2 and 'al')
4916   elseif ispar then -- Or error? Shouldn't happen
4917     save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
4918   else -- Or error? Shouldn't happen
4919     save_outer = ('TRT' == hdir) and 'r' or 'l'
4920   end
4921   -- when the callback is called, we are just _after_ the box,
4922   -- and the textdir is that of the surrounding text
4923   -- if not ispar and hdir ~= tex.textdir then
4924   --   save_outer = ('TRT' == hdir) and 'r' or 'l'
4925   -- end
4926   local outer = save_outer
4927   local last = outer
4928   -- 'al' is only taken into account in the first, current loop
4929   if save_outer == 'al' then save_outer = 'r' end
4930
4931   local fontmap = Babel.fontmap
4932
4933   for item in node.traverse(head) do
4934
4935     -- In what follows, #node is the last (previous) node, because the
4936     -- current one is not added until we start processing the neutrals.
4937
4938     -- three cases: glyph, dir, otherwise
4939     if item.id == GLYPH
4940        or (item.id == 7 and item.subtype == 2) then
4941
4942       local d_font = nil
4943       local item_r
4944       if item.id == 7 and item.subtype == 2 then
4945         item_r = item.replace -- automatic discs have just 1 glyph

```

```

4946     else
4947         item_r = item
4948     end
4949     local chardata = characters[item_r.char]
4950     d = chardata and chardata.d or nil
4951     if not d or d == 'nsm' then
4952         for nn, et in ipairs(ranges) do
4953             if item_r.char < et[1] then
4954                 break
4955             elseif item_r.char <= et[2] then
4956                 if not d then d = et[3]
4957                 elseif d == 'nsm' then d_font = et[3]
4958                 end
4959                 break
4960             end
4961         end
4962     end
4963     d = d or 'l'
4964
4965     -- A short 'pause' in bidi for mapfont
4966     d_font = d_font or d
4967     d_font = (d_font == 'l' and 0) or
4968             (d_font == 'nsm' and 0) or
4969             (d_font == 'r' and 1) or
4970             (d_font == 'al' and 2) or
4971             (d_font == 'an' and 2) or nil
4972     if d_font and fontmap and fontmap[d_font][item_r.font] then
4973         item_r.font = fontmap[d_font][item_r.font]
4974     end
4975
4976     if new_d then
4977         table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
4978         if inmath then
4979             attr_d = 0
4980         else
4981             attr_d = node.get_attribute(item, ATDIR)
4982             attr_d = attr_d % 3
4983         end
4984         if attr_d == 1 then
4985             outer_first = 'r'
4986             last = 'r'
4987         elseif attr_d == 2 then
4988             outer_first = 'r'
4989             last = 'al'
4990         else
4991             outer_first = 'l'
4992             last = 'l'
4993         end
4994         outer = last
4995         has_en = false
4996         first_et = nil
4997         new_d = false
4998     end
4999
5000     if glue_d then
5001         if (d == 'l' and 'l' or 'r') ~= glue_d then
5002             table.insert(nodes, {glue_i, 'on', nil})
5003         end
5004         glue_d = nil

```

```

5005         glue_i = nil
5006     end
5007
5008     elseif item.id == DIR then
5009         d = nil
5010         new_d = true
5011
5012     elseif item.id == node.id'glue' and item.subtype == 13 then
5013         glue_d = d
5014         glue_i = item
5015         d = nil
5016
5017     elseif item.id == node.id'math' then
5018         inmath = (item.subtype == 0)
5019
5020     else
5021         d = nil
5022     end
5023
5024     -- AL <= EN/ET/ES      -- W2 + W3 + W6
5025     if last == 'al' and d == 'en' then
5026         d = 'an'          -- W3
5027     elseif last == 'al' and (d == 'et' or d == 'es') then
5028         d = 'on'          -- W6
5029     end
5030
5031     -- EN + CS/ES + EN      -- W4
5032     if d == 'en' and #nodes >= 2 then
5033         if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
5034             and nodes[#nodes-1][2] == 'en' then
5035             nodes[#nodes][2] = 'en'
5036         end
5037     end
5038
5039     -- AN + CS + AN        -- W4 too, because uax9 mixes both cases
5040     if d == 'an' and #nodes >= 2 then
5041         if (nodes[#nodes][2] == 'cs')
5042             and nodes[#nodes-1][2] == 'an' then
5043             nodes[#nodes][2] = 'an'
5044         end
5045     end
5046
5047     -- ET/EN              -- W5 + W7->l / W6->on
5048     if d == 'et' then
5049         first_et = first_et or (#nodes + 1)
5050     elseif d == 'en' then
5051         has_en = true
5052         first_et = first_et or (#nodes + 1)
5053     elseif first_et then -- d may be nil here !
5054         if has_en then
5055             if last == 'l' then
5056                 temp = 'l'    -- W7
5057             else
5058                 temp = 'en'   -- W5
5059             end
5060         else
5061             temp = 'on'       -- W6
5062         end
5063         for e = first_et, #nodes do

```

```

5064         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
5065     end
5066     first_et = nil
5067     has_en = false
5068 end
5069
5070 if d then
5071     if d == 'al' then
5072         d = 'r'
5073         last = 'al'
5074     elseif d == 'l' or d == 'r' then
5075         last = d
5076     end
5077     prev_d = d
5078     table.insert(nodes, {item, d, outer_first})
5079 end
5080
5081     outer_first = nil
5082
5083 end
5084
5085 -- TODO -- repeated here in case EN/ET is the last node. Find a
5086 -- better way of doing things:
5087 if first_et then      -- dir may be nil here !
5088     if has_en then
5089         if last == 'l' then
5090             temp = 'l'    -- W7
5091         else
5092             temp = 'en'   -- W5
5093         end
5094     else
5095         temp = 'on'      -- W6
5096     end
5097     for e = first_et, #nodes do
5098         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
5099     end
5100 end
5101
5102 -- dummy node, to close things
5103 table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
5104
5105 ----- NEUTRAL -----
5106
5107 outer = save_outer
5108 last = outer
5109
5110 local first_on = nil
5111
5112 for q = 1, #nodes do
5113     local item
5114
5115     local outer_first = nodes[q][3]
5116     outer = outer_first or outer
5117     last = outer_first or last
5118
5119     local d = nodes[q][2]
5120     if d == 'an' or d == 'en' then d = 'r' end
5121     if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
5122

```

```

5123   if d == 'on' then
5124       first_on = first_on or q
5125   elseif first_on then
5126       if last == d then
5127           temp = d
5128       else
5129           temp = outer
5130       end
5131       for r = first_on, q - 1 do
5132           nodes[r][2] = temp
5133           item = nodes[r][1]    -- MIRRORING
5134           if Babel.mirroring_enabled and item.id == GLYPH
5135               and temp == 'r' and characters[item.char] then
5136               item.char = characters[item.char].m or item.char
5137           end
5138       end
5139       first_on = nil
5140   end
5141
5142   if d == 'r' or d == 'l' then last = d end
5143 end
5144
5145 -----  IMPLICIT, REORDER  -----
5146
5147 outer = save_outer
5148 last = outer
5149
5150 local state = {}
5151 state.has_r = false
5152
5153 for q = 1, #nodes do
5154
5155     local item = nodes[q][1]
5156
5157     outer = nodes[q][3] or outer
5158
5159     local d = nodes[q][2]
5160
5161     if d == 'nsm' then d = last end          -- W1
5162     if d == 'en' then d = 'an' end
5163     local isdir = (d == 'r' or d == 'l')
5164
5165     if outer == 'l' and d == 'an' then
5166         state.san = state.san or item
5167         state.ean = item
5168     elseif state.san then
5169         head, state = insert_numeric(head, state)
5170     end
5171
5172     if outer == 'l' then
5173         if d == 'an' or d == 'r' then      -- im -> implicit
5174             if d == 'r' then state.has_r = true end
5175             state.sim = state.sim or item
5176             state.eim = item
5177         elseif d == 'l' and state.sim and state.has_r then
5178             head, state = insert_implicit(head, state, outer)
5179         elseif d == 'l' then
5180             state.sim, state.eim, state.has_r = nil, nil, false
5181         end

```

```

5182     else
5183         if d == 'an' or d == 'l' then
5184             if nodes[q][3] then -- nil except after an explicit dir
5185                 state.sim = item -- so we move sim 'inside' the group
5186             else
5187                 state.sim = state.sim or item
5188             end
5189             state.eim = item
5190         elseif d == 'r' and state.sim then
5191             head, state = insert_implicit(head, state, outer)
5192         elseif d == 'r' then
5193             state.sim, state.eim = nil, nil
5194         end
5195     end
5196
5197     if isdir then
5198         last = d -- Don't search back - best save now
5199     elseif d == 'on' and state.san then
5200         state.san = state.san or item
5201         state.ean = item
5202     end
5203
5204 end
5205
5206 return node.prev(head) or head
5207 end
5208 </basic>

```

## 15 Data for CJK

It is a boring file and it is not shown here (see the generated file), but here is a sample:

```

└─[0x0021]={c='ex'},
└─[0x0024]={c='pr'},
└─[0x0025]={c='po'},
└─[0x0028]={c='op'},
└─[0x0029]={c='cp'},
└─[0x002B]={c='pr'},

```

For the meaning of these codes, see the Unicode standard.

## 16 The ‘nil’ language

This ‘language’ does nothing, except setting the hyphenation patterns to nohyphenation. For this language currently no special definitions are needed or available.

The macro `\LdfInit` takes care of preventing that this file is loaded more than once, checking the category code of the @ sign, etc.

```

5209 <nil>
5210 \ProvidesLanguage{nil}[<<date>> <<version>> Nil language]
5211 \LdfInit{nil}{datenil}

```

When this file is read as an option, i.e. by the `\usepackage` command, `nil` could be an ‘unknown’ language in which case we have to make it known.

```

5212 \ifx\l@nil\undefined
5213   \newlanguage\l@nil

```



```

5214 \@namedef{bbl@hyphendata@the\l@nil}{\{}} % Remove warning
5215 \let\bbl@elt\relax
5216 \edef\bbl@languages{% Add it to the list of languages
5217 \bbl@languages\bbl@elt{nil}{\the\l@nil}{\{}}
5218 \fi

```

This macro is used to store the values of the hyphenation parameters `\lefthyphenmin` and `\righthyphenmin`.

```
5219 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}
```

The next step consists of defining commands to switch to (and from) the ‘nil’ language.

```

\captionnil
\datenil 5220 \let\captionnil\@empty
5221 \let\datenil\@empty

```

The macro `\ldf@finish` takes care of looking for a configuration file, setting the main language to be switched on at `\begin{document}` and resetting the category code of `@` to its original value.

```

5222 \ldf@finish{nil}
5223 \</nil>

```

## 17 Support for Plain T<sub>E</sub>X (plain.def)

### 17.1 Not renaming hyphen.tex

As Don Knuth has declared that the filename `hyphen.tex` may only be used to designate *his* version of the american English hyphenation patterns, a new solution has to be found in order to be able to load hyphenation patterns for other languages in a plain-based T<sub>E</sub>X-format. When asked he responded:

That file name is “sacred”, and if anybody changes it they will cause severe upward/downward compatibility headaches.

People can have a file `locallyphen.tex` or whatever they like, but they mustn’t diddle with `hyphen.tex` (or `plain.tex` except to preload additional fonts).

The files `bplain.tex` and `blplain.tex` can be used as replacement wrappers around `plain.tex` and `lplain.tex` to achieve the desired effect, based on the `babel` package. If you load each of them with `iniTEX`, you will get a file called either `bplain.fmt` or `blplain.fmt`, which you can use as replacements for `plain.fmt` and `lplain.fmt`. As these files are going to be read as the first thing `iniTEX` sees, we need to set some category codes just to be able to change the definition of `\input`

```

5224 \<*bplain | blplain>
5225 \catcode`\{=1 % left brace is begin-group character
5226 \catcode`\}=2 % right brace is end-group character
5227 \catcode`\#=6 % hash mark is macro parameter character

```

Now let’s see if a file called `hyphen.cfg` can be found somewhere on T<sub>E</sub>X’s input path by trying to open it for reading...

```
5228 \openin 0 hyphen.cfg
```

If the file wasn’t found the following test turns out true.

```

5229 \ifeof0
5230 \else

```

When `hyphen.cfg` could be opened we make sure that *it* will be read instead of the file `hyphen.tex` which should (according to Don Knuth's ruling) contain the american English hyphenation patterns and nothing else.

We do this by first saving the original meaning of `\input` (and I use a one letter control sequence for that so as not to waste multi-letter control sequence on this in the format).

```
5231 \let\input
```

Then `\input` is defined to forget about its argument and load `hyphen.cfg` instead.

```
5232 \def\input #1 {%
5233   \let\input\input
5234   \input{hyphen.cfg}
```

Once that's done the original meaning of `\input` can be restored and the definition of `\input` can be forgotten.

```
5235 \let\input\undefined
5236 }
5237 \fi
5238 </bplain | bplain>
```

Now that we have made sure that `hyphen.cfg` will be loaded at the right moment it is time to load `plain.tex`.

```
5239 <bplain>\input{plain.tex}
5240 <bplain>\input{lplain.tex}
```

Finally we change the contents of `\fmtname` to indicate that this is *not* the plain format, but a format based on plain with the `babel` package preloaded.

```
5241 <bplain>\def\fmtname{babel-plain}
5242 <bplain>\def\fmtname{babel-lplain}
```

When you are using a different format, based on `plain.tex` you can make a copy of `blplain.tex`, rename it and replace `plain.tex` with the name of your format file.

## 17.2 Emulating some $\text{\LaTeX}$ features

The following code duplicates or emulates parts of  $\text{\LaTeX} 2_{\epsilon}$  that are needed for `babel`.

```
5243 <*plain>
5244 \def\@empty{}
5245 \def\loadlocalcfg#1{%
5246   \openin0#1.cfg
5247   \ifeof0
5248     \closein0
5249   \else
5250     \closein0
5251     {\immediate\write16{*****}%
5252      \immediate\write16{* Local config file #1.cfg used}%
5253      \immediate\write16{*}%
5254     }
5255     \input #1.cfg\relax
5256   \fi
5257   \@endofldf}
```

## 17.3 General tools

A number of  $\text{\LaTeX}$  macro's that are needed later on.

```
5258 \long\def\@firstofone#1{#1}
5259 \long\def\@firstoftwo#1#2{#1}
5260 \long\def\@secondoftwo#1#2{#2}
```

```

5261 \def\@nnil{\@nil}
5262 \def\@gobbletwo#1#2{}
5263 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
5264 \def\@star@or@long#1{%
5265   \@ifstar
5266   {\let\l@ngrel@x\relax#1}%
5267   {\let\l@ngrel@x\long#1}}
5268 \let\l@ngrel@x\relax
5269 \def\@car#1#2\@nil{#1}
5270 \def\@cdr#1#2\@nil{#2}
5271 \let\@typeset@protect\relax
5272 \let\protected@edef\edef
5273 \long\def\@gobble#1{}
5274 \edef\@backslashchar{\expandafter\@gobble\string\}
5275 \def\strip@prefix#1>{}
5276 \def\g@addto@macro#1#2{%
5277   \toks@{\expandafter{#1#2}}%
5278   \xdef#1{\the\toks@}}
5279 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
5280 \def\@nameuse#1{\csname #1\endcsname}
5281 \def\@ifundefined#1{%
5282   \expandafter\ifx\csname#1\endcsname\relax
5283     \expandafter\@firstoftwo
5284   \else
5285     \expandafter\@secondoftwo
5286   \fi}
5287 \def\@expandtwoargs#1#2#3{%
5288   \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
5289 \def\zap@space#1 #2{%
5290   #1%
5291   \ifx#2\@empty\else\expandafter\zap@space\fi
5292   #2}

```

$\LaTeX 2_{\epsilon}$  has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

5293 \ifx\@preamblecmds\@undefined
5294   \def\@preamblecmds{}
5295 \fi
5296 \def\@onlypreamble#1{%
5297   \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
5298     \@preamblecmds\do#1}}
5299 \@onlypreamble\@onlypreamble

```

Mimick  $\LaTeX$ 's `\AtBeginDocument`; for this to work the user needs to add `\begindocument` to his file.

```

5300 \def\begindocument{%
5301   \@begindocumenthook
5302   \global\let\@begindocumenthook\@undefined
5303   \def\do##1{\global\let##1\@undefined}%
5304   \@preamblecmds
5305   \global\let\do\noexpand}
5306 \ifx\@begindocumenthook\@undefined
5307   \def\@begindocumenthook{}
5308 \fi
5309 \@onlypreamble\@begindocumenthook
5310 \def\AtBeginDocument{\g@addto@macro\@begindocumenthook}

```

We also have to mimick  $\LaTeX$ 's `\AtEndOfPackage`. Our replacement macro is much simpler; it stores its argument in `\@endofldf`.

```

5311 \def\AtEndOfPackage#1{\g@addto@macro\@endofldf{#1}}
5312 \@onlypreamble\AtEndOfPackage
5313 \def\@endofldf{}
5314 \@onlypreamble\@endofldf
5315 \let\bbl@afterlang\@empty
5316 \chardef\bbl@opt@hyphenmap\z@

```

$\LaTeX$  needs to be able to switch off writing to its auxiliary files; plain doesn't have them by default.

```

5317 \ifx\if@files\@undefined
5318 \expandafter\let\csname if@files\expandafter\endcsname
5319 \csname iffalse\endcsname
5320 \fi

```

Mimick  $\LaTeX$ 's commands to define control sequences.

```

5321 \def\newcommand{\@star@or@long\new@command}
5322 \def\new@command#1{%
5323   \@testopt{\@newcommand#1}0}
5324 \def\@newcommand#1[#2]{%
5325   \@ifnextchar [{\@xargdef#1[#2]}%
5326     {\@argdef#1[#2]}}
5327 \long\def\@argdef#1[#2]#3{%
5328   \@yargdef#1\@ne{#2}{#3}}
5329 \long\def\@xargdef#1[#2][#3]#4{%
5330   \expandafter\def\expandafter#1\expandafter{%
5331     \expandafter\@protected@testopt\expandafter #1%
5332     \csname\string#1\expandafter\endcsname{#3}}%
5333   \expandafter\@yargdef \csname\string#1\endcsname
5334   \tw@{#2}{#4}}
5335 \long\def\@yargdef#1#2#3{%
5336   \@tempcnta#3\relax
5337   \advance \@tempcnta \@ne
5338   \let\@hash@\relax
5339   \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
5340   \@tempcntb #2%
5341   \@whilenum \@tempcntb < \@tempcnta
5342   \do{%
5343     \edef\reserved@a{\reserved@a\@hash@the\@tempcntb}%
5344     \advance \@tempcntb \@ne}%
5345   \let\@hash@###
5346   \l@ngrel@x\expandafter\def\expandafter#1\reserved@a}
5347 \def\providecommand{\@star@or@long\provide@command}
5348 \def\provide@command#1{%
5349   \begingroup
5350     \escapechar\m@ne\xdef\@gtempa{\string#1}%
5351   \endgroup
5352   \expandafter\@ifundefined\@gtempa
5353     {\def\reserved@a{\new@command#1}}%
5354     {\let\reserved@a\relax
5355     \def\reserved@a{\new@command\reserved@a}}%
5356   \reserved@a}%
5357 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
5358 \def\declare@robustcommand#1{%
5359   \edef\reserved@a{\string#1}%
5360   \def\reserved@b{#1}%
5361   \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
5362   \edef#1{%
5363     \ifx\reserved@a\reserved@b

```

```

5364      \noexpand\x@protect
5365      \noexpand#1%
5366      \fi
5367      \noexpand\protect
5368      \expandafter\noexpand\csname
5369      \expandafter\@gobble\string#1 \endcsname
5370    }%
5371    \expandafter\new@command\csname
5372      \expandafter\@gobble\string#1 \endcsname
5373  }
5374  \def\x@protect#1{%
5375    \ifx\protect\@typeset@protect\else
5376      \@x@protect#1%
5377    \fi
5378  }
5379  \def\@x@protect#1\fi#2#3{%
5380    \fi\protect#1%
5381  }

```

The following little macro `\in@` is taken from `latex.ltx`; it checks whether its first argument is part of its second argument. It uses the boolean `\in@`; allocating a new boolean inside conditionally executed code is not possible, hence the construct with the temporary definition of `\bbl@tempa`.

```

5382 \def\bbl@tempa{\csname newif\endcsname\ifin@}
5383 \ifx\in@\@undefined
5384   \def\in@#1#2{%
5385     \def\in@##1##2##3\in@{%
5386       \ifx\in@##2\in@false\else\in@true\fi}%
5387     \in@#2#1\in@\in@}
5388 \else
5389   \let\bbl@tempa\@empty
5390 \fi
5391 \bbl@tempa

```

$\text{\LaTeX}$  has a macro to check whether a certain package was loaded with specific options. The command has two extra arguments which are code to be executed in either the true or false case. This is used to detect whether the document needs one of the accents to be activated (`activegrave` and `activeacute`). For plain  $\text{\TeX}$  we assume that the user wants them to be active by default. Therefore the only thing we do is execute the third argument (the code for the true case).

```

5392 \def\@ifpackagewith#1#2#3#4{#3}

```

The  $\text{\LaTeX}$  macro `\@ifl@aded` checks whether a file was loaded. This functionality is not needed for plain  $\text{\TeX}$  but we need the macro to be defined as a no-op.

```

5393 \def\@ifl@aded#1#2#3#4{}

```

For the following code we need to make sure that the commands `\newcommand` and `\providecommand` exist with some sensible definition. They are not fully equivalent to their  $\text{\LaTeX 2}_\epsilon$  versions; just enough to make things work in plain  $\text{\TeX}$  environments.

```

5394 \ifx\@tempcnta\@undefined
5395   \csname newcount\endcsname\@tempcnta\relax
5396 \fi
5397 \ifx\@tempcntb\@undefined
5398   \csname newcount\endcsname\@tempcntb\relax
5399 \fi

```

To prevent wasting two counters in  $\text{\LaTeX 2.09}$  (because counters with the same name are allocated later by it) we reset the counter that holds the next free counter (`\count10`).

```

5400 \ifx\bye\@undefined
5401   \advance\count10 by -2\relax
5402 \fi
5403 \ifx\@ifnextchar\@undefined
5404   \def\@ifnextchar#1#2#3{%
5405     \let\reserved@d=#1%
5406     \def\reserved@a{#2}\def\reserved@b{#3}%
5407     \futurelet\@let@token\@ifnch}
5408 \def\@ifnch{%
5409   \ifx\@let@token\@sptoken
5410     \let\reserved@c\@xifnch
5411   \else
5412     \ifx\@let@token\reserved@d
5413       \let\reserved@c\reserved@a
5414     \else
5415       \let\reserved@c\reserved@b
5416     \fi
5417   \fi
5418   \reserved@c}
5419 \def\:{\let\@sptoken= } \: % this makes \@sptoken a space token
5420 \def\:{\@xifnch} \expandafter\def\:{\futurelet\@let@token\@ifnch}
5421 \fi
5422 \def\@testopt#1#2{%
5423   \@ifnextchar[{#1}]{#1[#2]}}
5424 \def\@protected@testopt#1{%
5425   \ifx\protect\@typeset@protect
5426     \expandafter\@testopt
5427   \else
5428     \@x@protect#1%
5429   \fi}
5430 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
5431   #2\relax}\fi}
5432 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
5433   \else\expandafter\@gobble\fi{#1}}

```

## 17.4 Encoding related macros

Code from `ltoutenc.dtx`, adapted for use in the plain  $\TeX$  environment.

```

5434 \def\DeclareTextCommand{%
5435   \@dec@text@cmd\providecommand
5436 }
5437 \def\ProvideTextCommand{%
5438   \@dec@text@cmd\providecommand
5439 }
5440 \def\DeclareTextSymbol#1#2#3{%
5441   \@dec@text@cmd\chardef#1{#2}#3\relax
5442 }
5443 \def\@dec@text@cmd#1#2#3{%
5444   \expandafter\def\expandafter#2%
5445     \expandafter{%
5446       \csname#3-cmd\expandafter\endcsname
5447       \expandafter#2%
5448       \csname#3\string#2\endcsname
5449     }%
5450 %   \let\@ifdefinable\@rc@ifdefinable
5451   \expandafter#1\csname#3\string#2\endcsname
5452 }
5453 \def\@current@cmd#1{%

```

```

5454 \ifx\protect\@typeset@protect\else
5455 \noexpand#1\expandafter\@gobble
5456 \fi
5457 }
5458 \def\@changed@cmd#1#2{%
5459 \ifx\protect\@typeset@protect
5460 \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
5461 \expandafter\ifx\csname ?\string#1\endcsname\relax
5462 \expandafter\def\csname ?\string#1\endcsname{%
5463 \@changed@x@err{#1}%
5464 }%
5465 \fi
5466 \global\expandafter\let
5467 \csname\cf@encoding \string#1\expandafter\endcsname
5468 \csname ?\string#1\endcsname
5469 \fi
5470 \csname\cf@encoding\string#1%
5471 \expandafter\endcsname
5472 \else
5473 \noexpand#1%
5474 \fi
5475 }
5476 \def\@changed@x@err#1{%
5477 \errhelp{Your command will be ignored, type <return> to proceed}%
5478 \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
5479 \def\DeclareTextCommandDefault#1{%
5480 \DeclareTextCommand#1?%
5481 }
5482 \def\ProvideTextCommandDefault#1{%
5483 \ProvideTextCommand#1?%
5484 }
5485 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
5486 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
5487 \def\DeclareTextAccent#1#2#3{%
5488 \DeclareTextCommand#1{#2}[1]{\accent#3 ##1}
5489 }
5490 \def\DeclareTextCompositeCommand#1#2#3#4{%
5491 \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
5492 \edef\reserved@b{\string##1}%
5493 \edef\reserved@c{%
5494 \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
5495 \ifx\reserved@b\reserved@c
5496 \expandafter\expandafter\expandafter\ifx
5497 \expandafter\@car\reserved@a\relax\relax\nil
5498 \@text@composite
5499 \else
5500 \edef\reserved@b##1{%
5501 \def\expandafter\noexpand
5502 \csname#2\string#1\endcsname###1{%
5503 \noexpand\@text@composite
5504 \expandafter\noexpand\csname#2\string#1\endcsname
5505 ###1\noexpand\@empty\noexpand\@text@composite
5506 {##1}%
5507 }%
5508 }%
5509 \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
5510 \fi
5511 \expandafter\def\csname\expandafter\string\csname
5512 #2\endcsname\string#1-\string#3\endcsname{#4}

```

```

5513 \else
5514 \errhelp{Your command will be ignored, type <return> to proceed}%
5515 \errmessage{\string\DeclareTextCompositeCommand\space used on
5516 inappropriate command \protect#1}
5517 \fi
5518 }
5519 \def\@text@composite#1#2#3\@text@composite{%
5520 \expandafter\@text@composite@x
5521 \csname\string#1-\string#2\endcsname
5522 }
5523 \def\@text@composite@x#1#2{%
5524 \ifx#1\relax
5525 #2%
5526 \else
5527 #1%
5528 \fi
5529 }
5530 %
5531 \def\@strip@args#1:#2-#3\@strip@args{#2}
5532 \def\DeclareTextComposite#1#2#3#4{%
5533 \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
5534 \bgroup
5535 \lcode` \@=#4%
5536 \lowercase{%
5537 \egroup
5538 \reserved@a @%
5539 }%
5540 }
5541 %
5542 \def\UseTextSymbol#1#2{%
5543 \let\@curr@enc\cf@encoding
5544 \@use@text@encoding{#1}%
5545 #2%
5546 \@use@text@encoding\@curr@enc
5547 }
5548 \def\UseTextAccent#1#2#3{%
5549 \let\@curr@enc\cf@encoding
5550 \@use@text@encoding{#1}%
5551 #2{\@use@text@encoding\@curr@enc\selectfont#3}%
5552 \@use@text@encoding\@curr@enc
5553 }
5554 \def\@use@text@encoding#1{%
5555 \edef\f@encoding{#1}%
5556 \xdef\font@name{%
5557 \csname\curr@fontshape/\f@size\endcsname
5558 }%
5559 \pickup@font
5560 \font@name
5561 \@@enc@update
5562 }
5563 \def\DeclareTextSymbolDefault#1#2{%
5564 \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
5565 }
5566 \def\DeclareTextAccentDefault#1#2{%
5567 \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
5568 }
5569 \def\cf@encoding{OT1}

```

Currently we only use the  $\text{\LaTeX} 2_{\epsilon}$  method for accents for those that are known to be made



active in *some* language definition file.

```
5570 \DeclareTextAccent{"}{OT1}{127}
5571 \DeclareTextAccent{'}{OT1}{19}
5572 \DeclareTextAccent{^}{OT1}{94}
5573 \DeclareTextAccent`{OT1}{18}
5574 \DeclareTextAccent{~}{OT1}{126}
```

The following control sequences are used in `babel.def` but are not defined for PLAIN  $\TeX$ .

```
5575 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
5576 \DeclareTextSymbol{\textquotedblright}{OT1}{`"}
5577 \DeclareTextSymbol{\textquoteleft}{OT1}{``}
5578 \DeclareTextSymbol{\textquoteright}{OT1}{``'}
5579 \DeclareTextSymbol{\i}{OT1}{16}
5580 \DeclareTextSymbol{\ss}{OT1}{25}
```

For a couple of languages we need the  $\TeX$ -control sequence `\scriptsize` to be available. Because plain  $\TeX$  doesn't have such a sophisticated font mechanism as  $\TeX$  has, we just `\let` it to `\sevenrm`.

```
5581 \ifx\scriptsize\@undefined
5582   \let\scriptsize\sevenrm
5583 \fi
5584 \</plain>
```

## 18 Acknowledgements

I would like to thank all who volunteered as  $\beta$ -testers for their time. Michel Goossens supplied contributions for most of the other languages. Nico Poppelier helped polish the text of the documentation and supplied parts of the macros for the Dutch language. Paul Wackers and Werenfried Spit helped find and repair bugs. During the further development of the babel system I received much help from Bernd Raichle, for which I am grateful.

## References

- [1] Huda Smitshuijzen Abifares, *Arabic Typography*, Saqi, 2001.
- [2] Donald E. Knuth, *The  $\TeX$ book*, Addison-Wesley, 1986.
- [3] Leslie Lamport,  *$\TeX$ , A document preparation System*, Addison-Wesley, 1986.
- [4] K.F. Treebus. *Tekstwijzer, een gids voor het grafisch verwerken van tekst*, SDU Uitgeverij ('s-Gravenhage, 1988).
- [5] Hubert Partl, *German  $\TeX$* , *TUGboat* 9 (1988) #1, p. 70–72.
- [6] Leslie Lamport, in:  $\TeX$ hax Digest, Volume 89, #13, 17 February 1989.
- [7] Johannes Braams, Victor Eijkhout and Nico Poppelier, *The development of national  $\TeX$  styles*, *TUGboat* 10 (1989) #3, p. 401–406.
- [8] Yannis Haralambous, *Fonts & Encodings*, O'Reilly, 2007.
- [9] Jukka K. Korpela, *Unicode Explained*, O'Reilly, 2006.
- [10] Ken Lunde, *CJKV Information Processing*, O'Reilly, 2nd ed., 2009.
- [11] Joachim Schrod, *International  $\TeX$  is ready to use*, *TUGboat* 11 (1990) #1, p. 87–90.

- [12] Apostolos Syropoulos, Antonis Tsolomitis and Nick Sofroniu, *Digital typography using  $\mathcal{E}T_X$* , Springer, 2002, p. 301–373.