

# Babel

Version 3.55.2319  
2021/03/22

Johannes L. Braams  
Original author

Javier Bezos  
Current maintainer

Localization and  
internationalization

Unicode

T<sub>E</sub>X

pdfT<sub>E</sub>X

LuaT<sub>E</sub>X

XeT<sub>E</sub>X

# 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	6
1.3	Mostly monolingual documents	8
1.4	Modifiers	8
1.5	Troubleshooting	9
1.6	Plain	9
1.7	Basic language selectors	9
1.8	Auxiliary language selectors	10
1.9	More on selection	11
1.10	Shorthands	12
1.11	Package options	16
1.12	The base option	18
1.13	ini files	18
1.14	Selecting fonts	26
1.15	Modifying a language	28
1.16	Creating a language	30
1.17	Digits and counters	33
1.18	Dates	35
1.19	Accessing language info	35
1.20	Hyphenation and line breaking	36
1.21	Transforms	38
1.22	Selection based on BCP 47 tags	39
1.23	Selecting scripts	41
1.24	Selecting directions	41
1.25	Language attributes	45
1.26	Hooks	46
1.27	Languages supported by babel with ldf files	47
1.28	Unicode character properties in luatex	48
1.29	Tweaking some features	49
1.30	Tips, workarounds, known issues and notes	49
1.31	Current and future work	50
1.32	Tentative and experimental code	50
<b>2</b>	<b>Loading languages with language.dat</b>	<b>51</b>
2.1	Format	51
<b>3</b>	<b>The interface between the core of babel and the language definition files</b>	<b>52</b>
3.1	Guidelines for contributed languages	53
3.2	Basic macros	54
3.3	Skeleton	55
3.4	Support for active characters	56
3.5	Support for saving macro definitions	56
3.6	Support for extending macros	57
3.7	Macros common to a number of languages	57
3.8	Encoding-dependent strings	57
<b>4</b>	<b>Changes</b>	<b>61</b>
4.1	Changes in babel version 3.9	61

<b>II</b>	<b>Source code</b>	<b>61</b>
<b>5</b>	<b>Identification and loading of required files</b>	<b>62</b>
<b>6</b>	<b>locale directory</b>	<b>62</b>
<b>7</b>	<b>Tools</b>	<b>62</b>
7.1	Multiple languages . . . . .	67
7.2	The Package File ( <code>\LaTeX</code> , <code>babel.sty</code> ) . . . . .	67
7.3	base . . . . .	69
7.4	Conditional loading of shorthands . . . . .	71
7.5	Cross referencing macros . . . . .	73
7.6	Marks . . . . .	75
7.7	Preventing clashes with other packages . . . . .	76
7.7.1	ifthen . . . . .	76
7.7.2	varioref . . . . .	77
7.7.3	hhline . . . . .	77
7.7.4	hyperref . . . . .	78
7.7.5	fancyhdr . . . . .	78
7.8	Encoding and fonts . . . . .	78
7.9	Basic bidi support . . . . .	80
7.10	Local Language Configuration . . . . .	85
<b>8</b>	<b>The kernel of Babel (<code>babel.def</code>, <code>common</code>)</b>	<b>89</b>
8.1	Tools . . . . .	89
<b>9</b>	<b>Multiple languages</b>	<b>90</b>
9.1	Selecting the language . . . . .	93
9.2	Errors . . . . .	101
9.3	Hooks . . . . .	104
9.4	Setting up language files . . . . .	106
9.5	Shorthands . . . . .	108
9.6	Language attributes . . . . .	117
9.7	Support for saving macro definitions . . . . .	119
9.8	Short tags . . . . .	120
9.9	Hyphens . . . . .	120
9.10	Multiencoding strings . . . . .	122
9.11	Macros common to a number of languages . . . . .	128
9.12	Making glyphs available . . . . .	128
9.12.1	Quotation marks . . . . .	128
9.12.2	Letters . . . . .	130
9.12.3	Shorthands for quotation marks . . . . .	131
9.12.4	Umlauts and tremas . . . . .	132
9.13	Layout . . . . .	133
9.14	Load engine specific macros . . . . .	134
9.15	Creating and modifying languages . . . . .	134
<b>10</b>	<b>Adjusting the Babel behavior</b>	<b>154</b>
<b>11</b>	<b>Loading hyphenation patterns</b>	<b>156</b>
<b>12</b>	<b>Font handling with fontspec</b>	<b>161</b>

<b>13</b>	<b>Hooks for XeTeX and LuaTeX</b>	<b>165</b>
13.1	XeTeX . . . . .	165
13.2	Layout . . . . .	167
13.3	LuaTeX . . . . .	168
13.4	Southeast Asian scripts . . . . .	174
13.5	CJK line breaking . . . . .	178
13.6	Automatic fonts and ids switching . . . . .	178
13.7	Layout . . . . .	190
13.8	Auto bidi with basic and basic-r . . . . .	194
<b>14</b>	<b>Data for CJK</b>	<b>205</b>
<b>15</b>	<b>The ‘nil’ language</b>	<b>205</b>
<b>16</b>	<b>Support for Plain T<sub>E</sub>X (plain.def)</b>	<b>206</b>
16.1	Not renaming hyphen.tex . . . . .	206
16.2	Emulating some L <sub>A</sub> T <sub>E</sub> X features . . . . .	206
16.3	General tools . . . . .	207
16.4	Encoding related macros . . . . .	210
<b>17</b>	<b>Acknowledgements</b>	<b>213</b>

## Troubleshooting

Paragraph ended before \UTFviii@three@octets was complete . . . . .	5
No hyphenation patterns were preloaded for (babel) the language ‘LANG’ into the format . . . . .	6
You are loading directly a language style . . . . .	9
Unknown language ‘LANG’ . . . . .	9
Argument of \language@active@arg” has an extra } . . . . .	13
Package fontspec Warning: ‘Language ‘LANG’ not available for font ‘FONT’ with script ‘SCRIPT’ ‘Default’ language used instead’ . . . . .	28
Package babel Info: The following fonts are not babel standard families . . . . .	28

# Part I

## User guide

**What is this document about?** This user guide focuses on internationalization and localization with  $\LaTeX$  and `pdftex`, `xetex` and `luatex` with the `babel` package. There are also some notes on its use with Plain  $\TeX$ . Part II describes the code, and usually it can be ignored.

**What if I'm interested only in the latest changes?** Changes and new features with relation to version 3.8 are highlighted with `New X.XX`, and there are some notes for the latest versions in [the babel repository](#). The most recent features can be still unstable.

**Can I help?** Sure! If you are interested in the  $\TeX$  multilingual support, please join the [kadingira mail list](#). You can follow the development of `babel` in [GitHub](#) and make suggestions; feel free to fork it and make pull requests. If you are the author of a package, send to me a few test files which I'll add to mine, so that possible issues can be caught in the development phase.

**It doesn't work for me!** You can ask for help in some forums like [tex.stackexchange](#), but if you have found a bug, I strongly beg you to report it in [GitHub](#), which is much better than just complaining on an e-mail list or a web forum. Remember *warnings are not errors* by themselves, they just warn about possible problems or incompatibilities.

**How can I contribute a new language?** See section 3.1 for contributing a language.

**I only need learn the most basic features.** The first subsections (1.1-1.3) describe the traditional way of loading a language (with `ldf` files), which is usually all you need. The alternative way based on `ini` files, which complements the previous one (it does *not* replace it, although it is still necessary in some languages), is described below; go to 1.13.

**I don't like manuals. I prefer sample files.** This manual contains lots of examples and tips, but in [GitHub](#) there are many [sample files](#).

## 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. Another approach is making the language a global option in order to let other packages detect and use it. This is the standard way in  $\LaTeX$  for an option – in this case a language – to be recognized by several packages.

Many languages are compatible with `xetex` and `luatex`. With them you can use `babel` to localize the documents. When these engines are used, the Latin script is covered by default in current  $\LaTeX$  (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** 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. It assumes UTF-8, the default encoding:

PDFTEX

```
\documentclass{article}

\usepackage[T1]{fontenc}

\usepackage[french]{babel}

\begin{document}

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

\end{document}
```

Now consider something like:

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

With this setting, the package `varioref` will also see the option `french` and will be able to use it.

**EXAMPLE** And now a simple monolingual document in Russian (text from the Wikipedia) with `xetex` or `luatex`. 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).

LUATEX/XETEX

```
\documentclass[russian]{article}

\usepackage{babel}

\babelfont{rm}{DejaVu Serif}

\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}
```

**TROUBLESHOOTING** A common source of trouble is a wrong setting of the input encoding. Depending on the  $\text{\LaTeX}$  version you can get the following somewhat cryptic error:

```
! Paragraph ended before \UTFviii@three@octets was complete.
```

Or the more explanatory:

```
! Package inputenc Error: Invalid UTF-8 byte ...
```

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

**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, TeXLive, etc.) for further info about how to configure it.

**NOTE** With hyperref you may want to set the document language with something like:

```
\usepackage[pdflang=es-MX]{hyperref}
```

This is not currently done by babel and you must set it by hand.

**NOTE** Although it has been customary to recommend placing `\title`, `\author` and other elements printed by `\maketitle` after `\begin{document}`, mainly because of shorthands, it is advisable to keep them in the preamble. Currently there is no real need to use shorthands in those macros.

## 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  $\LaTeX$ , the preamble of the document:

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

would tell  $\LaTeX$  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 can 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 with pdfTeX follows. The main language is french, which is activated when the document begins. It assumes UTF-8:

PDFTEX

```
\documentclass{article}  
  
\usepackage[T1]{fontenc}  
  
\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}
```

**EXAMPLE** With xetex and luatex, 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.

LUATEX/XETEX

```
\documentclass{article}  
  
\usepackage[vietnamese,danish]{babel}  
  
\begin{document}  
  
\prefacename{} -- \alsoname{} -- \today  
  
\selectlanguage{vietnamese}
```



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

\end{document}
```

**NOTE** Once loaded a language, you can select it with the corresponding BCP47 tag. See section 1.22 for further details.

### 1.3 Mostly monolingual documents

**New 3.39** Very often, multilingual documents consist of a main language with small pieces of text in another languages (words, idioms, short sentences). Typically, all you need is to set the line breaking rules and, perhaps, the font. In such a case, babel now does not require declaring these secondary languages explicitly, because the basic settings are loaded on the fly when the language is selected (and also when provided in the optional argument of `\babelfont`, if used.)

This is particularly useful, too, when there are short texts of this kind coming from an external source whose contents are not known on beforehand (for example, titles in a bibliography). At this regard, it is worth remembering that `\babelfont` does *not* load any font until required, so that it can be used just in case.

**EXAMPLE** A trivial document with the default font in English and Spanish, and FreeSerif in Russian is:

LUATEX/XETEX

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

\babelfont[russian]{rm}{FreeSerif}

\begin{document}

English. \foreignlanguage{russian}{Русский}.
\foreignlanguage{spanish}{Español}.

\end{document}
```

**NOTE** Instead of its name, you may prefer to select the language with the corresponding BCP47 tag. This alternative, however, must be activated explicitly, because a two- or tree-letter word is a valid name for a language (eg, `yi`). See section 1.22 for further details.

### 1.4 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>

```
\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.

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

## 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.

**New 3.43** However, if the macro name does not match any language, it will get expanded as expected.

**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** [*<option-list>*]{*<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).

**New 3.44** As already said, captions and dates are not switched. However, with the optional argument you can switch them, too. So, you can write:

```
\foreignlanguage[date]{polish}{\today}
```

In addition, captions can be switched with `captions` (or both, of course, with `date`, `captions`). Until 3.43 you had to write something like `{\selectlanguage{..} ..}`, which was not always the most convenient way.

## 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*}` [*<option-list>*] {*<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.

## 1.9 More on selection

`\babeltags` {*<tag1>* = *<language1>*, *<tag2>* = *<language2>*, ...}

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

**WARNING** There is a clear drawback to this feature, namely, the ‘prefix’ `\text...` is heavily overloaded in  $\text{\LaTeX}$  and conflicts with existing macros may arise (`\textlatin`, `\textbar`, `\textit`, `\textcolor` and many others). The same applies to environments, because `arabic` conflicts with `\arabic`. Except if there is a reason for this ‘syntactical sugar’, the best option is to stick to the default selectors or to define your own alternatives.

**EXAMPLE** With

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

you can write

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

and

```
text
\begin{de}
  German text
\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 the option `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 four levels of shorthands: *user*, *language*, *system*, and *language user* (by order of precedence). In most cases, you will use only shorthands provided by languages.

**NOTE** Keep in mind 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, 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 deactivated with, eg, `\string`).

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

**TROUBLESHOOTING** 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**     $\{\langle shorthands-list \rangle\}$   
**\shorthandoff**    $\ast\{\langle shorthands-list \rangle\}$

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.

If you do not need shorthands, or prefer an alternative approach of your own, you may want to switch them off with the package option `shorthands=off`, as described below.

**\useshorthands**    $\ast\{\langle char \rangle\}$

The command `\useshorthands` 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 `\useshorthands*\{\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 `\useshorthands`. 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 can 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 can then set:

```
\defineshorthand[*polish,*portuguese]{"-"}{\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.

**\languageshorthands**  $\{\langle language \rangle\}$

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}}
```

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

**EXAMPLE** 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]{\{\languageshorthands{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.)

**EXAMPLE** Since by default shorthands are not activated until `\begin{document}`, you may use this macro when defining the `\title` in the preamble:

<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.

```
\title{Documento científico\babelshorthand{"-}técnico}
```

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.

`\aliasshorthand`  $\{\langle original \rangle\}\{\langle alias \rangle\}$

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{"}{/}`. For the reasons in the warning below, usage of this macro is not recommended.

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

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

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

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

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



**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.

## 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.

**KeepShorthandsActive** Tells babel not to deactivate shorthands after loading a language file, so that they are also available in the preamble.

**activeacute** For some languages babel supports this options to set `'` as a shorthand in case it is not done by default.

**activegrave** Same for ```.

**shorthands=**  $\langle char \rangle \langle char \rangle \dots$  | off

The only language shorthands activated are those given, like, eg:

```
\usepackage[esperanto,french,shorthands=:;!]{babel}
```

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  $\TeX$  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.

**safe=** none | ref | bib

Some  $\TeX$  macros are redefined so that using shorthands is safe. With `safe=bib` only `\nocite`, `\bibcite` 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

**New 3.34**, in  $\epsilon\TeX$  based engines (ie, almost every engine except the oldest ones) shorthands can be used in these macros (formerly you could not).

**math=** active | normal

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  $\{a'\}$  (a closing brace after a shorthand) are not a source of trouble anymore.

**config=**  $\langle file \rangle$

Load  $\langle file \rangle$ .`cfg` instead of the default config file `bblopts.cfg` (the file is loaded even with `noconfigs`).

<b>main=</b>	<code>&lt;language&gt;</code> 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.
<b>headfoot=</b>	<code>&lt;language&gt;</code> 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.
<b>noconfigs</b>	Global and language default config files are not loaded, so you can make sure your document is not spoilt by an unexpected .cfg file. However, if the key config is set, this file is loaded.
<b>showlanguages</b>	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.
<b>nocase</b>	<b>New 3.9l</b> Language settings for uppercase and lowercase mapping (as set by <code>\SetCase</code> ) are ignored. Use only if there are incompatibilities with other packages.
<b>silent</b>	<b>New 3.9l</b> No warnings and no <i>infos</i> are written to the log file. <sup>9</sup>
<b>strings=</b>	<code>generic   unicode   encoded   &lt;label&gt;   &lt;font encoding&gt;</code> Selects the encoding of strings in languages supporting this feature. Predefined labels are <code>generic</code> (for traditional $\TeX$ , L $\text{\AA}$ CR and ASCII strings), <code>unicode</code> (for engines like xetex and luatex) and <code>encoded</code> (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 <code>\MakeUppercase</code> and the like (this feature misuses some internal $\text{\LaTeX}$ tools, so use it only as a last resort).
<b>hyphenmap=</b>	<code>off   first   select   other   other*</code> <b>New 3.9g</b> Sets the behavior of case mapping for hyphenation, provided the language defines it. <sup>10</sup> It can take the following values:  <b>off</b> deactivates this feature and no case mapping is applied; <b>first</b> sets it at the first switching commands in the current or parent scope (typically, when the aux file is first read and at <code>\begin{document}</code> }, but also the first <code>\selectlanguage</code> in the preamble), and it's the default if a single language option has been stated; <sup>11</sup> <b>select</b> sets it only at <code>\selectlanguage</code> ; <b>other</b> also sets it at <code>otherlanguage</code> ; <b>other*</b> also sets it at <code>otherlanguage*</code> as well as in heads and foots (if the option <code>headfoot</code> is used) and in auxiliary files (ie, at <code>\select@language</code> ), and it's the default if several language options have been stated. The option <code>first</code> can be regarded as an optimized version of <code>other*</code> for monolingual documents. <sup>12</sup>
<b>bidi=</b>	<code>default   basic   basic-r   bidi-l   bidi-r</code>

<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]

**New 3.14** Selects the bidi algorithm to be used in luatex and xetex. See sec. 1.24.

layout=

**New 3.16** Selects which layout elements are adapted in bidi documents. See sec. 1.24.

## 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 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 `\ldf@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}
```

**WARNING** Currently this option is not compatible with languages loaded on the fly.

## 1.13 ini files

An alternative approach to define a language (or, more precisely, a *locale*) is by means of an `ini` file. Currently babel provides about 200 of these files containing the basic data required for a locale.

`ini` files are not meant only for babel, and they have been devised as a resource for other packages. To easy interoperability between  $\text{T}_{\text{E}}\text{X}$  and other systems, they are identified with the BCP 47 codes as preferred by the Unicode Common Language Data Repository, which was used as source for most of the data provided by these files, too (the main exception being the `\...name` strings).

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 by means of `\babelprovide`. In other words, `\babelprovide` is mainly meant for auxiliary tasks, and as alternative when the `ldf`, for some reason, does work as expected.

---

think it isn't really useful, but who knows.

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

LUATEX/XETEX

```
\documentclass{book}

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

\babelfont{rm}[Renderer=Harfbuzz]{DejaVu Sans}

\begin{document}

\tableofcontents

\chapter{სამზარეულო და სუფრის ტრადიციები}

ქართული ტრადიციული სამზარეულო ერთ-ერთი უმდიდრესია მთელ მსოფლიოში.

\end{document}
```

**New 3.49** Alternatively, you can tell babel to load all or some languages passed as options with \babelprovide and not from the ldf file in a few typical cases. Thus, provide=\* means ‘load the main language with the \babelprovide mechanism instead of the ldf file’ applying the basic features, which in this case means import, main. There are (currently) three options:

- provide=\* is the option just explained, for the main language;
- provide+=\* is the same for additional languages (the main language is still the ldf file);
- provide\*=\* is the same for all languages, ie, main and additional.

**EXAMPLE** The preamble in the previous example can be more compactly written as:

```
\documentclass{book}
\usepackage[georgian, provide=*]{babel}
\babelfont{rm}[Renderer=Harfbuzz]{DejaVu Sans}
```

Or also:

```
\documentclass[georgian]{book}
\usepackage[provide=*]{babel}
\babelfont{rm}[Renderer=Harfbuzz]{DejaVu Sans}
```

**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 follow (which could no longer be valid when you read this manual, if the packages involved have been updated). The Harfbuzz renderer has still some issues, so as a rule of thumb prefer the default renderer, and resort to Harfbuzz only if the former does not work for you. Fortunately, fonts can be loaded twice with different renderers; for example:

```
\babelfont[spanish]{rm}{FreeSerif}
\babelfont[hindi]{rm}[Renderer=Harfbuzz]{FreeSerif}
```

**Arabic** Monolingual documents mostly work in luatex, but it must be fine tuned, particularly graphical elements like `picture`. In xetex babel resorts to the  `bidi`  package, which seems to work.

**Hebrew** Niqqud marks seem to work in both engines, but depending on the font cantillation marks might be misplaced (xetex or luatex with Harfbuzz seems better, but still problematic).

**Devanagari** In luatex and the the default renderer many fonts work, but some others do not, the main issue being the ‘`ra`’. You may need to set explicitly the script to either `deva` or `dev2`, eg:

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

Other Indic scripts are still under development in the default luatex renderer, but should work with `Renderer=Harfbuzz`. They also work with xetex, although unlike with luatex fine tuning the font behavior is not always possible.

**Southeast scripts** Thai works in both luatex and xetex, but line breaking differs (rules can be modified in luatex; they are hard-coded in xetex). Lao seems to work, too, but there are no patterns for the latter in luatex. Khemer clusters are rendered wrongly with the default renderer. The comment about Indic scripts and lualatex also applies here. Some quick patterns can help, with something similar to:

```
\babelprovide[import,hyphenrules=+]{lao}
\babelpatterns[lao]{\lŋ \lɯ \lə \ŋ \ŋ \ŋ} % Random
```

**East Asia scripts** Settings for either Simplified or Traditional should work out of the box, with basic line breaking with any renderer. Although for a few words and shorts 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[japanese]{ltjbook}
\usepackage{babel}
```

**Latin, Greek, Cyrillic** Combining chars with the default luatex font renderer might be wrong; on then other hand, with the Harfbuzz renderer diacritics are stacked correctly, but many hyphenations points are discarded (this bug seems related to kerning, so it depends on the font). With xetex both combining characters and hyphenation work as expected (not quite, but in most cases it works; the problem here are font clusters).

**NOTE** Wikipedia defines a *locale* as follows: “In computing, a locale is a set of parameters that defines the user’s language, region and any special variant preferences that the user wants to see in their user interface. Usually a locale identifier consists of at least a language code and a country/region code.” Babel is moving gradually from the old and fuzzy concept of *language* to the more modern of *locale*. Note each locale is by itself a separate “language”, which explains why there are so many files. This is on purpose, so that possible variants can be created and/or redefined easily.

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

af	Afrikaans <sup>ul</sup>	en-NZ	English <sup>ul</sup>
agq	Aghem	en-US	English <sup>ul</sup>
ak	Akan	en	English <sup>ul</sup>
am	Amharic <sup>ul</sup>	eo	Esperanto <sup>ul</sup>
ar	Arabic <sup>ul</sup>	es-MX	Spanish <sup>ul</sup>
ar-DZ	Arabic <sup>ul</sup>	es	Spanish <sup>ul</sup>
ar-MA	Arabic <sup>ul</sup>	et	Estonian <sup>ul</sup>
ar-SY	Arabic <sup>ul</sup>	eu	Basque <sup>ul</sup>
as	Assamese	ewo	Ewondo
asa	Asu	fa	Persian <sup>ul</sup>
ast	Asturian <sup>ul</sup>	ff	Fulah
az-Cyrl	Azerbaijani	fi	Finnish <sup>ul</sup>
az-Latn	Azerbaijani	fil	Filipino
az	Azerbaijani <sup>ul</sup>	fo	Faroese
bas	Basaa	fr	French <sup>ul</sup>
be	Belarusian <sup>ul</sup>	fr-BE	French <sup>ul</sup>
bem	Bemba	fr-CA	French <sup>ul</sup>
bez	Bena	fr-CH	French <sup>ul</sup>
bg	Bulgarian <sup>ul</sup>	fr-LU	French <sup>ul</sup>
bm	Bambara	fur	Friulian <sup>ul</sup>
bn	Bangla <sup>ul</sup>	fy	Western Frisian
bo	Tibetan <sup>u</sup>	ga	Irish <sup>ul</sup>
brx	Bodo	gd	Scottish Gaelic <sup>ul</sup>
bs-Cyrl	Bosnian	gl	Galician <sup>ul</sup>
bs-Latn	Bosnian <sup>ul</sup>	grc	Ancient Greek <sup>ul</sup>
bs	Bosnian <sup>ul</sup>	gsw	Swiss German
ca	Catalan <sup>ul</sup>	gu	Gujarati
ce	Chechen	guz	Gusii
cgg	Chiga	gv	Manx
chr	Cherokee	ha-GH	Hausa
ckb	Central Kurdish	ha-NE	Hausa <sup>l</sup>
cop	Coptic	ha	Hausa
cs	Czech <sup>ul</sup>	haw	Hawaiian
cu	Church Slavic	he	Hebrew <sup>ul</sup>
cu-Cyrs	Church Slavic	hi	Hindi <sup>u</sup>
cu-Glag	Church Slavic	hr	Croatian <sup>ul</sup>
cy	Welsh <sup>ul</sup>	hsb	Upper Sorbian <sup>ul</sup>
da	Danish <sup>ul</sup>	hu	Hungarian <sup>ul</sup>
dav	Taita	hy	Armenian <sup>u</sup>
de-AT	German <sup>ul</sup>	ia	Interlingua <sup>ul</sup>
de-CH	German <sup>ul</sup>	id	Indonesian <sup>ul</sup>
de	German <sup>ul</sup>	ig	Igbo
dje	Zarma	ii	Sichuan Yi
dsb	Lower Sorbian <sup>ul</sup>	is	Icelandic <sup>ul</sup>
dua	Duala	it	Italian <sup>ul</sup>
dyo	Jola-Fonyi	ja	Japanese
dz	Dzongkha	jgo	Ngomba
ebu	Embu	jmc	Machame
ee	Ewe	ka	Georgian <sup>ul</sup>
el	Greek <sup>ul</sup>	kab	Kabyle
el-polyton	Polytonic Greek <sup>ul</sup>	kam	Kamba
en-AU	English <sup>ul</sup>	kde	Makonde
en-CA	English <sup>ul</sup>	kea	Kabuverdianu
en-GB	English <sup>ul</sup>	khq	Koyra Chiini

ki	Kikuyu	om	Oromo
kk	Kazakh	or	Odia
kkj	Kako	os	Ossetic
kl	Kalaallisut	pa-Arab	Punjabi
kln	Kalenjin	pa-Guru	Punjabi
km	Khmer	pa	Punjabi
kn	Kannada <sup>ul</sup>	pl	Polish <sup>ul</sup>
ko	Korean	pms	Piedmontese <sup>ul</sup>
kok	Konkani	ps	Pashto
ks	Kashmiri	pt-BR	Portuguese <sup>ul</sup>
ksb	Shambala	pt-PT	Portuguese <sup>ul</sup>
ksf	Bafia	pt	Portuguese <sup>ul</sup>
ksh	Colognian	qu	Quechua
kw	Cornish	rm	Romansh <sup>ul</sup>
ky	Kyrgyz	rn	Rundi
lag	Langi	ro	Romanian <sup>ul</sup>
lb	Luxembourgish	rof	Rombo
lg	Ganda	ru	Russian <sup>ul</sup>
lkt	Lakota	rw	Kinyarwanda
ln	Lingala	rwk	Rwa
lo	Lao <sup>ul</sup>	sa-Beng	Sanskrit
lrc	Northern Luri	sa-Deva	Sanskrit
lt	Lithuanian <sup>ul</sup>	sa-Gujr	Sanskrit
lu	Luba-Katanga	sa-Knda	Sanskrit
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>

sw	Swahili	vai	Vai
ta	Tamil <sup>u</sup>	vi	Vietnamese <sup>ul</sup>
te	Telugu <sup>ul</sup>	vun	Vunjo
teo	Teso	wae	Walser
th	Thai <sup>ul</sup>	xog	Soga
ti	Tigrinya	yav	Yangben
tk	Turkmen <sup>ul</sup>	yi	Yiddish
to	Tongan	yo	Yoruba
tr	Turkish <sup>ul</sup>	yue	Cantonese
twq	Tasawaq	zgh	Standard Moroccan Tamazight
tzm	Central Atlas Tamazight	zh-Hans-HK	Chinese
ug	Uyghur	zh-Hans-MO	Chinese
uk	Ukrainian <sup>ul</sup>	zh-Hans-SG	Chinese
ur	Urdu <sup>ul</sup>	zh-Hans	Chinese
uz-Arab	Uzbek	zh-Hant-HK	Chinese
uz-Cyrl	Uzbek	zh-Hant-MO	Chinese
uz-Latn	Uzbek	zh-Hant	Chinese
uz	Uzbek	zh	Chinese
vai-Latn	Vai	zu	Zulu
vai-Vaii	Vai		

---

In some contexts (currently `\babel font`) an ini file may be loaded by its name. Here is the list of the names currently supported. With these languages, `\babel font` 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 `\babel provide` with a valueless `import`.

---

aghem	bambara
akan	basaa
albanian	basque
american	belarusian
amharic	bemba
ancientgreek	bena
arabic	bengali
arabic-algeria	bodo
arabic-DZ	bosnian-cyrillic
arabic-morocco	bosnian-cyrl
arabic-MA	bosnian-latin
arabic-syria	bosnian-latn
arabic-SY	bosnian
armenian	brazilian
assamese	breton
asturian	british
asu	bulgarian
australian	burmese
austrian	canadian
azerbaijani-cyrillic	cantonese
azerbaijani-cyrl	catalan
azerbaijani-latin	centralatlastamazight
azerbaijani-latn	centralkurdish
azerbaijani	chechen
bafia	cherokee



chiga	french-ch
chinese-hans-hk	french-lu
chinese-hans-mo	french-luxembourg
chinese-hans-sg	french-switzerland
chinese-hans	french
chinese-hant-hk	friulian
chinese-hant-mo	fulah
chinese-hant	galician
chinese-simplified-hongkongsarchina	ganda
chinese-simplified-macausarchina	georgian
chinese-simplified-singapore	german-at
chinese-simplified	german-austria
chinese-traditional-hongkongsarchina	german-ch
chinese-traditional-macausarchina	german-switzerland
chinese-traditional	german
chinese	greek
churchslavic	gujarati
churchslavic-cyrs	gusii
churchslavic-oldcyrillic <sup>13</sup>	hausa-gh
churchsslavic-glag	hausa-ghana
churchsslavic-glagolitic	hausa-ne
cognian	hausa-niger
cornish	hausa
croatian	hawaiian
czech	hebrew
danish	hindi
duala	hungarian
dutch	icelandic
dzongkha	igbo
embu	inarisami
english-au	indonesian
english-australia	interlingua
english-ca	irish
english-canada	italian
english-gb	japanese
english-newzealand	jolafonyi
english-nz	kabuverdianu
english-unitedkingdom	kabyle
english-unitedstates	kako
english-us	kalaallisut
english	kalenjin
esperanto	kamba
estonian	kannada
ewe	kashmiri
ewondo	kazakh
faroeese	khmer
filipino	kikuyu
finnish	kinyarwanda
french-be	konkani
french-belgium	korean
french-ca	koyraborosenni
french-canada	koyrachiini

<sup>13</sup>The name in the CLDR is Old Church Slavonic Cyrillic, but it has been shortened for practical reasons.

kwasio	ossetic
kyrgyz	pashto
lakota	persian
langi	piedmontese
lao	polish
latvian	polytonicgreek
lingala	portuguese-br
lithuanian	portuguese-brazil
lowersorbian	portuguese-portugal
lsorbian	portuguese-pt
lubakatanga	portuguese
luo	punjabi-arab
luxembourgish	punjabi-arabic
luyia	punjabi-gurmukhi
macedonian	punjabi-guru
machame	punjabi
makhuwameetto	quechua
makonde	romanian
malagasy	romansh
malay-bn	rombo
malay-brunei	rundi
malay-sg	russian
malay-singapore	rwa
malay	sakha
malayalam	samburu
maltese	samin
manx	sango
marathi	sangu
masai	sanskrit-beng
mazanderani	sanskrit-bengali
meru	sanskrit-deva
meta	sanskrit-devanagari
mexican	sanskrit-gujarati
mongolian	sanskrit-gujr
morisyen	sanskrit-kannada
mundang	sanskrit-knda
nama	sanskrit-malayalam
nepali	sanskrit-mlym
newzealand	sanskrit-telu
ngiemboon	sanskrit-telugu
ngomba	sanskrit
norsk	scottishgaelic
northernluri	sena
northernsami	serbian-cyrillic-bosniaherzegovina
northndebele	serbian-cyrillic-kosovo
norwegianbokmal	serbian-cyrillic-montenegro
norwegiannynorsk	serbian-cyrillic
nswissgerman	serbian-cyrl-ba
nuer	serbian-cyrl-me
nyankole	serbian-cyrl-xk
nynorsk	serbian-cyrl
occitan	serbian-latin-bosniaherzegovina
oriya	serbian-latin-kosovo
oromo	serbian-latin-montenegro

serbian-latin	tigrinya
serbian-latn-ba	tongan
serbian-latn-me	turkish
serbian-latn-xk	turkmen
serbian-latn	ukenglish
serbian	ukrainian
shambala	uppertsorbian
shona	urdu
sichuanyi	usenglish
sinhala	usorbian
slovak	uyghur
slovene	uzbek-arab
slovenian	uzbek-arabic
soga	uzbek-cyrillic
somali	uzbek-cyrl
spanish-mexico	uzbek-latin
spanish-mx	uzbek-latn
spanish	uzbek
standardmoroccantamazight	vai-latin
swahili	vai-latn
swedish	vai-vai
swissgerman	vai-vaii
tachelhit-latin	vai
tachelhit-latn	vietnam
tachelhit-tfng	vietnamese
tachelhit-tifinagh	vunjo
tachelhit	walser
taita	welsh
tamil	westernfrisian
tasawaq	yangben
telugu	yiddish
teso	yoruba
thai	zarma
tibetan	zulu afrikaans

---

### Modifying and adding values to ini files

**New 3.39** There is a way to modify the values of ini files when they get loaded with `\babelprovide` and `import`. To set, say, `digits.native` in the `numbers` section, use something like `numbers/digits.native=abcdefghijkl`. Keys may be added, too. Without `import` you may modify the identification keys. This can be used to create private variants easily. All you need is to import the same ini file with a different locale name and different parameters.

## 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>14</sup>

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

**NOTE** See the note in the previous section about some issues in specific languages.

---

<sup>14</sup>See also the package `combofont` for a complementary approach.

The main purpose of `\babelfont` 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, `\babelfont{rm}{FreeSerif}` defines 4 fonts (with their variants, of course), which are switched with the language by babel. It is a tool to make things easier and transparent to the user.

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 one. Alternatively, you may set a font for a script – just precede its name (lowercase) with a star (eg, `*devanagari`). With this optional argument, the font is *not* yet defined, but just predeclared. This means you may define as many fonts as you want ‘just in case’, because if the language is never selected, the corresponding `\babelfont` declaration is just ignored.

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.

LUATEX/XETEX

```
\documentclass{article}

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

\babelprovide[import]{hebrew}

\babelfont{rm}{FreeSerif}

\begin{document}

Svenska \foreignlanguage{hebrew}{עברית} svenska.

\end{document}
```

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

LUATEX/XETEX

```
\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:

LUATEX/XETEX

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

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

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

LUATEX/XETEX

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

This makes sure the OpenType script for Devanagari is deva and not dev2, in case it is not detected correctly. 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 with `\babelfont` (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 can be problematic, and also preserving 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** Using `\setxxxxfont` and `\babelfont` at the same time is discouraged, but very often works as expected. However, be aware with `\setxxxxfont` the language system will not be set by babel and should be set with `fontspec` if necessary.

**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 can 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 Info: The following fonts are not babel standard families.*

**This is *not* and error.** babel assumes that if you are using `\babelfont` 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.

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

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 `\babelfont` at all. But you must be aware that this may lead to some problems.

## 1.15 Modifying a language

Modifying the behavior of a language (say, the chapter “caption”), is sometimes necessary, but not always trivial. In the case of caption names a specific macro is provided, because this is perhaps the most frequent change:

`\setlocalecaption`  $\{\langle language-name \rangle\}\{\langle caption-name \rangle\}\{\langle string \rangle\}$

**New 3.51** Here *caption-name* is the name as string without the trailing name. An example, which also shows caption names are often a stylistic choice, is:

```
\setlocalecaption{english}{contents}{Table of Contents}
```

This works not only with existing caption names, because it also serves to define new ones by setting the *caption-name* to the name of your choice (name will be postpended). Captions so defined or redefined behave with the ‘new way’ described in the following note.

**NOTE** There are a few alternative methods:

- With data import’ed from ini files, you can modify the values of specific keys, like:

```
\babelprovide[import, captions/listtable = Lista de tablas]{spanish}
```

(In this particular case, instead of the captions group you may need to modify the captions.licr one.)

- 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 so. This redefinition is not activated until the language is selected.

- 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 \babelprovide and its key import, is:

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

This redefinition is immediate.

**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.

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

```
\addto\extrarussian{\mymacro}
```

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

**NOTE** These macros (`\captions<lang>`, `\extras<lang>`) 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. Without the optional argument it just loads some additional tools if provided by the ini file, like extra counters.

## 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` [*options*] {*language-name*}

If the language *language-name* has not been loaded as class or package option and there are no *options*, 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`, *language-name* is still relevant because in such a case the hyphenation and like breaking 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: \chaptername not set for 'mylang'. Please,
(babel)                define it after the language has been loaded
(babel)                (typically in the preamble) with:
(babel)                \setlocalecaption{mylang}{chapter}{..}
(babel)                Reported on input line 26.
```

In most cases, you will only need to define a few macros. Note languages loaded on the fly are not yet available in the preamble.

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

```
\usepackage[danish]{babel}
\babelprovide{arhinish}
\setlocalecaption{arhinish}{chapter}{Chapitula}
\setlocalecaption{arhinish}{refname}{Refirenke}
\renewcommand\arhinishhyphenmins{22}
```

**EXAMPLE** Locales with names based on BCP 47 codes can be created with something like:

```
\babelprovide[import=en-US]{enUS}
```

Note, however, mixing ways to identify locales can lead to problems. For example, is `yi` the name of the language spoken by the Yi people or is it the code for Yiddish?

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 and date (also line breaking rules in newly defined languages). 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 can be written:

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

There are about 250 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 may show a warning about the current lack of suitability of some features.

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}`. **New 3.44** More convenient is usually `\localedate`, which prints the date for the current locale.

**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 spanish 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  $\text{\TeX}$  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 (thus overriding that set when babel is loaded). Only in newly defined languages.

**EXAMPLE** Let's assume your document is mainly in Polytonic Greek, but with some sections in Italian. Then, the first attempt should be:

```
\usepackage[italian, greek.polutonic]{babel}
```

But if, say, accents in Greek are not shown correctly, you can try:

```
\usepackage[italian]{babel}
\babelprovide[import, main]{polytonicgreek}
```

Remember there is an alternative syntax for the latter:

```
\usepackage[italian, polytonicgreek, provide=*]{babel}
```

**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.

**alph=**  $\langle counter-name \rangle$

Assigns to `\alph` that counter. See the next section.

**Alph=**  $\langle counter-name \rangle$

Same for `\Alph`.

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.

**onchar=** ids | fonts

**New 3.38** This option is much like an ‘event’ called when a character belonging to the script of this locale is found (as its name implies, it acts on characters, not on spaces). There are currently two ‘actions’, which can be used at the same time (separated by a space): with `ids` the `\language` and the `\localeid` are set to the values of this locale; with `fonts`, the fonts are changed to those of this locale (as set with `\babelfont`). This option is not compatible with `mapfont`. Characters can be added or modified with `\babelcharproperty`.

**NOTE** An alternative approach with `luatex` and `Harfbuzz` is the `font` option `RawFeature={multiscript=auto}`. It does not switch the `babel` language and therefore the line breaking rules, but in many cases it can be enough.

**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).

**mapfont=** direction

Assigns the font for the writing direction of this language (only with `bidi=basic`). Whenever possible, instead of this option use `onchar`, based on the script, which usually makes more sense. 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. So, there should be at most 3 directives of this kind.

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

## 1.17 Digits and counters

**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 `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:

Arabic	Persian	Lao	Odia	Urdu
Assamese	Gujarati	Northern Luri	Punjabi	Uzbek
Bangla	Hindi	Malayalam	Pashto	Vai
Tibetar	Khmer	Marathi	Tamil	Cantonese
Bodo	Kannada	Burmese	Telugu	Chinese
Central Kurdish	Konkani	Mazanderani	Thai	
Dzongkha	Kashmiri	Nepali	Uyghur	

**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  $\TeX$  code). This means the local digits have the correct bidirectional behavior (unlike `Numbers=Arabic` in `fontspec`, which is not recommended).

**NOTE** With `xetex` you can use the option `Mapping` when defining a font.

**New 4.41** Many ‘ini’ locale files has been extended with information about non-positional numerical systems, based on those predefined in CSS. They only work with `xetex` and `luatex` and are fully expendable (even inside an unprotected `\edef`). Currently, they are limited to numbers below 10000. There are several ways to use them (for the available styles in each language, see the list below):

- `\localnumeral{<style>}{<number>}`, like `\localnumeral{abjad}{15}`
- `\localecounter{<style>}{<counter>}`, like `\localecounter{lower}{section}`
- In `\babelprovide`, as an argument to the keys `alph` and `Alph`, which redefine what `\alph` and `\Alph` print. For example:

```
\babelprovide[alph=alphabetic]{thai}
```

The styles are:

**Ancient Greek** lower.ancient, upper.ancient  
**Amharic** afar, agaw, ari, blin, dizi, gedeo, gumuz, hadiyya, harari, kaffa, kebena, kembata, konso, kunama, meen, oromo, saho, sidama, silti, tigre, wolaita, yemsa  
**Arabic** abjad, maghrebi.abjad  
**Belarusan, Bulgarian, Macedonian, Serbian** lower, upper  
**Bengali** alphabetic  
**Coptic** epact, lower.letters  
**Hebrew** letters (neither geresh nor gershayim yet)  
**Hindi** alphabetic  
**Armenian** lower.letter, upper.letter  
**Japanese** hiragana, hiragana.iroha, katakana, katakana.iroha, circled.katakana, informal, formal, cjk-earthly-branch, cjk-heavenly-stem, fullwidth.lower.alpha, fullwidth.upper.alpha  
**Georgian** letters

**Greek** lower.modern, upper.modern, lower.ancient, upper.ancient (all with keraia)  
**Khmer** consonant  
**Korean** consonant, syllable, hanja.informal, hanja.formal, hangul.formal,  
 cjk-earthly-branch, cjk-heavenly-stem, fullwidth.lower.alpha,  
 fullwidth.upper.alpha  
**Marathi** alphabetic  
**Persian** abjad, alphabetic  
**Russian** lower, lower.full, upper, upper.full  
**Syriac** letters  
**Tamil** ancient  
**Thai** alphabetic  
**Ukrainian** lower, lower.full, upper, upper.full  
**Chinese** cjk-earthly-branch, cjk-heavenly-stem, fullwidth.lower.alpha,  
 fullwidth.upper.alpha

**New 3.45** In addition, native digits (in languages defining them) may be printed with the numeral style digits.

## 1.18 Dates

**New 3.45** When the data is taken from an ini file, you may print the date corresponding to the Gregorian calendar and other lunisolar systems with the following command.

**\localedate** [ $\langle\text{calendar}=\dots, \text{variant}=\dots\rangle$ ]{ $\langle\text{year}\rangle$ }{ $\langle\text{month}\rangle$ }{ $\langle\text{day}\rangle$ }

By default the calendar is the Gregorian, but a ini files may define strings for other calendars (currently ar, ar-\*, he, fa, hi.) In the latter case, the three arguments are the year, the month, and the day in those in the corresponding calendar. They are *not* the Gregorian data to be converted (which means, say, 13 is a valid month number with calendar=hebrew).

Even with a certain calendar there may be variants. In Kurmanji the default variant prints something like *30. Çiley a Pêşîn 2019*, but with variant=iza fa it prints *31'ê Çiley a Pêşînê 2019*.

## 1.19 Accessing language info

**\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** { $\langle\text{language}\rangle$ }{ $\langle\text{true}\rangle$ }{ $\langle\text{false}\rangle$ }


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  $\TeX$ 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.

**\localeinfo** { $\langle\text{field}\rangle$ }

**New 3.38** If an ini file has been loaded for the current language, you may access the information stored in it. This macro is fully expandable, and the available fields are:

`name.english` as provided by the Unicode CLDR.  
`tag.ini` is the tag of the ini file (the way this file is identified in its name).  
`tag.bcp47` is the full BCP 47 tag (see the warning below).  
`language.tag.bcp47` is the BCP 47 language tag.  
`tag.opentype` is the tag used by OpenType (usually, but not always, the same as BCP 47).  
`script.name`, as provided by the Unicode CLDR.  
`script.tag.bcp47` is the BCP 47 tag of the script used by this locale.  
`script.tag.opentype` is the tag used by OpenType (usually, but not always, the same as BCP 47).

**WARNING** **New 3.46** As of version 3.46 `tag.bcp47` returns the full BCP 47 tag. Formerly it returned just the language subtag, which was clearly counterintuitive.

**`\getlocaleproperty`**  `{\macro}{\locale}{\property}`

**New 3.42** The value of any locale property as set by the ini files (or added/modified with `\babelprovide`) can be retrieved and stored in a macro with this command. For example, after:

```
\getlocaleproperty\hechap{hebrew}{captions/chapter}
```

the macro `\hechap` will contain the string פרק.

If the key does not exist, the macro is set to `\relax` and an error is raised. **New 3.47** With the starred version no error is raised, so that you can take your own actions with undefined properties.

Babel remembers which ini files have been loaded. There is a loop named `\LocaleForEach` to traverse the list, where #1 is the name of the current item, so that `\LocaleForEach{\message{ **#1** }}` just shows the loaded ini's.

**NOTE** ini files are loaded with `\babelprovide` and also when languages are selected if there is a `\babelfont`. To ensure the ini files are loaded (and therefore the corresponding data) even if these two conditions are not met, write `\BabelEnsureInfo` in the preamble.

**`\localeid`**

Each language in the babel sense has its own unique numeric identifier, which can be retrieved with `\localeid`.

**NOTE** The `\localeid` is not the same as the `\language` identifier, which refers to a set of hyphenation patterns (which, in turn, is just a component of the line breaking algorithm described in the next section). The data about preloaded patterns are stored in an internal macro named `\bbl@languages` (see the code for further details), but note several locales may share a single `\language`, so they are separated concepts. In `luatex`, the `\localeid` is saved in each node (where it makes sense) as an attribute, too.

## 1.20 Hyphenation and line breaking

Babel deals with three kinds of line breaking rules: Western, typically the LGC group, South East Asian, like Thai, and CJK, but support depends on the engine: `pdfTeX` only deals with the former, `xetex` also with the second one (although in a limited way), while `luatex` provides basic rules for the latter, too.

**`\babelhyphen`**  `{\type}`

`\babelhyphen` `*{<text>}`

**New 3.9a** It is customary to classify hyphens in two types: (1) *explicit* or *hard hyphens*, which in  $\TeX$  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  $\TeX$  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  $\LaTeX$ : (1) the character used is that set for the current font, while in  $\LaTeX$  it is hardwired to `-` (a typical value); (2) the hyphen to be used in fonts with a negative `\hyphenchar` is `-`, like in  $\LaTeX$ , 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.

**NOTE** To set hyphenation exceptions in the preamble before any language is explicitly set with a selector, use `\babelhyphenation` instead of `\hyphenation`. In the preamble the hyphenation rules are not always fully set up and an error can be raised.

`\begin{hyphenrules}`  $\{ \langle \textit{language} \rangle \} \dots \text{ } \code{\end{hyphenrules}}$

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 deprecated and `otherlanguage*` (the starred version) is preferred, because the former does not take into account possible changes in encodings of characters like, say, ‘`’`’ done by some languages (eg, *italian*, *french*, *ukraineb*).

`\babelpatterns`  $[ \langle \textit{language} \rangle , \langle \textit{language} \rangle , \dots ] \{ \langle \textit{patterns} \rangle \}$

**New 3.9m** *In *luatex* only,*<sup>15</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*).

## 1.21 Transforms

There are a couple of macros to define *transforms*.<sup>16</sup>

`\babelposthyphenation`  $\{ \langle \textit{hyphenrules-name} \rangle \} \{ \langle \textit{lua-pattern} \rangle \} \{ \langle \textit{replacement} \rangle \}$

**New 3.37-3.39** *With *luatex** it is now possible to define non-standard hyphenation rules, like `f-f → ff-f`, repeated hyphens, ranked ruled (or more precisely, ‘penalized’ hyphenation points), and so on. No rules are currently provided by default, but they can be defined as shown in the following example, where `{1}` is the first captured char (between `()` in the pattern):

<sup>15</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>16</sup>Similar in concept, but not the same, as those in Unicode.

```

\babelposthyphenation{german}{([fmtrp]) | {1}}
{
  { no = {1}, pre = {1}{1}- }, % Replace first char with disc
  remove,                      % Remove automatic disc (2nd node)
  {}                          % Keep last char, untouched
}

```

In the replacements, a captured char may be mapped to another, too. For example, if the first capture reads (`[îú]`), the replacement could be `{1|îú|íú}`, which maps `î` to `í`, and `ú` to `û`, so that the diaeresis is removed.

This feature is activated with the first `\babelposthyphenation` or `\babelprehyphenation`. See the [babel site](#) for a more detailed description and some examples. It also describes a few additional replacement types (string, penalty).

Although the main purpose of this command is non-standard hyphenation, it may actually be used for other transformations (after hyphenation is applied, so you must take discretionaries into account).

You are limited to substitutions as done by lua, although a future implementation may alternatively accept lpeg.

**`\babelprehyphenation`** `{<locale-name>}{<lua-pattern>}{<replacement>}`

**New 3.44-3-52** It is similar to the latter, but (as its name implies) applied before hyphenation. There are other differences: (1) the first argument is the locale instead the name of hyphenation patterns; (2) in the search patterns `=` has no special meaning, while `|` stands for an ordinary space; (3) in the replacement, discretionaries are not accepted. It handles glyphs and spaces.

This feature is activated with the first `\babelposthyphenation` or `\babelprehyphenation`.

**EXAMPLE** You can replace a character (or series of them) by another character (or series of them). Thus, to enter `ž` as `zh` and `š` as `sh` in a newly created locale for transliterated Russian:

```

\babelprovide[hyphenrules=+]{russian-latin} % Create locale
\babelprehyphenation{russian-latin}{([sz])h} % Create rule
{
  string = {1|sz|šž},
  remove
}

```

**EXAMPLE** The following rule prevent the word “a” from being at the end of a line:

```

\babelprehyphenation{english}{|a|}
{, {}, % Keep first space and a
{insert, penalty = 10000}, % Insert penalty
{} % Keep last space
}

```

## 1.22 Selection based on BCP 47 tags

**New 3.43** The recommended way to select languages is that described at the beginning of this document. However, BCP 47 tags are becoming customary, particularly in documents (or parts of documents) generated by external sources, and therefore babel will provide a set of tools to select the locales in different situations, adapted to the particular needs of



each case. Currently, babel provides autoloading of locales as described in this section. In these contexts autoloading is particularly important because we may not know on beforehand which languages will be requested.

It must be activated explicitly, because it is primarily meant for special tasks. Mapping from BCP 47 codes to locale names are not hardcoded in babel. Instead the data is taken from the ini files, which means currently about 250 tags are already recognized. Babel performs a simple lookup in the following way:  $\text{fr-Latn-FR} \rightarrow \text{fr-Latn} \rightarrow \text{fr-FR} \rightarrow \text{fr}$ . Languages with the same resolved name are considered the same. Case is normalized before, so that  $\text{fr-latn-fr} \rightarrow \text{fr-Latn-FR}$ . If a tag and a name overlap, the tag takes precedence.

Here is a minimal example:

```
\documentclass{article}

\usepackage[danish]{babel}

\babeladjust{
  autoload.bcp47 = on,
  autoload.bcp47.options = import
}

\begin{document}

Chapter in Danish: \chaptername.

\selectlanguage{de-AT}

\localedate{2020}{1}{30}

\end{document}
```

Currently the locales loaded are based on the ini files and decoupled from the main ldf files. This is by design, to ensure code generated externally produces the same result regardless of the languages requested in the document, but an option to use the ldf instead will be added in a future release, because both options make sense depending on the particular needs of each document (there will be some restrictions, however).

The behaviour is adjusted with `\babeladjust` with the following parameters:

`autoload.bcp47` with values on and off.

`autoload.bcp47.options`, which are passed to `\babelprovide`; empty by default, but you may add import (features defined in the corresponding `babel-...tex` file might not be available).

`autoload.bcp47.prefix`. Although the public name used in selectors is the tag, the internal name will be different and generated by prepending a prefix, which by default is `bcp47-`. You may change it with this key.

**New 3.46** If an ldf file has been loaded, you can enable the corresponding language tags as selector names with:

```
\babeladjust{ bcp47.toname = on }
```

(You can deactivate it with off.) So, if dutch is one of the package (or class) options, you can write `\selectlanguage{nl}`. Note the language name does not change (in this example is still dutch), but you can get it with `\localeinfo` or `\getlanguageproperty`. It must be turned on explicitly for similar reasons to those explained above.

## 1.23 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>17</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 is 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>18</sup>

`\ensureascii`  $\{\langle text \rangle\}$

**New 3.9i** This macro makes sure  $\langle text \rangle$  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).

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.

## 1.24 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 can 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 can 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).

**WARNING** If characters to be mirrored are shown without changes with `luatex`, try with the following line:

<sup>17</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>18</sup>But still defined for backwards compatibility.

```
\babeladjust{bidi.mirroring=off}
```

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.

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 is available in luatex only.

```
\documentclass{article}

\usepackage[bidi=basic]{babel}

\babelprovide[import, main]{arabic}

\babelfont{rm}{FreeSerif}

\begin{document}

    وقد عرفت شبه جزيرة العرب طيلة العصر الهيليني (الاجريقي) بـ
    Arabia أو Aravia (بالاغريقية Αραβία)، استخدم الرومان ثلاث
    بادئات 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[onchar=ids fonts]{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 \textit{fuṣḥā l-‘aṣr} (MSA) and
\textit{fuṣḥā t-turāth} (CA).

\end{document}

```

In this example, and thanks to `onchar=ids` fonts, 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`.`\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 can 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>19</sup>

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

<sup>19</sup>Next on the roadmap are counters and numeral systems in general. Expect some minor readjustments.

**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`, so that the first column is the right one (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. 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:

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

**\babelsublr** `{\lr-text}`

Digits in `pdftex` must be marked up explicitly (unlike `luatex` with `bidi=basic` or `bidi=basic-r` and, usually, `xetex`). This command is provided to set `{\lr-text}` 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-name \rangle}`

Mainly for bidi text, but it can 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-language \rangle}{\langle before \rangle}{\langle after \rangle}`

**New 3.17** Something like:

```
\BabelFootnote{\parsfootnote}{\language name}{\{ \}}
```

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

```
\footnote{(\foreignlanguage{\language name}{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 name}{\{ \}}%
\BabelFootnote{\localfootnote}{\language name}{\{ \}}%
\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.

## 1.25 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).

## 1.26 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 `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 just defines a few basic commands. It can be used to define different versions of them or to load a file.

**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).

## 1.27 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** indonesian (bahasa, indon, bahasai)

**Interlingua** interlingua

**Irish Gaelic** irish

**Italian** italian

**Latin** latin

**Lower Sorbian** lowersorbian

**Malay** malay, melayu (bahasam)

**North Sami** samin

**Norwegian** norsk, nynorsk

**Polish** polish

**Portuguese** portuguese, brazilian (portuges, brazil)<sup>20</sup>

<sup>20</sup>The two last name comes from the times when they had to be shortened to 8 characters



**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  $\LaTeX$ .

## 1.28 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).

**$\backslash\text{babelcharproperty}$**   $\{\langle char-code \rangle\}[\langle to-char-code \rangle]\{\langle property \rangle\}\{\langle value \rangle\}$

**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, and this command is allowed only in vertical mode (the preamble or between paragraphs). 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

```

**New 3.39** Another property is locale, which adds characters to the list used by onchar in  $\backslash\text{babelprovide}$ , or, if the last argument is empty, removes them. The last argument is the locale name:

```
\babelcharproperty{`,`}{locale}{english}
```

## 1.29 Tweaking some features

`\babeladjust` `{<key-value-list>}`

**New 3.36** Sometimes you might need to disable some babel features. Currently this macro understands the following keys (and only for 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 `luahtex` you may need `bidi.mirroring=off`. Use with care, because these options do not deactivate other related options (like paragraph direction with `bidi.text`).

## 1.30 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`),  $\TeX$  will keep complaining about an undefined label. To prevent such problems, you can 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 `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 (`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}}
```

- For the hyphenation to work correctly, `lccodes` cannot change, because  $\TeX$  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  $\TeX$ , not of `babel`. Alternatively, you may use `\usesorthands` to activate `'` and `\defineshortand`, 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.

<sup>21</sup>This explains why  $\TeX$  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.

- 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  $\TeX$  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.

### 1.31 Current and future work

The current work is focused on the so-called complex scripts in  $\text{luatex}$ . 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  $\text{\TeX}$  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. An option to manage bidirectional document layout in  $\text{luatex}$  (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).

### 1.32 Tentative and experimental code

See the code section for `\foreignlanguage*` (a new starred version of `\foreignlanguage`). For old an deprecated functions, see the wiki.

#### Options for locales loaded on the fly

**New 3.51** `\babeladjust{ autoload.options = ... }` sets the options when a language is loaded on the fly (by default, no options). A typical value would be `import`, which

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

defines captions, date, numerals, etc., but ignores the code in the tex file (for example, extended numerals in Greek).

### Labels

**New 3.48** There is some work in progress for babel to deal with labels, both with the relation to captions (chapters, part), and how counters are used to define them. It is still somewhat tentative because it is far from trivial – see the wiki for further details.

## 2 Loading languages with language.dat

T<sub>E</sub>X and most engines based on it (pdfT<sub>E</sub>X, xetex,  $\epsilon$ -T<sub>E</sub>X, the main exception being luatex) require hyphenation patterns to be preloaded when a format is created (eg, L<sup>A</sup>T<sub>E</sub>X, XeL<sup>A</sup>T<sub>E</sub>X, pdfL<sup>A</sup>T<sub>E</sub>X). 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>

### 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 iniTeX 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
```

<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.

<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.

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 can 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  $\text{\TeX}$  users, so the files have to be coded so that they can be read by both  $\text{\LaTeX}$  and plain  $\text{\TeX}$ . The current format can be checked by looking at the value of the macro `\fmtname`.
- 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 `\<lang>hyphenmins`, `\captions<lang>`, `\date<lang>`, `\extras<lang>` and `\noextras<lang>` (the last two may be left empty); where `<lang>` is either the name of the language definition file or the name of the  $\text{\LaTeX}$  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<lang>` but not `\captions<lang>` does not raise an error but can lead to unexpected results.
- When a language definition file is loaded, it can define `\l@<lang>` to be a dialect of `\language0` when `\l@<lang>` 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<lang>` 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<lang>`.
- 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.

### 3.1 Guidelines for contributed languages

Currently, the easiest way to contribute a new language is by taking one of the 500 or so `ini` templates available on GitHub as a basis. Just make a pull request or download it and then, after filling the fields, send it to me. Feel free to ask for help or to make feature requests.

As to `ldf` files, 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`, `otf`, `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.

---

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

The following page provides a starting point for ldf files:

<http://www.texnia.com/incubator.html>. See also

<https://github.com/latex3/babel/blob/master/news-guides/guides/list-of-locale-templates.md>.

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. Here “language” is used in the T<sub>E</sub>X 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 T<sub>E</sub>X 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.

**\noextras<lang>** Because we want to let the user switch between languages, but we do not know what state T<sub>E</sub>X might be in after the execution of `\extras<lang>`, a macro that brings T<sub>E</sub>X into a predefined state is needed. It will be no surprise that the name of this macro is `\noextras<lang>`.

**\bbl@declare@tribute** 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.

**\main@language** To postpone the activation of the definitions needed for a language until the beginning of a document, all language definition files should use `\main@language` instead of `\selectlanguage`. This will just store the name of the language, and the proper language will be activated at the start of the document.

**\ProvidesLanguage** The macro `\ProvidesLanguage` should be used to identify the language definition files. Its syntax is similar to the syntax of the L<sup>A</sup>T<sub>E</sub>X command `\ProvidesPackage`.

**\LdfInit** The macro `\LdfInit` 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 `@`-sign, preventing the `.ldf` 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, $\TeX$ can be instructed to load a local configuration file. This file can, for instance, be used to add strings to <code>\captions{lang}</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 $\TeX$ 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>}
    [2016/04/23 v0.0 <Language> support from the babel system]
\LdfInit{<language>}{captions<language>}

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

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

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

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

\StartBabelCommands*{<language>}{captions}
\SetString\chaptername{<chapter name>}
% More strings

\StartBabelCommands*{<language>}{date}
\SetString\monthinname{<name of first month>}
% More strings

\StartBabelCommands*{<dialect>}{captions}
\SetString\chaptername{<chapter name>}
% More strings

\StartBabelCommands*{<dialect>}{date}
\SetString\monthinname{<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  $\TeX$ book states: “Plain  $\TeX$  includes a macro called `\dospecials` that is essentially a set macro, representing the set of all characters that have a special category code.” [4, 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 `\dospecial`.  $\TeX$  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,

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

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 can 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 `TeX` 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@sf@q`** Sometimes it is necessary to preserve the `\spacefactor`. For this purpose the macro `\save@sf@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.

### 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 consist is 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`  $\{\langle language-list \rangle\}\{\langle category \rangle\}[\langle selector \rangle]$

The  $\langle language-list \rangle$  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 a encoded way).

The  $\langle category \rangle$  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]
\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\monthiiiname{März}
```

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

```

\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).

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

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>

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

**$\backslash AfterBabelCommands$**   $\{ \langle code \rangle \}$

The code is delayed and executed at the global scope just after  $\backslash EndBabelCommands$ .

**$\backslash SetString$**   $\{ \langle macro-name \rangle \} \{ \langle string \rangle \}$

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.

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

**\SetStringLoop** {<macro-name>}{<string-list>}

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** [*<map-list>*]{<toupper-code>}{<tolower-code>}

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 *<map-list>* 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 can 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}

\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 though the given uppercase codes, using the step, and assigns them the lccode, which is also increased (MM stands for *many-to-many*).

- `\BabelLowerMO{⟨ucode-from⟩}{⟨ucode-to⟩}{⟨step⟩}{⟨lcode⟩}` loops through the given uppercase codes, using the step, and assigns them the lcode, 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}{{"11F}{2}{{"101}}
```

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`name. 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 can happen. It worked incorrectly with `^` (if activated) and also if deactivated.
- Active chars were not reset at the end of language options, and that led 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 following description is no longer valid, because switch and plain have been merged into babel.def.**

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  $\LaTeX$  package, which sets options and loads language styles.

**plain.def** defines some  $\LaTeX$  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.

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 an 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 (which start always with a lowercase case). There is an exception, however: the section counters has been devised to have arbitrary keys, so you can add lowercased keys if you want.

## 7 Tools

```
1 <<version=3.55.2319>>
2 <<date=2021/03/22>>
```

**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@cl#1{\csname bbl@#1\language\endcsname}
14 \def\bbl@loop#1#2#3{\bbl@loop#1{#3}#2,\@nnil,}
15 \def\bbl@loopx#1#2{\expandafter\bbl@loop\expandafter#1\expandafter{#2}}
16 \def\bbl@loop#1#2#3,{%
17   \ifx\@nnil#3\relax\else
18     \def#1{#3}#2\bbl@afterfi\bbl@loop#1{#2}%
19   \fi}
20 \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.

```

21 \def\bbl@add@list#1#2{%
22   \edef#1{%
23     \bbl@ifunset{\bbl@stripslash#1}%
24     {}%
25     {\ifx#1@empty\else#1,\fi}%
26     #2}%

```

**\bbl@afterelse** **\bbl@afterfi** 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.

```

27 \long\def\bbl@afterelse#1\else#2\fi{\fi#1}
28 \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.

```

29 \def\bbl@exp#1{%
30   \begingroup
31   \let\>\noexpand
32   \def\<##1>{\expandafter\>\noexpand\csname##1\endcsname}%
33   \edef\bbl@exp@aux{\endgroup#1}%
34   \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.

```

35 \def\bbl@tempa#1{%
36   \long\def\bbl@trim##1##2{%
37     \futurelet\bbl@trim@a\bbl@trim@c##2\@nil\@nil#1\@nil\relax{##1}}%
38   \def\bbl@trim@c{%
39     \ifx\bbl@trim@a\@sptoken

```

<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.



```

40 \expandafter\bb1@trim@b
41 \else
42 \expandafter\bb1@trim@b\expandafter#1%
43 \fi}%
44 \long\def\bb1@trim@b#1##1 \nil{\bb1@trim@i##1}}
45 \bb1@tempa{ }
46 \long\def\bb1@trim@i#1\@nil#2\relax#3{#3{#1}}
47 \long\def\bb1@trim@def#1{\bb1@trim{\def#1}}

```

**\bb1@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.

```

48 \begingroup
49 \gdef\bb1@ifunset#1{%
50 \expandafter\ifx\csname#1\endcsname\relax
51 \expandafter\@firstoftwo
52 \else
53 \expandafter\@secondoftwo
54 \fi}
55 \bb1@ifunset{ifcsname}%
56 {}%
57 {\gdef\bb1@ifunset#1{%
58 \ifcsname#1\endcsname
59 \expandafter\ifx\csname#1\endcsname\relax
60 \bb1@afterelse\expandafter\@firstoftwo
61 \else
62 \bb1@afterfi\expandafter\@secondoftwo
63 \fi
64 \else
65 \expandafter\@firstoftwo
66 \fi}}
67 \endgroup

```

**\bb1@ifblank** A tool from url, by Donald Arseneau, which tests if a string is empty or space. The companion macros tests if a macro is defined with some ‘real’ value, ie, not `\relax` and not empty,

```

68 \def\bb1@ifblank#1{%
69 \bb1@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
70 \long\def\bb1@ifblank@i#1#2\@nil#3#4#5\@nil{#4}
71 \def\bb1@ifset#1#2#3{%
72 \bb1@ifunset{#1}{#3}{\bb1@exp{\bb1@ifblank{#1}}{#3}{#2}}}

```

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).

```

73 \def\bb1@forkv#1#2{%
74 \def\bb1@kvcmd##1##2##3{#2}%
75 \bb1@kvnext#1,\@nil,}
76 \def\bb1@kvnext#1,{%
77 \ifx\@nil#1\relax\else
78 \bb1@ifblank{#1}{\bb1@forkv@eq#1=\@empty=\@nil{#1}}%
79 \expandafter\bb1@kvnext
80 \fi}
81 \def\bb1@forkv@eq#1=#2=#3\@nil#4{%
82 \bb1@trim@def\bb1@forkv@a{#1}%
83 \bb1@trim{\expandafter\bb1@kvcmd\expandafter{\bb1@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).

```

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

```

\bbl@replace

```

93 \def\bbl@replace#1#2#3{% in #1 -> repl #2 by #3
94   \toks@{}%
95   \def\bbl@replace@aux##1#2##2#2{%
96     \ifx\bbl@nil##2%
97       \toks@\expandafter{\the\toks@##1}%
98     \else
99       \toks@\expandafter{\the\toks@##1#3}%
100     \bbl@afterfi
101     \bbl@replace@aux##2#2%
102   \fi}%
103   \expandafter\bbl@replace@aux#1#2\bbl@nil#2%
104   \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 are retokenized). It may change! (or even merged with \bbl@replace; I'm not sure ckecking the replacement is really necessary or just paranoia).

```

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

```

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  $\text{\LaTeX}$ , and 2 is  $\text{\XeTeX}$ . You may use the latter in your language style if you want.

```

132 \def\bbl@ifsamestring#1#2{%
133   \begingroup
134     \protected@edef\bbl@tempb{#1}%
135     \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
136     \protected@edef\bbl@tempc{#2}%
137     \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
138     \ifx\bbl@tempb\bbl@tempc
139       \aftergroup\@firstoftwo
140     \else
141       \aftergroup\@secondoftwo
142     \fi
143   \endgroup}
144 \chardef\bbl@engine=%
145 \ifx\directlua\@undefined
146   \ifx\XeTeXinputencoding\@undefined
147     \z@
148   \else
149     \tw@
150   \fi
151 \else
152   \@ne
153 \fi

```

A somewhat hackish tool (hence its name) to avoid spurious spaces in some contexts.

```

154 \def\bbl@bsphack{%
155   \ifhmode
156     \hskip\z@skip
157     \def\bbl@esphack{\loop\ifdim\lastskip>\z@\unskip\repeat\unskip}%
158   \else
159     \let\bbl@esphack\@empty
160   \fi}

```

Another hackish tool, to apply case changes inside a protected macros. It's based on the internal `\let's` made by `\MakeUppercase` and `\MakeLowercase` between things like `\oe` and `\OE`.

```

161 \def\bbl@cased{%
162   \ifx\oe\OE
163     \expandafter\in@\expandafter
164       {\expandafter\OE\expandafter}\expandafter{\oe}%
165     \ifin@
166       \bbl@afterelse\expandafter\MakeUppercase
167     \else
168       \bbl@afterfi\expandafter\MakeLowercase
169     \fi
170   \else
171     \expandafter\@firstofone
172   \fi}
173 <</Basic macros>>

```

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

```

174 <<*Make sure ProvidesFile is defined>> ≡
175 \ifx\ProvidesFile\@undefined
176   \def\ProvidesFile#1[#2 #3 #4]{%
177     \wlog{File: #1 #4 #3 <#2>}%
178     \let\ProvidesFile\@undefined}
179 \fi
180 <</Make sure ProvidesFile is defined>>

```

## 7.1 Multiple languages

`\language` Plain TeX 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.

```
181 <<*Define core switching macros>> ≡
182 <<ifx\language\undefined
183   \csname newcount\endcsname\language
184 \fi
185 <</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` This macro was introduced for TeX < 2. Preserved for compatibility.

```
186 <<*Define core switching macros>> ≡
187 <<*Define core switching macros>> ≡
188 \countdef\last@language=19 % TODO. why? remove?
189 \def\addlanguage{\csname newlanguage\endcsname}
190 <</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 L<sup>A</sup>T<sub>E</sub>X 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.

## 7.2 The Package File (L<sup>A</sup>T<sub>E</sub>X, `babel.sty`)

This file 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. The first two options are for debugging.

```
191 <*package>
192 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
193 \ProvidesPackage{babel}[\<<date>> \<<version>> The Babel package]
194 \@ifpackagewith{babel}{debug}
195   {\providecommand\bbl@trace[1]{\message{^^J[ #1 ]}}%
196    \let\bbl@debug\@firstofone
197    \ifx\directlua\@undefined\else
198      \directlua{ Babel = Babel or {}
199                Babel.debug = true }%
200    \fi}
201 {\providecommand\bbl@trace[1]{}%
202  \let\bbl@debug\@gobble
203  \ifx\directlua\@undefined\else
204    \directlua{ Babel = Babel or {}
205              Babel.debug = false }%
206  \fi}
207 <<Basic macros>>
208 % Temporarily repeat here the code for errors. TODO.
209 \def\bbl@error#1#2{%
210   \begingroup
```

```

211     \def\{\MessageBreak}%
212     \PackageError{babel}{#1}{#2}%
213   \endgroup}
214 \def\bbl@warning#1{%
215   \begingroup
216     \def\{\MessageBreak}%
217     \PackageWarning{babel}{#1}%
218   \endgroup}
219 \def\bbl@infowarn#1{%
220   \begingroup
221     \def\{\MessageBreak}%
222     \GenericWarning
223       {(babel) \@spaces\@spaces\@spaces}%
224       {Package babel Info: #1}%
225   \endgroup}
226 \def\bbl@info#1{%
227   \begingroup
228     \def\{\MessageBreak}%
229     \PackageInfo{babel}{#1}%
230   \endgroup}
231 \def\bbl@nocaption{\protect\bbl@nocaption@i}
232 % TODO - Wrong for \today !!! Must be a separate macro.
233 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
234   \global\@namedef{#2}{\textbf{?#1?}}%
235   \@nameuse{#2}%
236   \edef\bbl@tempa{#1}%
237   \bbl@sreplace\bbl@tempa{name}{}}%
238   \bbl@warning{%
239     \@backslashchar#1 not set for '\language'. Please,\%
240     define it after the language has been loaded\%
241     (typically in the preamble) with\%
242     \string\setlocalecaption{\language}{\bbl@tempa}{..\%
243     Reported}}
244 \def\bbl@tentative{\protect\bbl@tentative@i}
245 \def\bbl@tentative@i#1{%
246   \bbl@warning{%
247     Some functions for '#1' are tentative.\%
248     They might not work as expected and their behavior\%
249     may change in the future.\%
250     Reported}}
251 \def\@nolanerr#1{%
252   \bbl@error
253     {You haven't defined the language #1\space yet.\%
254     Perhaps you misspelled it or your installation\%
255     is not complete}%
256     {Your command will be ignored, type <return> to proceed}}
257 \def\@nopatterns#1{%
258   \bbl@warning
259     {No hyphenation patterns were preloaded for\%
260     the language `#1' into the format.\%
261     Please, configure your TeX system to add them and\%
262     rebuild the format. Now I will use the patterns\%
263     preloaded for \bbl@nulllanguage\space instead}}
264   % End of errors
265 \ifpackagewith{babel}{silent}
266   {\let\bbl@info\@gobble
267    \let\bbl@infowarn\@gobble
268    \let\bbl@warning\@gobble}
269   {}

```

```

270 %
271 \def\AfterBabelLanguage#1{%
272   \global\expandafter\bbbl@add\csname#1.ldf-h@@k\endcsname}%

```

If the format created a list of loaded languages (in `\bbbl@languages`), get the name of the 0-th to show the actual language used. Also available with `base`, because it just shows info.

```

273 \ifx\bbbl@languages\undefined\else
274   \begingroup
275     \catcode`\^^I=12
276     \@ifpackagewith{babel}{showlanguages}{%
277       \begingroup
278         \def\bbbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}%
279         \wlog{<*languages>}%
280         \bbbl@languages
281         \wlog{</languages>}%
282       \endgroup}{%
283     \endgroup
284     \def\bbbl@elt#1#2#3#4{%
285       \ifnum#2=\z@
286         \gdef\bbbl@nulllanguage{#1}%
287         \def\bbbl@elt##1##2##3##4{%
288           \fi}%
289       \bbbl@languages
290     \fi%

```

### 7.3 base

The first ‘real’ option to be processed is `base`, which set the hyphenation patterns then resets `ver@babel.sty` so that `TeX` forgets about the first loading. After a subset of `babel.def` has been loaded (the old `switch.def`) and `\AfterBabelLanguage` defined, it exits.

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`.

```

291 \bbbl@trace{Defining option 'base'}
292 \@ifpackagewith{babel}{base}{%
293   \let\bbbl@onlyswitch\@empty
294   \let\bbbl@provide@locale\relax
295   \input babel.def
296   \let\bbbl@onlyswitch\@undefined
297   \ifx\directlua\@undefined
298     \DeclareOption*{\bbbl@patterns{\CurrentOption}}%
299   \else
300     \input luababel.def
301     \DeclareOption*{\bbbl@patterns@lua{\CurrentOption}}%
302   \fi
303   \DeclareOption{base}{}%
304   \DeclareOption{showlanguages}{}%
305   \ProcessOptions
306   \global\expandafter\let\csname opt@babel.sty\endcsname\relax
307   \global\expandafter\let\csname ver@babel.sty\endcsname\relax
308   \global\let\@ifl@ter@\@ifl@ter
309   \def\@ifl@ter#1#2#3#4#5{\global\let\@ifl@ter\@ifl@ter@@}%
310   \endinput}{%
311 % \end{macrocode}
312 %
313 % \subsection{\texttt{key=value} options and other general option}
314 %
315 %   The following macros extract language modifiers, and only real
316 %   package options are kept in the option list. Modifiers are saved

```

```

317% and assigned to |\BabelModifiers| at |\bbl@load@language|; when
318% no modifiers have been given, the former is |\relax|. How
319% modifiers are handled are left to language styles; they can use
320% |\in@|, loop them with |\@for| or load |keyval|, for example.
321%
322% \begin{macrocode}
323 \bbl@trace{key=value and another general options}
324 \bbl@csarg\let{tempa\expandafter}\csname opt@babel.sty\endcsname
325 \def\bbl@tempb#1.#2{% Remove trailing dot
326   #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%
327 \def\bbl@tempd#1.#2@nnil{% TODO. Refactor lists?
328   \ifx\@empty#2%
329     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
330   \else
331     \in@{,provide,},{, #1,}%
332     \ifin@
333       \edef\bbl@tempc{%
334         \ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.\bbl@tempb#2}%
335     \else
336       \in@{=}{#1}%
337       \ifin@
338         \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
339       \else
340         \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
341         \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
342       \fi
343     \fi
344   \fi}
345 \let\bbl@tempc\@empty
346 \bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty\@nnil}
347 \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.

```

348 \DeclareOption{KeepShorthandsActive}{}
349 \DeclareOption{activeacute}{}
350 \DeclareOption{activegrave}{}
351 \DeclareOption{debug}{}
352 \DeclareOption{noconfigs}{}
353 \DeclareOption{showlanguages}{}
354 \DeclareOption{silent}{}
355 \DeclareOption{mono}{}
356 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
357 \chardef\bbl@iniflag\z@
358 \DeclareOption{provide=*}{\chardef\bbl@iniflag\@ne} % main -> +1
359 \DeclareOption{provide+=*}{\chardef\bbl@iniflag\tw@} % add = 2
360 \DeclareOption{provide*=*}{\chardef\bbl@iniflag\thr@@} % add + main
361 % A separate option
362 \let\bbl@autoload@options\@empty
363 \DeclareOption{provide@=*}{\def\bbl@autoload@options{import}}
364 % Don't use. Experimental. TODO.
365 \newif\ifbbl@single
366 \DeclareOption{selectors=off}{\bbl@singletrue}
367 <<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.

```
368 \let\bbl@opt@shorthands\@nnil
369 \let\bbl@opt@config\@nnil
370 \let\bbl@opt@main\@nnil
371 \let\bbl@opt@headfoot\@nnil
372 \let\bbl@opt@layout\@nnil
```

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

```
373 \def\bbl@tempa#1=#2\bbl@tempa{%
374   \bbl@csarg\ifx{opt@#1}\@nnil
375     \bbl@csarg\edef{opt@#1}{#2}%
376   \else
377     \bbl@error
378     {Bad option `#1=#2'. Either you have misspelled the\\%
379     key or there is a previous setting of `#1'. Valid\\%
380     keys are, among others, `shorthands', `main', `bidi',\\%
381     `strings', `config', `headfoot', `safe', `math'.}%
382     {See the manual for further details.}
383   \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.

```
384 \let\bbl@language@opts\@empty
385 \DeclareOption*{%
386   \bbl@xin@{\string=}{\CurrentOption}%
387   \ifin@
388     \expandafter\bbl@tempa\CurrentOption\bbl@tempa
389   \else
390     \bbl@add@list\bbl@language@opts{\CurrentOption}%
391   \fi}
```

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

```
392 \ProcessOptions*
```

## 7.4 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=...

```
393 \bbl@trace{Conditional loading of shorthands}
394 \def\bbl@sh@string#1{%
395   \ifx#1\@empty\else
396     \ifx#1t\string~%
397     \else\ifx#1c\string,%
398     \else\string#1%
399     \fi\fi
400   \expandafter\bbl@sh@string
401   \fi}
402 \ifx\bbl@opt@shorthands\@nnil
403   \def\bbl@ifshorthand#1#2#3{#2}%
404 \else\ifx\bbl@opt@shorthands\@empty
405   \def\bbl@ifshorthand#1#2#3{#3}%
406 \else
```

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

```
407   \def\bbl@ifshorthand#1{%
```



```

408 \bbl@xin@{\string#1}{\bbl@opt@shorthands}%
409 \ifin@
410 \expandafter\@firstoftwo
411 \else
412 \expandafter\@secondoftwo
413 \fi}

```

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

```

414 \edef\bbl@opt@shorthands{%
415 \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.

```

416 \bbl@ifshorthand{'}%
417 {\PassOptionsToPackage{activeacute}{babel}}{}
418 \bbl@ifshorthand{`}%
419 {\PassOptionsToPackage{activegrave}{babel}}{}
420 \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.

```

421 \ifx\bbl@opt@headfoot\@nnil\else
422 \g@addto@macro\@resetactivechars{%
423 \set@typeset@protect
424 \expandafter\select@language@x\expandafter{\bbl@opt@headfoot}%
425 \let\protect\noexpand}
426 \fi

```

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

```

427 \ifx\bbl@opt@safe\undefined
428 \def\bbl@opt@safe{BR}
429 \fi
430 \ifx\bbl@opt@main\@nnil\else
431 \edef\bbl@language@opts{%
432 \ifx\bbl@language@opts\@empty\else\bbl@language@opts,\fi
433 \bbl@opt@main}
434 \fi

```

For layout an auxiliary macro is provided, available for packages and language styles. Optimization: if there is no layout, just do nothing.

```

435 \bbl@trace{Defining IfBabelLayout}
436 \ifx\bbl@opt@layout\@nnil
437 \newcommand\IfBabelLayout[3]{#3}%
438 \else
439 \newcommand\IfBabelLayout[1]{%
440 \@expandtwoargs\in@{.#1.}{.\bbl@opt@layout.}%
441 \ifin@
442 \expandafter\@firstoftwo
443 \else
444 \expandafter\@secondoftwo
445 \fi}
446 \fi

```

**Common definitions.** *In progress.* Still based on babel.def, but the code should be moved here.

```

447 \input babel.def

```

## 7.5 Cross referencing macros

The  $\TeX$  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 following package options control which macros are to be redefined.

```
448 <<*More package options>> ≡
449 \DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
450 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
451 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
452 <</More package options>>
```

`\@newl@bel` 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.

```
453 \bbl@trace{Cross referencing macros}
454 \ifx\bbl@opt@safe\@empty\else
455   \def\@newl@bel#1#2#3{%
456     {\@safe@activestrue
457       \bbl@ifunset{#1@#2}%
458       \relax
459       {\gdef\@multiplelabels{%
460         \@latex@warning@no@line{There were multiply-defined labels}}%
461         \@latex@warning@no@line{Label `#2' multiply defined}}%
462       \global\@namedef{#1@#2}{#3}}}
```

`\@testdef` An internal  $\TeX$  macro used to test if the labels that have been written on the .aux file have changed. It is called by the `\enddocument` macro.

```
463 \CheckCommand*\@testdef[3]{%
464   \def\reserved@a{#3}%
465   \expandafter\ifx\csname#1@#2\endcsname\reserved@a
466   \else
467     \@tempswatrue
468   \fi}
```

Now that we made sure that `\@testdef` still has the same definition we can rewrite it. First we make the shorthands ‘safe’. Then we use `\bbl@tempa` as an ‘alias’ for the macro that contains the label which is being checked. Then we define `\bbl@tempb` just as `\@newl@bel` does it. When the label is defined we replace the definition of `\bbl@tempa` by its meaning. If the label didn’t change, `\bbl@tempa` and `\bbl@tempb` should be identical macros.

```
469 \def\@testdef#1#2#3{% TODO. With @samestring?
470   \@safe@activestrue
471   \expandafter\let\expandafter\bbl@tempa\csname #1@#2\endcsname
472   \def\bbl@tempb{#3}%
473   \@safe@activesfalse
474   \ifx\bbl@tempa\relax
475   \else
476     \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
477   \fi
478   \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
479   \ifx\bbl@tempa\bbl@tempb
480   \else
481     \@tempswatrue
```

```

482   \fi}
483 \fi

\ref    The same holds for the macro \ref that references a label and \pageref to reference a page. We
\pageref make them robust as well (if they weren't already) to prevent problems if they should become
         expanded at the wrong moment.

484 \bbl@xin@{R}\bbl@opt@safe
485 \ifin@
486   \bbl@redefineroast\ref#1{%
487     \@safe@activetrue\org@ref{#1}\@safe@activesfalse}
488   \bbl@redefineroast\pageref#1{%
489     \@safe@activetrue\org@pageref{#1}\@safe@activesfalse}
490 \else
491   \let\org@ref\ref
492   \let\org@pageref\pageref
493 \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.

494 \bbl@xin@{B}\bbl@opt@safe
495 \ifin@
496   \bbl@redefine\@citex[#1]#2{%
497     \@safe@activetrue\edef\@tempa{#2}\@safe@activesfalse
498     \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.

499 \AtBeginDocument{%
500   \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.)

501   \def\@citex[#1][#2]#3{%
502     \@safe@activetrue\edef\@tempa{#3}\@safe@activesfalse
503     \org@@citex[#1][#2]{\@tempa}}%
504   }{}}

The package cite has a definition of \@citex where the shorthands need to be turned off in both
arguments.

505 \AtBeginDocument{%
506   \ifpackageloaded{cite}{%
507     \def\@citex[#1]#2{%
508       \@safe@activetrue\org@@citex[#1]{#2}\@safe@activesfalse}%
509     }{}}

\nocite The macro \nocite which is used to instruct BiBTeX to extract uncited references from the database.

510 \bbl@redefine\nocite#1{%
511   \@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 \bibtex is needed we define \bibtex in such a way that it redefines itself with the proper definition. We call \bbl@cite@choice to select the proper definition for \bibtex. This new definition is then activated.

```
512 \bbl@redefine\bibtex{%
513   \bbl@cite@choice
514   \bibtex}
```

\bbl@bibtex The macro \bbl@bibtex holds the definition of \bibtex needed when neither natbib nor cite is loaded.

```
515 \def\bbl@bibtex#1#2{%
516   \org@bibtex{#1}{\@safe@activesfalse#2}}
```

\bbl@cite@choice The macro \bbl@cite@choice determines which definition of \bibtex is needed. First we give \bibtex its default definition.

```
517 \def\bbl@cite@choice{%
518   \global\let\bibtex\bbl@bibtex
519   \@ifpackageloaded{natbib}{\global\let\bibtex\org@bibtex}{}%
520   \@ifpackageloaded{cite}{\global\let\bibtex\org@bibtex}{}%
521   \global\let\bbl@cite@choice\relax}
```

When a document is run for the first time, no .aux file is available, and \bibtex will not yet be properly defined. In this case, this has to happen before the document starts.

```
522 \AtBeginDocument{\bbl@cite@choice}
```

\@bibitem One of the two internal  $\TeX$  macros called by \bibitem that write the citation label on the .aux file.

```
523 \bbl@redefine\@bibitem#1{%
524   \@safe@activestrue\org@bibitem{#1}\@safe@activesfalse}
525 \else
526   \let\org@nocite\nocite
527   \let\org@citex\citex
528   \let\org@bibtex\bibtex
529   \let\org@bibitem\@bibitem
530 \fi
```

## 7.6 Marks

\markright Because the output routine is asynchronous, we must pass the current language attribute to the head lines. To achieve this we need to adapt the definition of \markright and \markboth somewhat. However, headlines and footlines can contain text outside marks; for that we must take some actions in the output routine if the 'headfoot' options is used. We need to make some redefinitions to the output routine to avoid an endless loop and to correctly handle the page number in bidi documents.

```
531 \bbl@trace{Marks}
532 \IfBabelLayout{sectioning}
533   {\ifx\bbl@opt@headfoot\@nnil
534     \g@addto@macro\resetactivechars{%
535       \set@typeset@protect
536       \expandafter\select@language@x\expandafter{\bbl@main@language}%
537       \let\protect\noexpand
538       \ifcase\bbl@bidimode\else % Only with bidi. See also above
539         \edef\thepage{%
540           \noexpand\babelsublr{\unexpanded\expandafter{\thepage}}}%
541       \fi}%
542   \fi}
543 {\ifbbl@single\else
544   \bbl@ifunset{markright }{\bbl@redefine\bbl@redefineroobust
545     \markright#1{%
```

```

546      \bbl@ifblank{#1}%
547      {\org@markright{}}%
548      {\toks@{#1}}%
549      \bbl@exp{%
550          \\org@markright{\\protect\\foreignlanguage{\language}%
551              {\\protect\\bbl@restore@actives\the\toks@}}}%
\markboth The definition of \markboth is equivalent to that of \markright, except that we need two token
\@mkboth 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 macro, so it's not necessary anymore,
but it's preserved for older versions.)

552      \ifx\@mkboth\markboth
553          \def\bbl@tempc{\let\@mkboth\markboth}
554      \else
555          \def\bbl@tempc{
556              \fi
557              \bbl@ifunset{markboth } \bbl@redefine\bbl@redefineroobust
558              \markboth#1#2{%
559                  \protected@edef\bbl@tempb##1{%
560                      \protect\foreignlanguage
561                          {\language}{\protect\bbl@restore@actives##1}}%
562                      \bbl@ifblank{#1}%
563                      {\toks@{}}%
564                      {\toks@\expandafter{\bbl@tempb{#1}}}%
565                      \bbl@ifblank{#2}%
566                      {\@temptokena{}}%
567                      {\@temptokena\expandafter{\bbl@tempb{#2}}}%
568                      \bbl@exp{\\org@markboth{\the\toks@}{\the\@temptokena}}
569                      \bbl@tempc
570              \fi} % end ifbbl@single, end \IfBabelLayout

```

## 7.7 Preventing clashes with other packages

### 7.7.1 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.

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.

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.

```

571 \bbl@trace{Preventing clashes with other packages}
572 \bbl@xin@{R}\bbl@opt@safe
573 \ifin@
574 \AtBeginDocument{%
575     \@ifpackageloaded{ifthen}{%
576         \bbl@redefine@long\ifthenelse#1#2#3{%

```

```

577 \let\bbl@temp@pref\pageref
578 \let\pageref\org@pageref
579 \let\bbl@temp@ref\ref
580 \let\ref\org@ref
581 \@safe@activetrue
582 \org@ifthenelse{#1}%
583   {\let\pageref\bbl@temp@pref
584    \let\ref\bbl@temp@ref
585    \@safe@activesfalse
586    #2}%
587   {\let\pageref\bbl@temp@pref
588    \let\ref\bbl@temp@ref
589    \@safe@activesfalse
590    #3}%
591   }%
592   }{}%
593 }

```

### 7.7.2 varioref

`\@vpageref` When the package `varioref` is in use we need to modify its internal command `\@vpageref` in order to prevent problems when an active character ends up in the argument of `\vref`. The same needs to happen for `\vrefpagenum`.

```

594 \AtBeginDocument{%
595   \@ifpackageloaded{varioref}{%
596     \bbl@redefine\@vpageref#1[#2]#3{%
597       \@safe@activetrue
598       \org@@@vpageref{#1}[#2]{#3}%
599       \@safe@activesfalse}%
600     \bbl@redefine\vrefpagenum#1#2{%
601       \@safe@activetrue
602       \org@vrefpagenum{#1}[#2]%
603       \@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.

```

604   \expandafter\def\csname Ref \endcsname#1{%
605     \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
606   }{}%
607 }
608 \fi

```

### 7.7.3 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. 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.

```

609 \AtEndOfPackage{%
610   \AtBeginDocument{%
611     \@ifpackageloaded{hhline}%
612     {\expandafter\ifx\csname normal@char\string\endcsname\relax
613       \else
614         \makeatletter

```

```

615         \def\@currname{hhline}\input{hhline.sty}\makeatother
616         \fi}%
617     {}}}

```

#### 7.7.4 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. TODO. Still true? Commented out in 2020/07/27.

```

618% \AtBeginDocument{%
619%     \ifx\pdfstringdefDisableCommands\@undefined\else
620%         \pdfstringdefDisableCommands{\languageshorthands{system}}%
621%     \fi}

```

#### 7.7.5 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.

```

622 \DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
623     \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. This command is deprecated. Use the tools provided by  $\TeX$ .

```

624 \def\substitutefontfamily#1#2#3{%
625     \lowercase{\immediate\openout15=#1#2.fd\relax}%
626     \immediate\write15{%
627         \string\ProvidesFile{#1#2.fd}%
628         [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
629         \space generated font description file]^{}
630         \string\DeclareFontFamily{#1}{#2}{ }^{}
631         \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{ }^{}
632         \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{ }^{}
633         \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{ }^{}
634         \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{ }^{}
635         \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{ }^{}
636         \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{ }^{}
637         \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{ }^{}
638         \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{ }^{}
639     }%
640     \closeout15
641 }
642 \@onlypreamble\substitutefontfamily

```

### 7.8 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 have been loaded by traversing `\@filelist` to search for `<enc>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
643 \bb1@trace{Encoding and fonts}

```

```

644 \newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU,PU,PD1}
645 \newcommand\BabelNonText{TS1,T3,TS3}
646 \let\org@TeX\TeX
647 \let\org@LaTeX\LaTeX
648 \let\ensureascii\@firstofone
649 \AtBeginDocument{%
650   \in@false
651   \bbl@foreach\BabelNonASCII{% is there a text non-ascii enc?
652     \ifin@%
653       \lowercase{\bbl@xin@{,#1enc.def,},{, \@filelist,}}%
654     \fi}%
655   \ifin@ % if a text non-ascii has been loaded
656     \def\ensureascii#1{{\fontencoding{OT1}\selectfont#1}}%
657     \DeclareTextCommandDefault{\TeX}{\org@TeX}%
658     \DeclareTextCommandDefault{\LaTeX}{\org@LaTeX}%
659     \def\bbl@tempb#1\@{\uppercase{\bbl@tempc#1}ENC.DEF\@empty\@}%
660     \def\bbl@tempc#1ENC.DEF#2\@{\%
661       \ifx\@empty#2\else
662         \bbl@ifunset{T@#1}%
663         {}%
664         {\bbl@xin@{,#1,},{, \BabelNonASCII, \BabelNonText,}%
665         \ifin@
666           \DeclareTextCommand{\TeX}{#1}{\ensureascii{\org@TeX}}%
667           \DeclareTextCommand{\LaTeX}{#1}{\ensureascii{\org@LaTeX}}%
668         \else
669           \def\ensureascii#1{{\fontencoding{#1}\selectfont#1}}%
670         \fi}%
671       \fi}%
672   \bbl@foreach\@filelist{\bbl@tempb#1\@}% TODO - \@ de mas??
673   \bbl@xin@{,\cf@encoding,},{, \BabelNonASCII, \BabelNonText,}%
674   \ifin@%
675     \edef\ensureascii#1{{%
676       \noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}}%
677   \fi
678 \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.

```

679 \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.

```

680 \AtBeginDocument{%
681   \@ifpackageloaded{fontspec}%
682   {\xdef\latinencoding{%
683     \ifx\UTFfencname\@undefined
684       EU\ifcase\bbl@engine\or2\or1\fi
685     \else
686       \UTFfencname
687     \fi}}%
688   {\gdef\latinencoding{OT1}}%
689   \ifx\cf@encoding\bbl@t@one
690     \xdef\latinencoding{\bbl@t@one}%

```



```

691 \else
692 \ifx\@fontenc@load@list\@undefined
693 \@ifl@aded{def}{t1enc}{\xdef\latinencoding{\bbl@t@one}}{}%
694 \else
695 \def\@elt#1{,#1,}%
696 \edef\bbl@tempa{\expandafter\@gobbletwo\@fontenc@load@list}%
697 \let\@elt\relax
698 \bbl@xin@{,T1,}\bbl@tempa
699 \ifin@
700 \xdef\latinencoding{\bbl@t@one}%
701 \fi
702 \fi
703 \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.

```

704 \DeclareRobustCommand{\latintext}{%
705 \fontencoding{\latinencoding}\selectfont
706 \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.

```

707 \ifx\@undefined\DeclareTextFontCommand
708 \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
709 \else
710 \DeclareTextFontCommand{\textlatin}{\latintext}
711 \fi

```

## 7.9 Basic bidi support

**Work in progress.** This code is currently placed here for practical reasons. It will be moved to the correct place soon, I hope.

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 $\TeX$ -ja` shows, vertical typesetting is possible, too.

As a first step, add a handler for bidi and digits (and potentially other processes) just before `luaotfload` is applied, which is loaded by default by  $\LaTeX$ . Just in case, consider the possibility it has not been loaded.

```

712 \ifodd\bbl@engine
713 \def\bbl@activate@preotf{%
714 \let\bbl@activate@preotf\relax % only once
715 \directlua{

```

```

716     Babel = Babel or {}
717     %
718     function Babel.pre_otfload_v(head)
719         if Babel.numbers and Babel.digits_mapped then
720             head = Babel.numbers(head)
721         end
722         if Babel.bidi_enabled then
723             head = Babel.bidi(head, false, dir)
724         end
725         return head
726     end
727     %
728     function Babel.pre_otfload_h(head, gc, sz, pt, dir)
729         if Babel.numbers and Babel.digits_mapped then
730             head = Babel.numbers(head)
731         end
732         if Babel.bidi_enabled then
733             head = Babel.bidi(head, false, dir)
734         end
735         return head
736     end
737     %
738     luatexbase.add_to_callback('pre_linebreak_filter',
739         Babel.pre_otfload_v,
740         'Babel.pre_otfload_v',
741         luatexbase.priority_in_callback('pre_linebreak_filter',
742             'luaotfload.node_processor') or nil)
743     %
744     luatexbase.add_to_callback('hpack_filter',
745         Babel.pre_otfload_h,
746         'Babel.pre_otfload_h',
747         luatexbase.priority_in_callback('hpack_filter',
748             'luaotfload.node_processor') or nil)
749     }}
750 \fi

```

The basic setup. In luatex, the output is modified at a very low level to set the `\bodydir` to the `\pagedir`.

```

751 \bbl@trace{Loading basic (internal) bidi support}
752 \ifodd\bbl@engine
753   \ifnum\bbl@bidimode>100 \ifnum\bbl@bidimode<200
754     \let\bbl@beforeforeign\leavevmode
755     \AtEndOfPackage{\EnableBabelHook{babel-bidi}}
756     \RequirePackage{luatexbase}
757     \bbl@activate@preotf
758     \directlua{
759       require('babel-data-bidi.lua')
760       \ifcase\expandafter\@gobbletwo\the\bbl@bidimode\or
761         require('babel-bidi-basic.lua')
762       \or
763         require('babel-bidi-basic-r.lua')
764     }
765     % TODO - to locale_props, not as separate attribute
766     \newattribute\bbl@attr@dir
767     % TODO. I don't like it, hackish:
768     \bbl@exp{\output{\bodydir\pagedir\the\output}}
769     \AtEndOfPackage{\EnableBabelHook{babel-bidi}}
770   \fi\fi
771 \else

```

```

772 \ifnum\bbl@bidimode>100 \ifnum\bbl@bidimode<200
773   \bbl@error
774   {The bidi method 'basic' is available only in\%
775     luatex. I'll continue with 'bidi=default', so\%
776     expect wrong results}%
777   {See the manual for further details.}%
778   \let\bbl@beforeforeign\leavevmode
779   \AtEndOfPackage{%
780     \EnableBabelHook{babel-bidi}%
781     \bbl@xebidipar}
782 \fi\fi
783 \def\bbl@loadxebidi#1{%
784   \ifx\RTLfootnotetext\@undefined
785     \AtEndOfPackage{%
786       \EnableBabelHook{babel-bidi}%
787       \ifx\fontspec\@undefined
788         \bbl@loadfontspec % bidi needs fontspec
789       \fi
790       \usepackage#1{bidi}}%
791   \fi}
792 \ifnum\bbl@bidimode>200
793   \ifcase\expandafter\@gobbletwo\the\bbl@bidimode\or
794     \bbl@tentative{bidi=bidi}
795     \bbl@loadxebidi{}
796   \or
797     \bbl@loadxebidi{[rldocument]}
798   \or
799     \bbl@loadxebidi{}
800   \fi
801 \fi
802 \fi
803 \ifnum\bbl@bidimode=\@ne
804   \let\bbl@beforeforeign\leavevmode
805   \ifodd\bbl@engine
806     \newattribute\bbl@attr@dir
807     \bbl@exp{\output{\bodydir\pagedir\the\output}}}%
808   \fi
809   \AtEndOfPackage{%
810     \EnableBabelHook{babel-bidi}%
811     \ifodd\bbl@engine\else
812       \bbl@xebidipar
813     \fi}
814 \fi

```

Now come the macros used to set the direction when a language is switched. First the (mostly) common macros.

```

815 \bbl@trace{Macros to switch the text direction}
816 \def\bbl@alscripts{,Arabic,Syriac,Thaana,}
817 \def\bbl@rscripts{% TODO. Base on codes ??
818   ,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
819   Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaeen,%
820   Manichaeen,Meroitic Cursive,Meroitic,Old North Arabian,%
821   Nabataean,N'Ko,Orkhon,Palmyrene,Inscriptional Pahlavi,%
822   Psalter Pahlavi,Phoenician,Inscriptional Parthian,Samaritan,%
823   Old South Arabian,}%
824 \def\bbl@provide@dirs#1{%
825   \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts\bbl@rscripts}%
826   \ifin@
827     \global\bbl@csarg\chardef{wdir@#1}\@ne

```

```

828 \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts}%
829 \ifin@
830 \global\bbl@csarg\chardef{wdir@#1}\tw@ % useless in xetex
831 \fi
832 \else
833 \global\bbl@csarg\chardef{wdir@#1}\z@
834 \fi
835 \ifodd\bbl@engine
836 \bbl@csarg\ifcase{wdir@#1}%
837 \directlua{ Babel.locale_props[\the\localeid].texmdir = 'l' }%
838 \or
839 \directlua{ Babel.locale_props[\the\localeid].texmdir = 'r' }%
840 \or
841 \directlua{ Babel.locale_props[\the\localeid].texmdir = 'al' }%
842 \fi
843 \fi}
844 \def\bbl@switchdir{%
845 \bbl@ifunset{bbl@sys@\language name}{\bbl@provide@sys{\language name}}}%
846 \bbl@ifunset{bbl@wdir@\language name}{\bbl@provide@dirs{\language name}}}%
847 \bbl@exp{\bbl@setdirs\bbl@cl{wdir}}%
848 \def\bbl@setdirs#1{% TODO - math
849 \ifcase\bbl@select@type % TODO - strictly, not the right test
850 \bbl@bodydir{#1}%
851 \bbl@pardir{#1}%
852 \fi
853 \bbl@texmdir{#1}}
854 % TODO. Only if \bbl@bidimode > 0?:
855 \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
856 \DisableBabelHook{babel-bidi}

```

Now the engine-dependent macros. TODO. Must be moved to the engine files?

```

857 \ifodd\bbl@engine % luatex=1
858 \chardef\bbl@thetexmdir\z@
859 \chardef\bbl@thepardir\z@
860 \def\bbl@getluadir#1{%
861 \directlua{
862 if tex.#1dir == 'TLT' then
863 tex.sprint('0')
864 elseif tex.#1dir == 'TRT' then
865 tex.sprint('1')
866 end}}
867 \def\bbl@setluadir#1#2#3{% 1=text/par.. 2=\texmdir.. 3=0 lr/1 r1
868 \ifcase#3\relax
869 \ifcase\bbl@getluadir{#1}\relax\else
870 #2 TLT\relax
871 \fi
872 \else
873 \ifcase\bbl@getluadir{#1}\relax
874 #2 TRT\relax
875 \fi
876 \fi}
877 \def\bbl@texmdir#1{%
878 \bbl@setluadir{tex}\texmdir{#1}%
879 \chardef\bbl@thetexmdir#1\relax
880 \setattribute\bbl@attr@dir{\numexpr\bbl@thepardir*3+#1}}
881 \def\bbl@pardir#1{%
882 \bbl@setluadir{par}\pardir{#1}%
883 \chardef\bbl@thepardir#1\relax}
884 \def\bbl@bodydir{\bbl@setluadir{body}\bodydir}

```

```

885 \def\bbl@pagedir{\bbl@setluadir{page}\pagedir}
886 \def\bbl@dirparastext{\paddir\the\textdir\relax}%   %%%
887 % Sadly, we have to deal with boxes in math with basic.
888 % Activated every math with the package option bidi=:
889 \def\bbl@mathboxdir{%
890   \ifcase\bbl@thetextdir\relax
891     \everyhbox{\textdir TLT\relax}%
892   \else
893     \everyhbox{\textdir TRT\relax}%
894   \fi}
895 \frozen@everymath\expandafter{%
896   \expandafter\bbl@mathboxdir\the\frozen@everymath}
897 \frozen@everydisplay\expandafter{%
898   \expandafter\bbl@mathboxdir\the\frozen@everydisplay}
899 \else % pdftex=0, xetex=2
900   \newcount\bbl@dirlevel
901   \chardef\bbl@thetextdir\z@
902   \chardef\bbl@thepaddir\z@
903   \def\bbl@textdir#1{%
904     \ifcase#1\relax
905       \chardef\bbl@thetextdir\z@
906       \bbl@textdir@i\beginL\endL
907     \else
908       \chardef\bbl@thetextdir\@ne
909       \bbl@textdir@i\beginR\endR
910     \fi}
911   \def\bbl@textdir@i#1#2{%
912     \ifhmode
913       \ifnum\currentgrouplevel>\z@
914         \ifnum\currentgrouplevel=\bbl@dirlevel
915           \bbl@error{Multiple bidi settings inside a group}%
916           {I'll insert a new group, but expect wrong results.}%
917           \bgroup\aftergroup#2\aftergroup\egroup
918         \else
919           \ifcase\currentgrouptype\or % 0 bottom
920             \aftergroup#2% 1 simple {}
921           \or
922             \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
923           \or
924             \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
925           \or\or % vbox vtop align
926           \or
927             \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
928           \or\or\or\or\or\or % output math disc insert vcent mathchoice
929           \or
930             \aftergroup#2% 14 \begingroup
931           \else
932             \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
933           \fi
934         \fi
935         \bbl@dirlevel\currentgrouplevel
936       \fi
937       #1%
938     \fi}
939   \def\bbl@paddir#1{\chardef\bbl@thepaddir#1\relax}
940   \let\bbl@bodydir@gobble
941   \let\bbl@pagedir@gobble
942   \def\bbl@dirparastext{\chardef\bbl@thepaddir\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).

```

943 \def\bbl@xebidipar{%
944   \let\bbl@xebidipar\relax
945   \TeXeTstate\@ne
946   \def\bbl@xeeverypar{%
947     \ifcase\bbl@thepardir
948       \ifcase\bbl@thetextdir\else\beginR\fi
949     \else
950       {\setbox\z@\lastbox\beginR\box\z@}%
951     \fi}%
952   \let\bbl@severypar\everypar
953   \newtoks\everypar
954   \everypar=\bbl@severypar
955   \bbl@severypar{\bbl@xeeverypar\the\everypar}}
956 \ifnum\bbl@bidimode>200
957   \let\bbl@textdir\i\@gobbletwo
958   \let\bbl@xebidipar\@empty
959   \AddBabelHook{bidi}{foreign}{%
960     \def\bbl@tempa{\def\BabelText####1}%
961     \ifcase\bbl@thetextdir
962       \expandafter\bbl@tempa\expandafter{\BabelText{\LR{##1}}}%
963     \else
964       \expandafter\bbl@tempa\expandafter{\BabelText{\RL{##1}}}%
965     \fi}
966   \def\bbl@pardir#1{\ifcase#1\relax\setLR\else\setRL\fi}
967 \fi
968 \fi

```

A tool for weak L (mainly digits). We also disable warnings with `hyperref`.

```

969 \DeclareRobustCommand\babelsublr[1]{\leavevmode{\bbl@textdir\z@#1}}
970 \AtBeginDocument{%
971   \ifx\pdfstringdefDisableCommands\@undefined\else
972     \ifx\pdfstringdefDisableCommands\relax\else
973       \pdfstringdefDisableCommands{\let\babelsublr\@firstofone}%
974     \fi
975   \fi}

```

## 7.10 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`.

```

976 \bbl@trace{Local Language Configuration}
977 \ifx\loadlocalcfg\@undefined
978   \@ifpackagewith{babel}{noconfigs}%
979   {\let\loadlocalcfg\@gobble}%
980   {\def\loadlocalcfg#1{%
981     \InputIfFileExists{#1.cfg}%
982     {\typeout{*****^J%
983               * Local config file #1.cfg used^^J%
984               *}}}%
985     \@empty}}
986 \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. TODO. Necessary? Correct place? Used by some ldf file?

```

987 \ifx\@unexpandable@protect\@undefined
988   \def\@unexpandable@protect{\noexpand\protect\noexpand}
989   \long\def\protected@write#1#2#3{%
990     \begingroup
991       \let\thepage\relax
992       #2%
993       \let\protect\@unexpandable@protect
994       \edef\reserved@a{\write#1{#3}}%
995       \reserved@a
996     \endgroup
997     \if@nobreak\ifvmode\nobreak\fi\fi}
998 \fi
999 %
1000 % \subsection{Language options}
1001 %
1002 % Languages are loaded when processing the corresponding option
1003 % \textit{except} if a |main| language has been set. In such a
1004 % case, it is not loaded until all options has been processed.
1005 % The following macro inputs the ldf file and does some additional
1006 % checks (|\input| works, too, but possible errors are not caught).
1007 %
1008 %   \begin{macrocode}
1009 \bbl@trace{Language options}
1010 \let\bbl@afterlang\relax
1011 \let\BabelModifiers\relax
1012 \let\bbl@loaded\@empty
1013 \def\bbl@load@language#1{%
1014   \InputIfFileExists{#1.ldf}%
1015   {\edef\bbl@loaded{\CurrentOption
1016     \ifx\bbl@loaded\@empty\else,\bbl@loaded\fi}%
1017     \expandafter\let\expandafter\bbl@afterlang
1018       \csname\CurrentOption.ldf-h@@k\endcsname
1019     \expandafter\let\expandafter\BabelModifiers
1020       \csname bbl@mod@\CurrentOption\endcsname}%
1021   {\bbl@error{%
1022     Unknown option '\CurrentOption'. Either you misspelled it\\%
1023     or the language definition file \CurrentOption.ldf was not found}}%
1024     Valid options are, among others: shorthands=, KeepShorthandsActive,\\%
1025     activeacute, activegrave, noconfigs, safe=, main=, math=\\%
1026     headfoot=, strings=, config=, hyphenmap=, or a language name.}}}

```

Now, we set a few language options whose names are different from ldf files. These declarations are preserved for backwards compatibility, but they must be eventually removed. Use proxy files instead.

```

1027 \def\bbl@try@load@lang#1#2#3{%
1028   \IfFileExists{\CurrentOption.ldf}%
1029   {\bbl@load@language{\CurrentOption}}%
1030   {#1\bbl@load@language{#2}#3}}
1031 \DeclareOption{hebrew}{%
1032   \input{rlbabel.def}%
1033   \bbl@load@language{hebrew}}
1034 \DeclareOption{hungarian}{\bbl@try@load@lang{}{magyar}{}}
1035 \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}{}}
1036 \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}{}}
1037 \DeclareOption{polutonikogreek}{%
1038   \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}}
1039 \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}{}}

```

```

1040 \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}{}}
1041 \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.

```

1042 \ifx\bbl@opt@config\@nnil
1043   \@ifpackagewith{babel}{noconfigs}{}%
1044   {\InputIfFileExists{bblopts.cfg}%
1045     {\typeout{*****^^J%
1046               * Local config file bblopts.cfg used^^J%
1047               *}}%
1048     {}}%
1049 \else
1050   \InputIfFileExists{\bbl@opt@config.cfg}%
1051   {\typeout{*****^^J%
1052             * Local config file \bbl@opt@config.cfg used^^J%
1053             *}}%
1054   {\bbl@error{%
1055     Local config file '\bbl@opt@config.cfg' not found}%
1056     Perhaps you misspelled it.}}%
1057 \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.

```

1058 \let\bbl@tempc\relax
1059 \bbl@foreach\bbl@language@opts{%
1060   \ifcase\bbl@iniflag % Default
1061     \bbl@ifunset{ds@#1}%
1062     {\DeclareOption{#1}{\bbl@load@language{#1}}}%
1063     {}%
1064   \or % provide=*
1065     \@gobble % case 2 same as 1
1066   \or % provide+=*
1067     \bbl@ifunset{ds@#1}%
1068     {\IfFileExists{#1.ldf}{}%
1069      {\IfFileExists{babel-#1.tex}{}{\@namedef{ds@#1}{}}}}%
1070     {}%
1071   \bbl@ifunset{ds@#1}%
1072   {\def\bbl@tempc{#1}%
1073    \DeclareOption{#1}{%
1074      \ifnum\bbl@iniflag>\@ne
1075        \bbl@ldfinit
1076        \babelprovide[import]{#1}%
1077        \bbl@afterldf}%
1078      \else
1079        \bbl@load@language{#1}%
1080      \fi}%
1081   {}%
1082   \or % provide*=*
1083     \def\bbl@tempc{#1}%
1084     \bbl@ifunset{ds@#1}%
1085     {\DeclareOption{#1}{%
1086       \bbl@ldfinit
1087       \babelprovide[import]{#1}%
1088       \bbl@afterldf}}}%

```



```

1089     {}%
1090 \fi}

```

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.

```

1091 \let\bbl@tempb\@nnil
1092 \bbl@foreach\@classoptionslist{%
1093   \bbl@ifunset{ds@#1}%
1094     {\IfFileExists{#1.ldf}}{}%
1095     {\IfFileExists{babel-#1.tex}}{\@namedef{ds@#1}}{}%
1096   }%
1097   \bbl@ifunset{ds@#1}%
1098     {\def\bbl@tempb{#1}%
1099       \DeclareOption{#1}%
1100       \ifnum\bbl@iniflag>\@ne
1101         \bbl@ldfinit
1102         \babelprovide[import]{#1}%
1103         \bbl@afterldf}%
1104     \else
1105       \bbl@load@language{#1}%
1106     \fi}%
1107   {}%

```

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

```

1108 \ifnum\bbl@iniflag=\z@ \else
1109   \ifx\bbl@opt@main\@nnil
1110     \ifx\bbl@tempc\relax
1111       \let\bbl@opt@main\bbl@tempb
1112     \else
1113       \let\bbl@opt@main\bbl@tempc
1114     \fi
1115   \fi
1116 \fi
1117 \ifx\bbl@opt@main\@nnil \else
1118   \expandafter
1119   \let\expandafter\bbl@loadmain\csname ds@\bbl@opt@main\endcsname
1120   \expandafter\let\csname ds@\bbl@opt@main\endcsname\@empty
1121 \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  $\TeX$  processes before):

```

1122 \def\AfterBabelLanguage#1{%
1123   \bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang}}{}
1124 \DeclareOption*{}
1125 \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.

```

1126 \bbl@trace{Option 'main'}
1127 \ifx\bbl@opt@main\@nnil
1128   \edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
1129   \let\bbl@tempc\@empty
1130   \bbl@for\bbl@tempb\bbl@tempa{%
1131     \bbl@xin@{\bbl@tempb,}{,\bbl@loaded,}%

```

```

1132 \ifin@vdef\bbl@tempc{\bbl@tempb}\fi}
1133 \def\bbl@tempa#1,#2\@nnil{\def\bbl@tempb{#1}}
1134 \expandafter\bbl@tempa\bbl@loaded,\@nnil
1135 \ifx\bbl@tempb\bbl@tempc\else
1136 \bbl@warning{%
1137     Last declared language option is '\bbl@tempc',\%
1138     but the last processed one was '\bbl@tempb'.\%
1139     The main language cannot be set as both a global\%
1140     and a package option. Use 'main=\bbl@tempc' as\%
1141     option. Reported}%
1142 \fi
1143 \else
1144 \ifodd\bbl@iniflag % case 1,3
1145 \bbl@ldfinit
1146 \let\CurrentOption\bbl@opt@main
1147 \bbl@exp{\bbl@babelprovide[import,main]{\bbl@opt@main}}
1148 \bbl@afterldf}%
1149 \else % case 0,2
1150 \chardef\bbl@iniflag\z@ % Force ldf
1151 \expandafter\let\csname ds@\bbl@opt@main\endcsname\bbl@loadmain
1152 \ExecuteOptions{\bbl@opt@main}
1153 \DeclareOption*{}%
1154 \ProcessOptions*
1155 \fi
1156 \fi
1157 \def\AfterBabelLanguage{%
1158 \bbl@error
1159 {Too late for \string\AfterBabelLanguage}%
1160 {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.

1161 \ifx\bbl@main@language\@undefined
1162 \bbl@info{%
1163     You haven't specified a language. I'll use 'nil'\%
1164     as the main language. Reported}
1165 \bbl@load@language{nil}
1166 \fi
1167 \</package>
1168 \<core>

```

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

The kernel of the babel system is currently stored in babel.def. The file babel.def contains most of the code. 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.

Because plain TeX users might want to use some of the features of the babel system too, care has to be taken that plain 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 TeX and L<sup>A</sup>T<sub>E</sub>X, some of it is for the L<sup>A</sup>T<sub>E</sub>X 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.

### 8.1 Tools

```

1169 \ifx\ldf@quit\@undefined\else

```

```

1170 \endinput\fi % Same line!
1171 <<Make sure ProvidesFile is defined>>
1172 \ProvidesFile{babel.def}[\<date>] \<version>] Babel common definitions]

```

The file `babel.def` expects some definitions made in the  $\text{\TeX}$  style file. So, In  $\text{\TeX}$ 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 are no package options, and therefore an 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).

```

1173 \ifx\AtBeginDocument\@undefined % TODO. change test.
1174 <<Emulate LaTeX>>
1175 \def\language{english}%
1176 \let\bbl@opt@shorthands\@nnil
1177 \def\bbl@ifshorthand#1#2#3{#2}%
1178 \let\bbl@language@opts\@empty
1179 \ifx\babeloptionstrings\@undefined
1180   \let\bbl@opt@strings\@nnil
1181 \else
1182   \let\bbl@opt@strings\babeloptionstrings
1183 \fi
1184 \def\BabelStringsDefault{generic}
1185 \def\bbl@tempa{normal}
1186 \ifx\babeloptionmath\bbl@tempa
1187   \def\bbl@mathnormal{\noexpand\textormath}
1188 \fi
1189 \def\AfterBabelLanguage#1#2{}
1190 \ifx\BabelModifiers\@undefined\let\BabelModifiers\relax\fi
1191 \let\bbl@afterlang\relax
1192 \def\bbl@opt@safe{BR}
1193 \ifx\@uclclist\@undefined\let\@uclclist\@empty\fi
1194 \ifx\bbl@trace\@undefined\def\bbl@trace#1{}\fi
1195 \expandafter\newif\csname ifbbl@single\endcsname
1196 \chardef\bbl@bidimode\z@
1197 \fi

```

Exit immediately with 2.09. An error is raised by the `sty` file, but also try to minimize the number of errors.

```

1198 \ifx\bbl@trace\@undefined
1199   \let\LdfInit\endinput
1200   \def\ProvidesLanguage#1{\endinput}
1201 \endinput\fi % Same line!

```

And continue.

## 9 Multiple languages

This is not a separate file (`switch.def`) anymore.

Plain  $\text{\TeX}$  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.

```

1202 <<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.

```

1203 \def\bbl@version{\<version>}
1204 \def\bbl@date{\<date>}
1205 \def\adddialect#1#2{%
1206   \global\chardef#1#2\relax
1207   \bbl@usehooks{adddialect}{\#1}{\#2}%
1208   \begingroup
1209     \count@#1\relax

```

```

1210 \def\bbl@elt##1##2##3##4{%
1211 \ifnum\count=##2\relax
1212 \bbl@info{\string#1 = using hyphenrules for ##1\%
1213 (\string\language\the\count))}%
1214 \def\bbl@elt####1####2####3####4{%
1215 \fi}%
1216 \bbl@cs{languages}%
1217 \endgroup}

```

\bbl@iflanguage executes code only if the language l@ exists. Otherwise raises and 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.

```

1218 \def\bbl@fixname#1{%
1219 \begingroup
1220 \def\bbl@tempe{l@}%
1221 \edef\bbl@tempd{\noexpand\@ifundefined{\noexpand\bbl@tempe#1}}%
1222 \bbl@tempd
1223 {\lowercase\expandafter{\bbl@tempd}%
1224 {\uppercase\expandafter{\bbl@tempd}%
1225 \@empty
1226 {\edef\bbl@tempd{\def\noexpand#1{#1}}%
1227 \uppercase\expandafter{\bbl@tempd}}}%
1228 {\edef\bbl@tempd{\def\noexpand#1{#1}}%
1229 \lowercase\expandafter{\bbl@tempd}}}%
1230 \@empty
1231 \edef\bbl@tempd{\endgroup\def\noexpand#1{#1}}%
1232 \bbl@tempd
1233 \bbl@exp{\bbl@usehooks{language}{\language}{#1}}%
1234 \def\bbl@iflanguage#1{%
1235 \@ifundefined{l@#1}{\@nolanerr{#1}\@gobble}\@firstofone}

```

After a name has been ‘fixed’, the selectors will try to load the language. If even the fixed name is not defined, will load it on the fly, either based on its name, or if activated, its BCP47 code.

We first need a couple of macros for a simple BCP 47 look up. It also makes sure, with \bbl@bcpcase, casing is the correct one, so that sr-latn-ba becomes fr-Latn-BA. Note #4 may contain some \@empty’s, but they are eventually removed. \bbl@bcpllookup either returns the found ini or it is \relax.

```

1236 \def\bbl@bcpcase#1#2#3#4\@#5{%
1237 \ifx\@empty#3%
1238 \uppercase{\def#5{#1#2}}%
1239 \else
1240 \uppercase{\def#5{#1}}%
1241 \lowercase{\edef#5{#5#2#3#4}}%
1242 \fi}
1243 \def\bbl@bcpllookup#1-#2-#3-#4\@{%
1244 \let\bbl@bcp\relax
1245 \lowercase{\def\bbl@tempa{#1}}%
1246 \ifx\@empty#2%
1247 \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
1248 \else\ifx\@empty#3%
1249 \bbl@bcpcase#2\@empty\@empty\@empty\bbl@tempb
1250 \IfFileExists{babel-\bbl@tempa-\bbl@tempb.ini}%
1251 {\edef\bbl@bcp{\bbl@tempa-\bbl@tempb}}%
1252 {}%
1253 \ifx\bbl@bcp\relax
1254 \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
1255 \fi

```

```

1256 \else
1257   \bbl@bcpcase#2\@empty\@empty\@bbl@tempb
1258   \bbl@bcpcase#3\@empty\@empty\@bbl@tempc
1259   \IfFileExists{babel-\bbl@tempa-\bbl@tempb-\bbl@tempc.ini}%
1260     {\edef\bbl@bcp{\bbl@tempa-\bbl@tempb-\bbl@tempc}}%
1261     {}%
1262   \ifx\bbl@bcp\relax
1263     \IfFileExists{babel-\bbl@tempa-\bbl@tempc.ini}%
1264       {\edef\bbl@bcp{\bbl@tempa-\bbl@tempc}}%
1265       {}%
1266   \fi
1267   \ifx\bbl@bcp\relax
1268     \IfFileExists{babel-\bbl@tempa-\bbl@tempc.ini}%
1269       {\edef\bbl@bcp{\bbl@tempa-\bbl@tempc}}%
1270       {}%
1271   \fi
1272   \ifx\bbl@bcp\relax
1273     \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
1274   \fi
1275 \fi\fi}
1276 \let\bbl@initoload\relax
1277 \def\bbl@provide@locale{%
1278   \ifx\babelprovide\undefined
1279     \bbl@error{For a language to be defined on the fly 'base'\\%
1280               is not enough, and the whole package must be\\%
1281               loaded. Either delete the 'base' option or\\%
1282               request the languages explicitly}%
1283     {See the manual for further details.}%
1284   \fi
1285 % TODO. Option to search if loaded, with \LocaleForEach
1286 \let\bbl@auxname\language % Still necessary. TODO
1287 \bbl@ifunset{bbl@bcp@map@\language}{}% Move uplevel??
1288 {\edef\language{\@nameuse{bbl@bcp@map@\language}}}%
1289 \ifbbl@bcpallowed
1290   \expandafter\ifx\csname date\language\endcsname\relax
1291     \expandafter
1292     \bbl@bcplookup\language-\@empty-\@empty-\@empty\@@
1293     \ifx\bbl@bcp\relax\else % Returned by \bbl@bcplookup
1294       \edef\language{\bbl@bcp@prefix\bbl@bcp}%
1295       \edef\localename{\bbl@bcp@prefix\bbl@bcp}%
1296       \expandafter\ifx\csname date\language\endcsname\relax
1297         \let\bbl@initoload\bbl@bcp
1298         \bbl@exp{\bbl@babelprovide[\bbl@autoload@bcptoptions]{\language}}%
1299         \let\bbl@initoload\relax
1300       \fi
1301       \bbl@csarg\xdef{bcp@map@\bbl@bcp}{\localename}%
1302     \fi
1303   \fi
1304 \fi
1305 \expandafter\ifx\csname date\language\endcsname\relax
1306   \IfFileExists{babel-\language.tex}%
1307   {\bbl@exp{\bbl@babelprovide[\bbl@autoload@options]{\language}}}%
1308   {}%
1309 \fi}

```

`\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.

```

1310 \def\iflanguage#1{%
1311   \bbl@iflanguage{#1}{%
1312     \ifnum\csname l@#1\endcsname=\language
1313       \expandafter\@firstoftwo
1314     \else
1315       \expandafter\@secondoftwo
1316     \fi}}

```

## 9.1 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.

```

1317 \let\bbl@select@type\z@
1318 \edef\selectlanguage{%
1319   \noexpand\protect
1320   \expandafter\noexpand\csname selectlanguage \endcsname}

```

Because the command `\selectlanguage` could be used in a moving argument it expands to `\protect\selectlanguageU`. Therefore, we have to make sure that a macro `\protect` exists. If it doesn't it is `\let` to `\relax`.

```
1321 \ifx\@undefined\protect\let\protect\relax\fi
```

The following definition is preserved for backwards compatibility. It is related to a trick for 2.09.

```
1322 \let\xstring\string
```

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 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.

```
1323 \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`

```

1324 \def\bbl@push@language{%
1325   \ifx\language\@undefined\else
1326     \xdef\bbl@language@stack{\language+\bbl@language@stack}%
1327   \fi}

```

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 in `\bbl@language@stack`.

```

1328 \def\bbl@pop@lang#1+#2\@@{%
1329   \edef\language{#1}%
1330   \xdef\bbl@language@stack{#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).

```
1331 \let\bbl@ifrestoring\@secondoftwo
1332 \def\bbl@pop@language{%
1333   \expandafter\bbl@pop@lang\bbl@language@stack\@@
1334   \let\bbl@ifrestoring\@firstoftwo
1335   \expandafter\bbl@set@language\expandafter{\language}%
1336   \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 (so that two locales can share the same rules).

```
1337 \chardef\localeid\z@
1338 \def\bbl@id@last{0} % No real need for a new counter
1339 \def\bbl@id@assign{%
1340   \bbl@ifunset{bbl@id@\language}%
1341   {\count@bbl@id@last\relax
1342     \advance\count@\@ne
1343     \bbl@csarg\chardef{id@\language}\count@
1344     \edef\bbl@id@last{\the\count@}%
1345     \ifcase\bbl@engine\or
1346       \directlua{
1347         Babel = Babel or {}
1348         Babel.locale_props = Babel.locale_props or {}
1349         Babel.locale_props[\bbl@id@last] = {}
1350         Babel.locale_props[\bbl@id@last].name = '\language'
1351       }%
1352     \fi}%
1353   }%
1354   \chardef\localeid\bbl@c{id@}}
```

The unprotected part of `\selectlanguage`.

```
1355 \expandafter\def\csname selectlanguage \endcsname#1{%
1356   \ifnum\bbl@hymapset=\@cclv\let\bbl@hymapset\tw@fi
1357   \bbl@push@language
1358   \aftergroup\bbl@pop@language
1359   \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.

```
1360 \def\BabelContentsFiles{toc,lof,lot}
1361 \def\bbl@set@language#1{% from selectlanguage, pop@
1362   % The old buggy way. Preserved for compatibility.
1363   \edef\language{%
1364     \ifnum\escapechar=\expandafter`\string#1\@empty
1365     \else\string#1\@empty\fi}%

```

```

1366 \ifcat\relax\noexpand#1%
1367 \expandafter\ifx\csname date\language\endcsname\relax
1368 \edef\language{#1}%
1369 \let\localname\language
1370 \else
1371 \bbl@info{Using '\string\language' instead of 'language' is\\%
1372 deprecated. If what you want is to use a\\%
1373 macro containing the actual locale, make\\%
1374 sure it does not not match any language.\\%
1375 Reported}%
1376 % I'll\\%
1377 % try to fix '\string\localname', but I cannot promise\\%
1378 % anything. Reported}%
1379 \ifx\scantokens\@undefined
1380 \def\localname{??}%
1381 \else
1382 \scantokens\expandafter{\expandafter
1383 \def\expandafter\localname\expandafter{\language}}%
1384 \fi
1385 \fi
1386 \else
1387 \def\localname{#1}% This one has the correct catcodes
1388 \fi
1389 \select@language{\language}%
1390 % write to aux
1391 \expandafter\ifx\csname date\language\endcsname\relax\else
1392 \if@files
1393 \ifx\babel@aux\@gobbletwo\else % Set if single in the first, redundant
1394 % \bbl@savelastskip
1395 \protected@write\@auxout{\string\babel@aux{\bbl@auxname}}}%
1396 % \bbl@restorelastskip
1397 \fi
1398 \bbl@usehooks{write}}%
1399 \fi
1400 \fi}
1401 % The following is used above to deal with skips before the write
1402 % whatsit. Adapted from hyperref, but it might fail, so for the moment
1403 % it's not activated. TODO.
1404 \def\bbl@savelastskip{%
1405 \let\bbl@restorelastskip\relax
1406 \ifvmode
1407 \ifdim\lastskip=\z@
1408 \let\bbl@restorelastskip\nobreak
1409 \else
1410 \bbl@exp{%
1411 \def\\bbl@restorelastskip{%
1412 \skip@=\the\lastskip
1413 \\nobreak \vskip-\skip@ \vskip\skip@}}%
1414 \fi
1415 \fi}
1416 \newif\ifbbl@bcpallowed
1417 \bbl@bcpallowedfalse
1418 \def\select@language#1{% from set@, babel@aux
1419 % set hymap
1420 \ifnum\bbl@hymapset=\@cclv\chardef\bbl@hymapset4\relax\fi
1421 % set name
1422 \edef\language{#1}%
1423 \bbl@fixname\language
1424 % TODO. name@map must be here?

```





```

1471 \fi
1472 \bbl@xin@{,date,}{,\bbl@select@opts,}%
1473 \ifin@ % if \foreign... within \<lang>date
1474 \csname date#1\endcsname\relax
1475 \fi
1476 \fi
1477 \bbl@esphack
1478 % switch extras
1479 \bbl@usehooks{beforeextras}{}%
1480 \csname extras#1\endcsname\relax
1481 \bbl@usehooks{afterextras}{}%
1482 % > babel-ensure
1483 % > babel-sh-<short>
1484 % > babel-bidi
1485 % > babel-fontspec
1486 % hyphenation - case mapping
1487 \ifcase\bbl@opt@hyphenmap\or
1488 \def\BabelLower##1##2{\lccode##1=##2\relax}%
1489 \ifnum\bbl@hymapsel>4\else
1490 \csname\language\name @bbl@hyphenmap\endcsname
1491 \fi
1492 \chardef\bbl@opt@hyphenmap\z@
1493 \else
1494 \ifnum\bbl@hymapsel>\bbl@opt@hyphenmap\else
1495 \csname\language\name @bbl@hyphenmap\endcsname
1496 \fi
1497 \fi
1498 \let\bbl@hymapsel\@cclv
1499 % hyphenation - select patterns
1500 \bbl@patterns{#1}%
1501 % hyphenation - allow stretching with babelnohyphens
1502 \ifnum\language=\l@babelnohyphens
1503 \babel@savevariable\emergencystretch
1504 \emergencystretch\maxdimen
1505 \babel@savevariable\hbadness
1506 \hbadness\@M
1507 \fi
1508 % hyphenation - mins
1509 \babel@savevariable\lefthyphenmin
1510 \babel@savevariable\righthyphenmin
1511 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
1512 \set@hyphenmins\tw@\thr@\@relax
1513 \else
1514 \expandafter\expandafter\expandafter\set@hyphenmins
1515 \csname #1hyphenmins\endcsname\relax
1516 \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.

```

1517 \long\def\otherlanguage#1{%
1518 \ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel\thr@\@fi
1519 \csname selectlanguage \endcsname{#1}%
1520 \ignorespaces}

```

The `\endotherlanguage` part of the environment tries to hide itself when it is called in horizontal

mode.

```
1521 \long\def\endotherlanguage{%
1522   \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`.

```
1523 \expandafter\def\csname otherlanguage*\endcsname{%
1524   \@ifnextchar[\bbl@otherlanguage@s{\bbl@otherlanguage@s{}}
1525   \def\bbl@otherlanguage@s[#1]#2{%
1526     \ifnum\bbl@hymapsel=\@cc1v\chardef\bbl@hymapsel4\relax\fi
1527     \def\bbl@select@opts{#1}%
1528     \foreign@language{#2}}
```

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”.

```
1529 \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`.

```
1530 \providecommand\bbl@beforeforeign{}
1531 \edef\foreignlanguage{%
1532   \noexpand\protect
1533   \expandafter\noexpand\csname foreignlanguage \endcsname}
1534 \expandafter\def\csname foreignlanguage \endcsname{%
1535   \@ifstar\bbl@foreign@s\bbl@foreign@x}
1536 \providecommand\bbl@foreign@x[3][]{%
1537   \begingroup
1538     \def\bbl@select@opts{#1}%
1539     \let\BabelText\@firstofone
1540     \bbl@beforeforeign
1541     \foreign@language{#2}%
1542     \bbl@usehooks{foreign}{}%
1543     \BabelText{#3}% Now in horizontal mode!
1544   \endgroup}
1545 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \@setpar, ?\@par
1546   \begingroup
1547     {\par}%
1548     \let\bbl@select@opts\@empty
```

```

1549 \let\BabelText\@firstofone
1550 \foreign@language{#1}%
1551 \bbl@usehooks{foreign*}{}%
1552 \bbl@dirparastext
1553 \BabelText{#2}% Still in vertical mode!
1554 {\par}%
1555 \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`.

```

1556 \def\foreign@language#1{%
1557 % set name
1558 \edef\language{#1}%
1559 \ifbbl@usedategroup
1560 \bbl@add\bbl@select@opts{,date,}%
1561 \bbl@usedategroupfalse
1562 \fi
1563 \bbl@fixname\language
1564 % TODO. name@map here?
1565 \bbl@provide@locale
1566 \bbl@iflanguage\language{%
1567 \expandafter\ifx\csname date\language\endcsname\relax
1568 \bbl@warning % TODO - why a warning, not an error?
1569 {Unknown language `#1'. Either you have\\%
1570 misspelled its name, it has not been installed,\\%
1571 or you requested it in a previous run. Fix its name,\\%
1572 install it or just rerun the file, respectively. In\\%
1573 some cases, you may need to remove the aux file.\\%
1574 I'll proceed, but expect wrong results.\\%
1575 Reported}%
1576 \fi
1577 % set type
1578 \let\bbl@select@type\@ne
1579 \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 `\lcode`'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.

```

1580 \let\bbl@hyphlist\@empty
1581 \let\bbl@hyphenation\relax
1582 \let\bbl@pttnlist\@empty
1583 \let\bbl@patterns\relax
1584 \let\bbl@hymapset\@cclv
1585 \def\bbl@patterns#1{%
1586 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
1587 \csname l@#1\endcsname
1588 \edef\bbl@tempa{#1}%
1589 \else
1590 \csname l@#1:\f@encoding\endcsname
1591 \edef\bbl@tempa{#1:\f@encoding}%
1592 \fi
1593 \@expandtwoargs\bbl@usehooks{patterns}{#{1}}{\bbl@tempa}}%

```

```

1594 % > luatex
1595 \@ifundefined{bbl@hyphenation@}{}{% Can be \relax!
1596   \begingroup
1597     \bbl@xin@{\number\language,}{,\bbl@hyphlist}%
1598     \ifin\else
1599       \@expandtwoargs\bbl@usehooks{hyphenation}{\#1}{\bbl@tempa}}%
1600     \hyphenation{%
1601       \bbl@hyphenation@
1602       \@ifundefined{bbl@hyphenation@#1}%
1603       \@empty
1604       {\space\csname bbl@hyphenation@#1\endcsname}}%
1605     \xdef\bbl@hyphlist{\bbl@hyphlist\number\language,}%
1606   \fi
1607 \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 `otherlanguage*`.

```

1608 \def\hyphenrules#1{%
1609   \edef\bbl@tempf{#1}%
1610   \bbl@fixname\bbl@tempf
1611   \bbl@iflanguage\bbl@tempf{%
1612     \expandafter\bbl@patterns\expandafter{\bbl@tempf}%
1613     \ifx\languageshorthands\@undefined\else
1614       \languageshorthands{none}%
1615     \fi
1616     \expandafter\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
1617       \set@hyphenmins\tw@\thr@\relax
1618     \else
1619       \expandafter\expandafter\expandafter\set@hyphenmins
1620       \csname\bbl@tempf hyphenmins\endcsname\relax
1621     \fi}}
1622 \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.

```

1623 \def\providehyphenmins#1#2{%
1624   \expandafter\ifx\csname #1hyphenmins\endcsname\relax
1625     \@namedef{#1hyphenmins}{#2}%
1626   \fi}

```

**\set@hyphenmins** This macro sets the values of `\lefthyphenmin` and `\righthyphenmin`. It expects two values as its argument.

```

1627 \def\set@hyphenmins#1#2{%
1628   \lefthyphenmin#1\relax
1629   \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.

```

1630 \ifx\ProvidesFile\@undefined
1631   \def\ProvidesLanguage#1[#2 #3 #4]{%
1632     \wlog{Language: #1 #4 #3 <#2>}%
1633   }
1634 \else

```

```

1635 \def\ProvidesLanguage#1{%
1636     \begingroup
1637     \catcode`\ 10 %
1638     \@makeother\/%
1639     \@ifnextchar[%]
1640         {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}
1641 \def\@provideslanguage#1[#2]{%
1642     \wlog{Language: #1 #2}%
1643     \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
1644     \endgroup}
1645 \fi

```

`\originalTeX` The macro `\originalTeX` should be known to  $\TeX$  at this moment. As it has to be expandable we `\let` it to `\@empty` instead of `\relax`.

```
1646 \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 initializes the save mechanism, `\babel@beginsave`, is not considered to be undefined.

```
1647 \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’:

```

1648 \providecommand\setlocale{%
1649     \bbl@error
1650     {Not yet available}%
1651     {Find an armchair, sit down and wait}}
1652 \let\uselocale\setlocale
1653 \let\locale\setlocale
1654 \let\selectlocale\setlocale
1655 \let\localename\setlocale
1656 \let\textlocale\setlocale
1657 \let\textlanguage\setlocale
1658 \let\languagetext\setlocale

```

## 9.2 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  $\LaTeX 2\epsilon$ , so we can safely use its error handling interface. Otherwise we’ll have to ‘keep it simple’.  
Infos are not written to the console, but on the other hand many people think warnings are errors, so a further message type is defined: an important info which is sent to the console.

```

1659 \edef\bbl@nulllanguage{\string\language=0}
1660 \ifx\PackageError\undefined % TODO. Move to Plain
1661     \def\bbl@error#1#2{%
1662         \begingroup
1663         \newlinechar=`^^J
1664         \def\{^^J(babel) }%
1665         \errhelp{#2}\errmessage{\{#1}%
1666         \endgroup}
1667 \def\bbl@warning#1{%
1668     \begingroup
1669     \newlinechar=`^^J
1670     \def\{^^J(babel) }%

```

```

1671     \message{\#1}%
1672   \endgroup}
1673   \let\bbl@infowarn\bbl@warning
1674   \def\bbl@info#1{%
1675     \begingroup
1676       \newlinechar=`^^J
1677       \def\{^^J}%
1678       \wlog{#1}%
1679     \endgroup}
1680 \fi
1681 \def\bbl@nocaption{\protect\bbl@nocaption@i}
1682 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
1683   \global\@namedef{#2}{\textbf{?#1?}}%
1684   \@nameuse{#2}%
1685   \edef\bbl@tempa{#1}%
1686   \bbl@sreplace\bbl@tempa{name}{}%
1687   \bbl@warning{% TODO.
1688     \@backslashchar#1 not set for '\language'. Please,\%
1689     define it after the language has been loaded\%
1690     (typically in the preamble) with:\%
1691     \string\setlocalecaption{\language}{\bbl@tempa}{..\%
1692     Reported}}
1693 \def\bbl@tentative{\protect\bbl@tentative@i}
1694 \def\bbl@tentative@i#1{%
1695   \bbl@warning{%
1696     Some functions for '#1' are tentative.\%
1697     They might not work as expected and their behavior\%
1698     could change in the future.\%
1699     Reported}}
1700 \def\@nolanerr#1{%
1701   \bbl@error
1702   {You haven't defined the language #1\space yet.\%
1703     Perhaps you misspelled it or your installation\%
1704     is not complete}%
1705   {Your command will be ignored, type <return> to proceed}}
1706 \def\@nopatterns#1{%
1707   \bbl@warning
1708   {No hyphenation patterns were preloaded for\%
1709     the language '#1' into the format.\%
1710     Please, configure your TeX system to add them and\%
1711     rebuild the format. Now I will use the patterns\%
1712     preloaded for \bbl@nulllanguage\space instead}}
1713 \let\bbl@usehooks\@gobbletwo
1714 \ifx\bbl@onlyswitch\@empty\endinput\fi
1715 % Here ended switch.def

    Here ended switch.def.

1716 \ifx\directlua\@undefined\else
1717   \ifx\bbl@luapatterns\@undefined
1718     \input luabel.def
1719   \fi
1720 \fi
1721 <<Basic macros>>
1722 \bbl@trace{Compatibility with language.def}
1723 \ifx\bbl@languages\@undefined
1724   \ifx\directlua\@undefined
1725     \openin1 = language.def % TODO. Remove hardcoded number
1726     \ifeof1
1727       \closein1

```

```

1728     \message{I couldn't find the file language.def}
1729   \else
1730     \closein1
1731     \begingroup
1732     \def\addlanguage#1#2#3#4#5{%
1733       \expandafter\ifx\csname lang@#1\endcsname\relax\else
1734         \global\expandafter\let\csname l@#1\expandafter\endcsname
1735           \csname lang@#1\endcsname
1736       \fi}%
1737     \def\uselanguage#1{%
1738       \input language.def
1739     \endgroup
1740   \fi
1741 \fi
1742 \chardef\l@english\z@
1743 \fi

```

`\addto` 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. Note there is an inconsistency, because the assignment in the last branch is global.

```

1744 \def\addto#1#2{%
1745   \ifx#1\undefined
1746     \def#1{#2}%
1747   \else
1748     \ifx#1\relax
1749       \def#1{#2}%
1750     \else
1751       {\toks@\expandafter{#1#2}%
1752        \xdef#1{\the\toks@}}%
1753     \fi
1754   \fi}

```

The macro `\initiate@active@char` below takes all the necessary actions to make its argument a shorthand character. The real work is performed once for each character. But first we define a little tool. `TODO`. Always used with additional expansions. Move them here? Move the macro to basic?

```

1755 \def\bbl@withactive#1#2{%
1756   \begingroup
1757   \lccode`~=`#2\relax
1758   \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 T<sub>E</sub>X macros completely in case their definitions change (they have changed in the past). A macro named `\macro` will be saved new control sequences named `\org@macro`.

```

1759 \def\bbl@redefine#1{%
1760   \edef\bbl@tempa{\bbl@stripslash#1}%
1761   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
1762   \expandafter\def\csname\bbl@tempa\endcsname}
1763 \@onlypreamble\bbl@redefine

```

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

```

1764 \def\bbl@redefine@long#1{%
1765   \edef\bbl@tempa{\bbl@stripslash#1}%
1766   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
1767   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
1768 \@onlypreamble\bbl@redefine@long

```



`\bbl@redefinero bust` 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`.

```
1769 \def\bbl@redefinero bust#1{%
1770   \edef\bbl@tempa{\bbl@stripslash#1}%
1771   \bbl@ifunset{\bbl@tempa\space}%
1772   {\expandafter\let\csname org\bbl@tempa\endcsname#1%
1773    \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
1774   {\bbl@exp{\let\<org\bbl@tempa>\<\bbl@tempa\space>}}}%
1775   \@namedef{\bbl@tempa\space}}
1776 \@onlypreamble\bbl@redefinero bust
```

### 9.3 Hooks

Admittedly, the current implementation is a somewhat simplistic and does very little to catch errors, but it is meant for developers, after all. `\bbl@usehooks` is the commands used by babel to execute hooks defined for an event.

```
1777 \bbl@trace{Hooks}
1778 \newcommand\AddBabelHook[3][{}]{%
1779   \bbl@ifunset{\bbl@hk@#2}{\EnableBabelHook{#2}}{}%
1780   \def\bbl@tempa##1,##2,##3\@empty{\def\bbl@tempb{##2}}%
1781   \expandafter\bbl@tempa\bbl@evargs,##3,\@empty
1782   \bbl@ifunset{\bbl@ev@#2@#3@#1}%
1783   {\bbl@csarg\bbl@add{\ev@#3@#1}{\bbl@elth{#2}}}%
1784   {\bbl@csarg\let{\ev@#2@#3@#1}\relax}%
1785   \bbl@csarg\newcommand{\ev@#2@#3@#1}{\bbl@tempb}}
1786 \newcommand\EnableBabelHook[1]{\bbl@csarg\let{\hk@#1}\@firstofone}
1787 \newcommand\DisableBabelHook[1]{\bbl@csarg\let{\hk@#1}\@gobble}
1788 \def\bbl@usehooks#1#2{%
1789   \def\bbl@elth##1{%
1790     \bbl@cs{\hk@##1}{\bbl@cs{\ev@##1@#1}{#2}}%
1791     \bbl@cs{\ev@#1@}%
1792     \ifx\language\@undefined\else % Test required for Plain (?)
1793       \def\bbl@elth##1{%
1794         \bbl@cs{\hk@##1}{\bbl@cl{\ev@##1@#1}{#2}}%
1795         \bbl@cl{\ev@#1}%
1796       \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).

```
1797 \def\bbl@evargs{,% <- don't delete this comma
1798   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
1799   adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
1800   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
1801   hyphenation=2,initiateactive=3,afterreset=0,foreign=0,foreign*=0,%
1802   beforestart=0,language=2}
```

`\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.

```

1803 \bbl@trace{Defining babelensure}
1804 \newcommand\babelensure[2][{}]{% TODO - revise test files
1805   \AddBabelHook{babel-ensure}{afterextras}{%
1806     \ifcase\bbl@select@type
1807       \bbl@cl{e}%
1808     \fi}%
1809   \beginingroup
1810     \let\bbl@ens@include\@empty
1811     \let\bbl@ens@exclude\@empty
1812     \def\bbl@ens@fontenc{\relax}%
1813     \def\bbl@tempb##1{%
1814       \ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}%
1815     \edef\bbl@tempa{\bbl@tempb##1\@empty}%
1816     \def\bbl@tempb##1=##2\@{\@namedef{bbl@ens@##1}{##2}}%
1817     \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
1818     \def\bbl@tempc{\bbl@ensure}%
1819     \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
1820       \expandafter{\bbl@ens@include}}%
1821     \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
1822       \expandafter{\bbl@ens@exclude}}%
1823     \toks@\expandafter{\bbl@tempc}%
1824     \bbl@exp{%
1825   \endgroup
1826   \def\<bbl@e@#2>{\the\toks@{\bbl@ens@fontenc}}}%
1827 \def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
1828   \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
1829     \ifx##1\undefined % 3.32 - Don't assume the macro exists
1830       \edef##1{\noexpand\bbl@nocaption
1831         {\bbl@stripslash##1}{\language\name\bbl@stripslash##1}}%
1832     \fi
1833     \ifx##1\@empty\else
1834       \in@{##1}{#2}%
1835       \ifin@ \else
1836         \bbl@ifunset{bbl@ensure@\language\name}%
1837         {\bbl@exp{%
1838           \\\DeclareRobustCommand\<bbl@ensure@\language\name>[1]{%
1839             \\\foreignlanguage{\language\name}%
1840             {\ifx\relax#3\else
1841               \\\fontencoding{#3}\selectfont
1842               \fi
1843             #####1}}}%
1844         }%
1845         \toks@\expandafter{##1}%
1846         \edef##1{%
1847           \bbl@csarg\noexpand{ensure@\language\name}%
1848           {\the\toks@}}%
1849       \fi
1850       \expandafter\bbl@tempb
1851     \fi}%
1852   \expandafter\bbl@tempb\bbl@captionslist\today\@empty
1853   \def\bbl@tempa##1{% elt for include list
1854     \ifx##1\@empty\else
1855       \bbl@csarg\in@{ensure@\language\name\expandafter}\expandafter{##1}%
1856     \ifin@ \else
1857       \bbl@tempb##1\@empty
1858     \fi
1859     \expandafter\bbl@tempa
1860     \fi}%
1861   \bbl@tempa#1\@empty}

```

```

1862 \def\bbl@captionslist{%
1863   \prefacename\refname\abstractname\bibname\chaptername\appendixname
1864   \contentsname\listfigurename\listtablename\indexname\figurename
1865   \tablename\partname\enclname\ccname\headtoname\pagename\seenname
1866   \alsiname\proofname\glossaryname}

```

## 9.4 Setting up language files

**\LdfInit**    \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 @-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.

```

1867 \bbl@trace{Macros for setting language files up}
1868 \def\bbl@ldfinit{%
1869   \let\bbl@screset\@empty
1870   \let\BabelStrings\bbl@opt@string
1871   \let\BabelOptions\@empty
1872   \let\BabelLanguages\relax
1873   \ifx\originalTeX\@undefined
1874     \let\originalTeX\@empty
1875   \else
1876     \originalTeX
1877   \fi}
1878 \def\LdfInit#1#2{%
1879   \chardef\atcatcode=\catcode`\@
1880   \catcode`\@=11\relax
1881   \chardef\eqcatcode=\catcode`\=
1882   \catcode`\==12\relax
1883   \expandafter\if\expandafter\@backslashchar
1884     \expandafter\@car\string#2\@nil
1885   \ifx#2\@undefined\else
1886     \ldf@quit{#1}%
1887   \fi
1888 \else
1889   \expandafter\ifx\csname#2\endcsname\relax\else
1890     \ldf@quit{#1}%
1891   \fi
1892 \fi
1893 \bbl@ldfinit}

```

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

```

1894 \def\ldf@quit#1{%
1895   \expandafter\main@language\expandafter{#1}%
1896   \catcode`\@=\atcatcode \let\atcatcode\relax

```

```

1897 \catcode`\==\eqcatcode \let\eqcatcode\relax
1898 \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.

```

1899 \def\bbl@afterldf#1{% TODO. Merge into the next macro? Unused elsewhere
1900 \bbl@afterlang
1901 \let\bbl@afterlang\relax
1902 \let\BabelModifiers\relax
1903 \let\bbl@screset\relax}%
1904 \def\ldf@finish#1{%
1905 \ifx\loadlocalcfg@undefined\else % For LaTeX 209
1906 \loadlocalcfg{#1}%
1907 \fi
1908 \bbl@afterldf{#1}%
1909 \expandafter\main@language\expandafter{#1}%
1910 \catcode`\@=\atcatcode \let\atcatcode\relax
1911 \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 `ltxex`.

```

1912 \@onlypreamble\LdfInit
1913 \@onlypreamble\ldf@quit
1914 \@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.

```

1915 \def\main@language#1{%
1916 \def\bbl@main@language{#1}%
1917 \let\language\bbl@main@language % TODO. Set localename
1918 \bbl@id@assign
1919 \bbl@patterns{\language}}

```

We also have to make sure that some code gets executed at the beginning of the document, either when the aux file is read or, if it does not exist, when the `\AtBeginDocument` is executed. Languages do not set `\pagedir`, so we set here for the whole document to the main `\bodydir`.

```

1920 \def\bbl@beforestart{%
1921 \bbl@usehooks{beforestart}}}%
1922 \global\let\bbl@beforestart\relax}
1923 \AtBeginDocument{%
1924 \@nameuse{bbl@beforestart}%
1925 \if@filesw
1926 \providecommand\babel@aux[2]{}%
1927 \immediate\write\@mainaux{%
1928 \string\providecommand\string\babel@aux[2]{}%
1929 \immediate\write\@mainaux{\string\@nameuse{bbl@beforestart}}}%
1930 \fi
1931 \expandafter\selectlanguage\expandafter{\bbl@main@language}%
1932 \ifbbl@single % must go after the line above.
1933 \renewcommand\selectlanguage[1]{}%
1934 \renewcommand\foreignlanguage[2]{#2}%
1935 \global\let\babel@aux\@gobbles % Also as flag
1936 \fi
1937 \ifcase\bbl@engine\or\pagedir\bodydir\fi} % TODO - a better place

```

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

```

1938 \def\select@language#1{%
1939   \ifcase\bbbl@select@type
1940     \bbbl@ifsamestring\language#1\{\select@language{#1}}%
1941   \else
1942     \select@language{#1}%
1943   \fi}

```

## 9.5 Shorthands

**\bbbl@add@special** The macro `\bbbl@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.

```

1944 \bbbl@trace{Shorhands}
1945 \def\bbbl@add@special#1{% 1:a macro like "\", "\?", etc.
1946   \bbbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
1947   \bbbl@ifunset{@sanitize}\{\bbbl@add\@sanitize{\@makeother#1}}%
1948   \ifx\nfss@catcodes\undefined\else % TODO - same for above
1949     \begingroup
1950       \catcode`#1\active
1951       \nfss@catcodes
1952       \ifnum\catcode`#1=\active
1953         \endgroup
1954         \bbbl@add\nfss@catcodes{\@makeother#1}%
1955       \else
1956         \endgroup
1957     \fi
1958   \fi}

```

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

```

1959 \def\bbbl@remove@special#1{%
1960   \begingroup
1961     \def\x##1##2{\ifnum`#1=`##2\noexpand\@empty
1962       \else\noexpand##1\noexpand##2\fi}%
1963     \def\do{\x\do}%
1964     \def\@makeother{\x\@makeother}%
1965     \edef\x{\endgroup
1966       \def\noexpand\dospecials{\dospecials}%
1967       \expandafter\ifx\curname @sanitize\endcsname\relax\else
1968         \def\noexpand\@sanitize{\@sanitize}%
1969       \fi}%
1970     \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 `\bbbl@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).

```

1971 \def\bbl@active@def#1#2#3#4{%
1972   \namedef{#3#1}{%
1973     \expandafter\ifx\csname#2@sh@#1\endcsname\relax
1974       \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
1975     \else
1976       \bbl@afterfi\csname#2@sh@#1\endcsname
1977     \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.

```

1978   \long\@namedef{#3@arg#1}##1{%
1979     \expandafter\ifx\csname#2@sh@#1\string##1\endcsname\relax
1980       \bbl@afterelse\csname#4#1\endcsname##1%
1981     \else
1982       \bbl@afterfi\csname#2@sh@#1\string##1\endcsname
1983     \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.

```

1984 \def\@initiate@active@char#1{%
1985   \bbl@ifunset{active@char\string#1}%
1986   {\bbl@withactive
1987    {\expandafter\@initiate@active@char\expandafter}#1\string#1}%
1988   {}}

```

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`).

```

1989 \def\@initiate@active@char#1#2#3{%
1990   \bbl@csarg\edef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%
1991   \ifx#1\@undefined
1992     \bbl@csarg\edef{oridef@#2}{\let\noexpand#1\noexpand\@undefined}%
1993   \else
1994     \bbl@csarg\let{oridef@#2}#1%
1995     \bbl@csarg\edef{oridef@#2}{%
1996       \let\noexpand#1%
1997       \expandafter\noexpand\csname bbl@oridef@@#2\endcsname}%
1998   \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*”).

```

1999   \ifx#1#3\relax
2000     \expandafter\let\csname normal@char#2\endcsname#3%
2001   \else
2002     \bbl@info{Making #2 an active character}%
2003     \ifnum\mathcode`#2=\ifodd\bbl@engine"1000000 \else"8000 \fi
2004     \@namedef{normal@char#2}{%
2005       \textormath{#3}{\csname bbl@oridef@@#2\endcsname}}%
2006   \else

```

```

2007 \namedef{normal@char#2}{#3}%
2008 \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).

```

2009 \bbl@restoreactive{#2}%
2010 \AtBeginDocument{%
2011 \catcode`#2\active
2012 \if@filesw
2013 \immediate\write\@mainaux{\catcode`\string#2\active}%
2014 \fi}%
2015 \expandafter\bbl@add@special\csname#2\endcsname
2016 \catcode`#2\active
2017 \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}).

```

2018 \let\bbl@tempa\@firstoftwo
2019 \if\string^#2%
2020 \def\bbl@tempa{\noexpand\textormath}%
2021 \else
2022 \ifx\bbl@mathnormal\@undefined\else
2023 \let\bbl@tempa\bbl@mathnormal
2024 \fi
2025 \fi
2026 \expandafter\edef\csname active@char#2\endcsname{%
2027 \bbl@tempa
2028 {\noexpand\if@safe@actives
2029 \noexpand\expandafter
2030 \expandafter\noexpand\csname normal@char#2\endcsname
2031 \noexpand\else
2032 \noexpand\expandafter
2033 \expandafter\noexpand\csname bbl@doactive#2\endcsname
2034 \noexpand\fi}%
2035 {\expandafter\noexpand\csname normal@char#2\endcsname}}%
2036 \bbl@csarg\edef{doactive#2}{%
2037 \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!).

```

2038 \bbl@csarg\edef{active@#2}{%
2039 \noexpand\active@prefix\noexpand#1%
2040 \expandafter\noexpand\csname active@char#2\endcsname}%
2041 \bbl@csarg\edef{normal@#2}{%
2042 \noexpand\active@prefix\noexpand#1%
2043 \expandafter\noexpand\csname normal@char#2\endcsname}%
2044 \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.

```
2045 \bbl@active@def#2\user@group{user@active}{language@active}%
2046 \bbl@active@def#2\language@group{language@active}{system@active}%
2047 \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.

```
2048 \expandafter\edef\csname\user@group @sh#2@@\endcsname
2049 {\expandafter\noexpand\csname normal@char#2\endcsname}%
2050 \expandafter\edef\csname\user@group @sh#2@\string\protect\endcsname
2051 {\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.

```
2052 \if\string'#2%
2053 \let\prim@s\bbl@prim@s
2054 \let\active@math@prime#1%
2055 \fi
2056 \bbl@usehooks{initiateactive}{\{#1\}{#2\}{#3\}}
```

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

```
2057 <<(*More package options)>> ≡
2058 \DeclareOption{math=active}{}
2059 \DeclareOption{math=normal}{{\def\bbl@mathnormal{\noexpand\textormath}}}
2060 <</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* the end of the ldf.

```
2061 \@ifpackagewith{babel}{KeepShorthandsActive}%
2062 {\let\bbl@restoreactive\@gobble}%
2063 {\def\bbl@restoreactive#1{%
2064 \bbl@exp{%
2065 \\\AfterBabelLanguage\\CurrentOption
2066 {\catcode`#1=\the\catcode`#1\relax}%
2067 \\\AtEndOfPackage
2068 {\catcode`#1=\the\catcode`#1\relax}}}%
2069 \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.

```
2070 \def\bbl@sh@select#1#2{%
2071 \expandafter\ifx\csname#1@sh#2@sel\endcsname\relax
2072 \bbl@afterelse\bbl@scndcs
2073 \else
2074 \bbl@afterfi\csname#1@sh#2@sel\endcsname
2075 \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 `\protect`s 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.

```

2076 \begingroup
2077 \bbl@ifunset{ifincsname}% TODO. Ugly. Correct?
2078 {\gdef\active@prefix#1{%
2079   \ifx\protect\@typeset@protect
2080   \else
2081     \ifx\protect\@unexpandable@protect
2082       \noexpand#1%
2083     \else
2084       \protect#1%
2085     \fi
2086     \expandafter\@gobble
2087   \fi}}
2088 {\gdef\active@prefix#1{%
2089   \ifincsname
2090     \string#1%
2091     \expandafter\@gobble
2092   \else
2093     \ifx\protect\@typeset@protect
2094     \else
2095       \ifx\protect\@unexpandable@protect
2096         \noexpand#1%
2097       \else
2098         \protect#1%
2099       \fi
2100       \expandafter\expandafter\expandafter\@gobble
2101     \fi
2102   \fi}}
2103 \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⟩`.

```

2104 \newif\if@safe@actives
2105 \@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.

```

2106 \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 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`.

```

2107 \def\bbl@activate#1{%
2108   \bbl@withactive{\expandafter\let\expandafter}#1%
2109   \csname bbl@active@\string#1\endcsname}
2110 \def\bbl@deactivate#1{%
2111   \bbl@withactive{\expandafter\let\expandafter}#1%
2112   \csname bbl@normal@\string#1\endcsname}

```

`\bbl@firstcs` These macros are used only as a trick when declaring shorthands.

```

\bbl@scndcs
2113 \def\bbl@firstcs#1#2{\csname#1\endcsname}
2114 \def\bbl@scndcs#1#2{\csname#2\endcsname}

```

`\declare@shorthand` The command `\declare@shorthand` is used to declare a shorthand on a certain level. It takes three arguments:

1. a name for the collection of shorthands, i.e. ‘system’, or ‘dutch’;
2. the character (sequence) that makes up the shorthand, i.e. ~ or "a;
3. the code to be executed when the shorthand is encountered.

The auxiliary macro `\babel@texpdf` improves the interoperativity with `hyperref` and takes 4 arguments: (1) The  $\TeX$  code in text mode, (2) the string for `hyperref`, (3) the  $\TeX$  code in math mode, and (4), which is currently ignored, but it's meant for a string in math mode, like a minus sign instead of an hyphen (currently `hyperref` doesn't discriminate the mode). This macro may be used in `ldf` files.

```

2115 \def\babel@texpdf#1#2#3#4{%
2116   \ifx\texorpdfstring\@undefined
2117     \textormath{#1}{#2}%
2118   \else
2119     \texorpdfstring{\textormath{#1}{#3}}{#2}%
2120     % \texorpdfstring{\textormath{#1}{#3}}{\textormath{#2}{#4}}%
2121   \fi}
2122 %
2123 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil}
2124 \def\@decl@short#1#2#3\@nil#4{%
2125   \def\bbl@tempa{#3}%
2126   \ifx\bbl@tempa\@empty
2127     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs
2128     \bbl@ifunset{#1@sh@\string#2@}{}%
2129     {\def\bbl@tempa{#4}%
2130      \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa
2131      \else
2132        \bbl@info
2133        {Redefining #1 shorthand \string#2\\%
2134         in language \CurrentOption}%
2135      \fi}%
2136     \@namedef{#1@sh@\string#2@}{#4}%
2137   \else
2138     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs
2139     \bbl@ifunset{#1@sh@\string#2@\string#3@}{}%
2140     {\def\bbl@tempa{#4}%
2141      \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
2142      \else
2143        \bbl@info
2144        {Redefining #1 shorthand \string#2\string#3\\%
2145         in language \CurrentOption}%
2146      \fi}%
2147     \@namedef{#1@sh@\string#2@\string#3@}{#4}%
2148   \fi}

```

`\textormath` 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 `\textormath` is provided.

```

2149 \def\textormath{%
2150   \ifmmode
2151     \expandafter\@secondoftwo
2152   \else
2153     \expandafter\@firstoftwo
2154   \fi}

```

`\user@group` 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’.

```

2155 \def\user@group{user}
2156 \def\language@group{english} % TODO. I don't like defaults
2157 \def\system@group{system}

\useshorthands This is the user level macro. It initializes and activates 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.

2158 \def\useshorthands{%
2159   \@ifstar\bbl@usesh@s{\bbl@usesh@x{}}
2160 \def\bbl@usesh@s#1{%
2161   \bbl@usesh@x
2162   {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bbl@activate{#1}}}%
2163   {#1}}
2164 \def\bbl@usesh@x#1#2{%
2165   \bbl@ifshorthand{#2}%
2166   {\def\user@group{user}%
2167    \initiate@active@char{#2}%
2168    #1%
2169    \bbl@activate{#2}}%
2170   {\bbl@error
2171    {Cannot declare a shorthand turned off (\string#2)}
2172    {Sorry, but you cannot use shorthands which have been\\%
2173     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 \bbl@set@user@generic); we make also sure {} and
\protect are taken into account in this new top level.

2174 \def\user@language@group{user@\language@group}
2175 \def\bbl@set@user@generic#1#2{%
2176   \bbl@ifunset{user@generic@active#1}%
2177   {\bbl@active@def#1\user@language@group{user@active}{user@generic@active}%
2178    \bbl@active@def#1\user@group{user@generic@active}{language@active}%
2179    \expandafter\edef\csname#2@sh@#1@@\endcsname{%
2180     \expandafter\noexpand\csname normal@char#1\endcsname}%
2181    \expandafter\edef\csname#2@sh@#1@\string\protect@\endcsname{%
2182     \expandafter\noexpand\csname user@active#1\endcsname}}%
2183   \@empty}
2184 \newcommand\defineshorthand[3][user]{%
2185   \edef\bbl@tempa{\zap@space#1 \@empty}%
2186   \bbl@for\bbl@tempb\bbl@tempa{%
2187     \if*\expandafter\@car\bbl@tempb\@nil
2188       \edef\bbl@tempb{user@\expandafter\@gobble\bbl@tempb}%
2189       \@expandtwoargs
2190       \bbl@set@user@generic{\expandafter\string\@car#2\@nil}\bbl@tempb
2191     \fi
2192     \declare@shorthand{\bbl@tempb}{#2}{#3}}

\languageshorthands 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 to fix it in the same way languages names are fixed. [TODO].

2193 \def\languageshorthands#1{\def\language@group{#1}}

\aliasshorthand First the new shorthand needs to be initialized. 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".

2194 \def\aliasshorthand#1#2{%

```

```

2195 \bbl@ifshorthand{#2}%
2196 {\expandafter\ifx\csname active@char\string#2\endcsname\relax
2197   \ifx\document\@notprerr
2198     \notshorthand{#2}%
2199   \else
2200     \initiate@active@char{#2}%
2201     \expandafter\let\csname active@char\string#2\expandafter\endcsname
2202       \csname active@char\string#1\endcsname
2203     \expandafter\let\csname normal@char\string#2\expandafter\endcsname
2204       \csname normal@char\string#1\endcsname
2205     \bbl@activate{#2}%
2206   \fi
2207 \fi}%
2208 {\bbl@error
2209   {Cannot declare a shorthand turned off (\string#2)}
2210   {Sorry, but you cannot use shorthands which have been\\%
2211     turned off in the package options}}}

```

\@notshorthand

```

2212 \def\@notshorthand#1{%
2213   \bbl@error{%
2214     The character '\string #1' should be made a shorthand character;\\%
2215     add the command \string\usesshorthands\string{#1\string} to
2216     the preamble.\\%
2217     I will ignore your instruction}%
2218   {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  
\shorthandoff \@nil at the end to denote the end of the list of characters.

```

2219 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
2220 \DeclareRobustCommand*\shorthandoff{%
2221   \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
2222 \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.

```

2223 \def\bbl@switch@sh#1#2{%
2224   \ifx#2\@nnil\else
2225     \bbl@ifunset{\bbl@active@\string#2}%
2226     {\bbl@error
2227       {I cannot switch '\string#2' on or off--not a shorthand}%
2228       {This character is not a shorthand. Maybe you made\\%
2229         a typing mistake? I will ignore your instruction}}}%
2230     {\ifcase#1%
2231       \catcode`#2\relax
2232     \or
2233       \catcode`#2\active
2234     \or
2235       \csname bbl@oricat@\string#2\endcsname
2236       \csname bbl@oridef@\string#2\endcsname
2237     \fi}%
2238     \bbl@afterfi\bbl@switch@sh#1%
2239   \fi}

```

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

```

2240 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
2241 \def\bbl@putsh#1{%
2242   \bbl@ifunset{\bbl@active@\string#1}%
2243   {\bbl@putsh@i#1\@empty\@nnil}%
2244   {\csname bbl@active@\string#1\endcsname}}
2245 \def\bbl@putsh@i#1#2\@nnil{%
2246   \csname\language@group @sh@\string#1@%
2247   \ifx\@empty#2\else\string#2\fi\endcsname}
2248 \ifx\bbl@opt@shorthands\@nnil\else
2249   \let\bbl@s@initiate@active@char\initiate@active@char
2250   \def\initiate@active@char#1{%
2251     \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
2252   \let\bbl@s@switch@sh\bbl@switch@sh
2253   \def\bbl@switch@sh#1#2{%
2254     \ifx#2\@nnil\else
2255       \bbl@afterfi
2256       \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
2257       \fi}
2258   \let\bbl@s@activate\bbl@activate
2259   \def\bbl@activate#1{%
2260     \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
2261   \let\bbl@s@deactivate\bbl@deactivate
2262   \def\bbl@deactivate#1{%
2263     \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
2264 \fi

```

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

```

2265 \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  
**\bbl@pr@m@s** 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.

```

2266 \def\bbl@prim@s{%
2267   \prime\futurelet\@let@token\bbl@pr@m@s}
2268 \def\bbl@if@primes#1#2{%
2269   \ifx#1\@let@token
2270     \expandafter\@firstoftwo
2271   \else\ifx#2\@let@token
2272     \bbl@afterelse\expandafter\@firstoftwo
2273   \else
2274     \bbl@afterfi\expandafter\@secondoftwo
2275   \fi\fi}
2276 \begingroup
2277   \catcode`\^=7 \catcode`\*=\active \lccode`\*=\^
2278   \catcode`\'=12 \catcode`\"=\active \lccode`\"=\'
2279   \lowercase{%
2280     \gdef\bbl@pr@m@s{%
2281       \bbl@if@primes"'"%
2282       \pr@@@s
2283       {\bbl@if@primes*\^pr@@@t\egroup}}}
2284 \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).

```
2285 \initiate@active@char{~}
2286 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
2287 \bbl@activate{~}
```

\OT1dpos The position of the double quote character is different for the OT1 and T1 encodings. It will later be selected using the \f@encoding macro. Therefore we define two macros here to store the position of the character in these encodings.

```
2288 \expandafter\def\csname OT1dpos\endcsname{127}
2289 \expandafter\def\csname T1dpos\endcsname{4}
```

When the macro \f@encoding is undefined (as it is in plain T<sub>E</sub>X) we define it here to expand to OT1

```
2290 \ifx\f@encoding\undefined
2291 \def\f@encoding{OT1}
2292 \fi
```

## 9.6 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.

```
2293 \bbl@trace{Language attributes}
2294 \newcommand\languageattribute[2]{%
2295 \def\bbl@tempc{#1}%
2296 \bbl@fixname\bbl@tempc
2297 \bbl@iflanguage\bbl@tempc{%
2298 \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@attrs. When that control sequence is not yet defined this attribute is certainly not selected before.

```
2299 \ifx\bbl@known@attrs\undefined
2300 \in@false
2301 \else
2302 \bbl@xin@{,\bbl@tempc-##1,}{,\bbl@known@attrs,}%
2303 \fi
2304 \ifin@
2305 \bbl@warning{%
2306 You have more than once selected the attribute '##1'\%
2307 for language #1. Reported}%
2308 \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 T<sub>E</sub>X-code.

```
2309 \bbl@exp{%
2310 \bbl@add@list\bbl@known@attrs{\bbl@tempc-##1}}%
2311 \edef\bbl@tempa{\bbl@tempc-##1}%
2312 \expandafter\bbl@ifknown@ttrib\expandafter{\bbl@tempa}\bbl@attributes%
2313 {\csname\bbl@tempc @attr@##1\endcsname}%
2314 {\@attrerr{\bbl@tempc}{##1}}%
2315 \fi}}
2316 \@onlypreamble\languageattribute
```

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

```
2317 \newcommand*{\@attrerr}[2]{%
2318   \bbl@error
2319   {The attribute #2 is unknown for language #1.}%
2320   {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}`.

```
2321 \def\bbl@declare@ttribute#1#2#3{%
2322   \bbl@xin@{,#2,}{,\BabelModifiers,}%
2323   \ifin@
2324     \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
2325   \fi
2326   \bbl@add@list\bbl@attributes{#1-#2}%
2327   \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.

```
2328 \def\bbl@ifattributeset#1#2#3#4{%
2329   \ifx\bbl@known@attribs\@undefined
2330     \in@false
2331   \else
2332     \bbl@xin@{,#1-#2,}{,\bbl@known@attribs,}%
2333   \fi
2334   \ifin@
2335     \bbl@afterelse#3%
2336   \else
2337     \bbl@afterfi#4%
2338   \fi}
```

**\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. We first assume the attribute is unknown. Then we loop over the list of known attributes, trying to find a match.

```
2339 \def\bbl@ifknown@ttrib#1#2{%
2340   \let\bbl@tempa\@secondoftwo
2341   \bbl@loopx\bbl@tempb{#2}{%
2342     \expandafter\in@\expandafter{\expandafter,\bbl@tempb,}{,#1,}%
2343   \ifin@
2344     \let\bbl@tempa\@firstoftwo
2345   \else
2346     \fi}%
2347   \bbl@tempa}
```

**\bbl@clear@ttribs** This macro removes all the attribute code from  $\TeX$ 's memory at `\begin{document}` time (if any is present).

```
2348 \def\bbl@clear@ttribs{%
2349   \ifx\bbl@attributes\@undefined\else
2350     \bbl@loopx\bbl@tempa{\bbl@attributes}{%
2351       \expandafter\bbl@clear@ttrib\bbl@tempa.
2352     }%
2353   \let\bbl@attributes\@undefined}
```

```

2354 \fi}
2355 \def\bbl@clear@ttrib#1-#2.{%
2356 \expandafter\let\csname#1@attr@#2\endcsname\undefined}
2357 \AtBeginDocument{\bbl@clear@ttribs}

```

## 9.7 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
2358 \bbl@trace{Macros for saving definitions}
2359 \def\babel@beginsave{\babel@savecnt\z@}

```

Before it's forgotten, allocate the counter and initialize all.

```

2360 \newcount\babel@savecnt
2361 \babel@beginsave

```

`\babel@save` The macro `\babel@save<cname>` saves the current meaning of the control sequence `<cname>` to `\originalTeX`<sup>32</sup>. To do this, we let the current meaning to a temporary control sequence, the restore commands are appended to `\originalTeX` and the counter is incremented. The macro `\babel@savevariable<variable>` saves the value of the variable. `<variable>` can be anything allowed after the `\the` primitive.

```

2362 \def\babel@save#1{%
2363 \expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax
2364 \toks@\expandafter{\originalTeX\let#1=}%
2365 \bbl@exp{%
2366 \def\\originalTeX{\the\toks@<\babel@number\babel@savecnt>\relax}}%
2367 \advance\babel@savecnt@ne}
2368 \def\babel@savevariable#1{%
2369 \toks@\expandafter{\originalTeX #1=}%
2370 \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 command `\bbl@nonfrenchspacing` switches it on when it isn't already in effect and `\bbl@nonfrenchspacing` switches it off if necessary. A more refined way to switch the catcodes is done with ini files. Here an auxiliary macro is defined, but the main part is in `\babelprovide`. This new method should be ideally the default one.

```

2371 \def\bbl@frenchspacing{%
2372 \ifnum\the\sfcode`.=\@m
2373 \let\bbl@nonfrenchspacing\relax
2374 \else
2375 \frenchspacing
2376 \let\bbl@nonfrenchspacing\nonfrenchspacing
2377 \fi}
2378 \let\bbl@nonfrenchspacing\nonfrenchspacing
2379 \let\bbl@elt\relax
2380 \edef\bbl@fs@chars{%
2381 \bbl@elt{\string.}\@m{3000}\bbl@elt{\string?}\@m{3000}%
2382 \bbl@elt{\string!}\@m{3000}\bbl@elt{\string:}\@m{2000}%
2383 \bbl@elt{\string;}\@m{1500}\bbl@elt{\string,}\@m{1250}}

```

<sup>32</sup>`\originalTeX` has to be expandable, i.e. you shouldn't let it to `\relax`.



## 9.8 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.

```

2384 \bbl@trace{Short tags}
2385 \def\babeltags#1{%
2386   \edef\bbl@tempa{\zap@space#1 \@empty}%
2387   \def\bbl@tempb##1=##2\@{
2388     \edef\bbl@tempc{%
2389       \noexpand\newcommand
2390       \expandafter\noexpand\csname ##1\endcsname{%
2391         \noexpand\protect
2392         \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
2393       \noexpand\newcommand
2394       \expandafter\noexpand\csname text##1\endcsname{%
2395         \noexpand\foreignlanguage{##2}}
2396       \bbl@tempc}%
2397   \bbl@for\bbl@tempa\bbl@tempa{%
2398     \expandafter\bbl@tempb\bbl@tempa\@{

```

## 9.9 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.

```

2399 \bbl@trace{Hyphens}
2400 \@onlypreamble\babelhyphenation
2401 \AtEndOfPackage{%
2402   \newcommand\babelhyphenation[2][\@empty]{%
2403     \ifx\bbl@hyphenation@relax
2404       \let\bbl@hyphenation@\@empty
2405     \fi
2406     \ifx\bbl@hyphlist\@empty\else
2407       \bbl@warning{%
2408         You must not intermingle \string\selectlanguage\space and\%
2409         \string\babelhyphenation\space or some exceptions will not\%
2410         be taken into account. Reported}%
2411     \fi
2412     \ifx\@empty#1%
2413       \protected@edef\bbl@hyphenation@{\bbl@hyphenation@\space#2}%
2414     \else
2415       \bbl@vforeach{#1}{%
2416         \def\bbl@tempa{##1}%
2417         \bbl@fixname\bbl@tempa
2418         \bbl@iflanguage\bbl@tempa{%
2419           \bbl@csarg\protected@edef{hyphenation@\bbl@tempa}{%
2420             \bbl@ifunset{bbl@hyphenation@\bbl@tempa}%
2421             {}%
2422             {\csname bbl@hyphenation@\bbl@tempa\endcsname\space}%
2423             #2}}}%
2424       \fi}

```

`\bbl@allowhyphens` This macro makes hyphenation possible. Basically its definition is nothing more than `\nobreak \hskip 0pt plus 0pt`<sup>33</sup>.

```

2425 \def\bbl@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}

```

<sup>33</sup> $\TeX$  begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.

```

2426 \def\bbl@t@one{T1}
2427 \def\allowhyphens{\ifx\cf@encoding\bbl@t@one\else\bbl@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.

```

2428 \newcommand\babellnullhyphen{\char\hyphenchar\font}
2429 \def\babelhyphen{\active@prefix\babelhyphen\bbl@hyphen}
2430 \def\bbl@hyphen{%
2431   \@ifstar{\bbl@hyphen@i @}{\bbl@hyphen@i \@empty}}
2432 \def\bbl@hyphen@i#1#2{%
2433   \bbl@ifunset{\bbl@hy@#1#2@empty}%
2434   {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%
2435   {\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.

```

2436 \def\bbl@usehyphen#1{%
2437   \leavevmode
2438   \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
2439   \nobreak\hskip\z@skip}
2440 \def\bbl@@usehyphen#1{%
2441   \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

The following macro inserts the hyphen char.

```

2442 \def\bbl@hyphenchar{%
2443   \ifnum\hyphenchar\font=\m@ne
2444     \babellnullhyphen
2445   \else
2446     \char\hyphenchar\font
2447   \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@nobreak is redundant.

```

2448 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
2449 \def\bbl@hy@@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
2450 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
2451 \def\bbl@hy@@hard{\bbl@usehyphen\bbl@hyphenchar}
2452 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}
2453 \def\bbl@hy@@nobreak{\mbox{\bbl@hyphenchar}}
2454 \def\bbl@hy@repeat{%
2455   \bbl@usehyphen{%
2456     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}
2457 \def\bbl@hy@@repeat{%
2458   \bbl@usehyphen{%
2459     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}
2460 \def\bbl@hy@empty{\hskip\z@skip}
2461 \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.

```

2462 \def\bbl@disc#1#2{\nobreak\discretionary{#2-}{}{#1}\bbl@allowhyphens}

```

## 9.10 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.

```
2463 \bbl@trace{Multiencoding strings}
2464 \def\bbl@tglobal#1{\global\let#1#1}
2465 \def\bbl@recatcode#1{% TODO. Used only once?
2466   \@tempcnta="7F
2467   \def\bbl@tempa{%
2468     \ifnum\@tempcnta>"FF\else
2469       \catcode\@tempcnta=#1\relax
2470       \advance\@tempcnta\@ne
2471       \expandafter\bbl@tempa
2472     \fi}%
2473   \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).

```
2474 \@@ifpackagewith{babel}{nocase}%
2475   {\let\bbl@patchuclc\relax}%
2476   {\def\bbl@patchuclc{%
2477     \global\let\bbl@patchuclc\relax
2478     \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}%
2479     \gdef\bbl@uclc##1{%
2480       \let\bbl@encoded\bbl@encoded@uclc
2481       \bbl@ifunset{\language @bbl@uclc}% and resumes it
2482       {##1}%
2483       {\let\bbl@tempa##1\relax % Used by LANG@bbl@uclc
2484         \csname\language @bbl@uclc\endcsname}%
2485       {\bbl@tolower\@empty}{\bbl@toupper\@empty}}%
2486     \gdef\bbl@tolower{\csname\language @bbl@lc\endcsname}%
2487     \gdef\bbl@toupper{\csname\language @bbl@uc\endcsname}}%
2488 \<<*More package options>> \equiv
2489 \DeclareOption{nocase}{}
2490 \<</More package options>>
```

The following package options control the behavior of `\SetString`.

```
2491 \<<*More package options>> \equiv
2492 \let\bbl@opt@strings\@nnil % accept strings=value
2493 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
2494 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
2495 \def\BabelStringsDefault{generic}
2496 \<</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.

```

2497 \@onlypreamble\StartBabelCommands
2498 \def\StartBabelCommands{%
2499   \begingroup
2500   \bbl@recatcode{11}%
2501   <<Macros local to BabelCommands>>
2502   \def\bbl@provstring##1##2{%
2503     \providecommand##1{##2}%
2504     \bbl@tglobal##1}%
2505   \global\let\bbl@scafter\@empty
2506   \let\StartBabelCommands\bbl@startcmds
2507   \ifx\BabelLanguages\relax
2508     \let\BabelLanguages\CurrentOption
2509   \fi
2510   \begingroup
2511   \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
2512   \StartBabelCommands}
2513 \def\bbl@startcmds{%
2514   \ifx\bbl@screset\@nnil\else
2515     \bbl@usehooks{stopcommands}{}%
2516   \fi
2517   \endgroup
2518   \begingroup
2519   \@ifstar
2520     {\ifx\bbl@opt@strings\@nnil
2521       \let\bbl@opt@strings\BabelStringsDefault
2522     \fi
2523     \bbl@startcmds@i}%
2524   \bbl@startcmds@i}
2525 \def\bbl@startcmds@i#1#2{%
2526   \edef\bbl@L{\zap@space#1 \@empty}%
2527   \edef\bbl@G{\zap@space#2 \@empty}%
2528   \bbl@startcmds@ii}
2529 \let\bbl@startcmds\StartBabelCommands

```

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.

```

2530 \newcommand\bbl@startcmds@ii[1][\@empty]{%
2531   \let\SetString\@gobbletwo
2532   \let\bbl@stringdef\@gobbletwo
2533   \let\AfterBabelCommands\@gobble
2534   \ifx\@empty#1%
2535     \def\bbl@sc@label{generic}%
2536     \def\bbl@encstring##1##2{%
2537       \ProvideTextCommandDefault##1{##2}%
2538       \bbl@tglobal##1%
2539       \expandafter\bbl@tglobal\csname\string?\string##1\endcsname}%
2540     \let\bbl@sctest\in@true
2541   \else
2542     \let\bbl@sc@charset\space % <- zapped below

```

```

2543 \let\bbl@sc@fontenc\space % <- " "
2544 \def\bbl@tempa##1=##2\@nil{%
2545 \bbl@csarg\edef{sc@zap@space##1 \@empty}{##2 }}%
2546 \bbl@vforeach{label=#1}{\bbl@tempa##1\@nil}%
2547 \def\bbl@tempa##1 ##2{% space -> comma
2548 ##1%
2549 \ifx\@empty##2\else\ifx,##1,\else,\fi\bbl@afterfi\bbl@tempa##2\fi}%
2550 \edef\bbl@sc@fontenc{\expandafter\bbl@tempa\bbl@sc@fontenc\@empty}%
2551 \edef\bbl@sc@label{\expandafter\zap@space\bbl@sc@label\@empty}%
2552 \edef\bbl@sc@charset{\expandafter\zap@space\bbl@sc@charset\@empty}%
2553 \def\bbl@encstring##1##2{%
2554 \bbl@foreach\bbl@sc@fontenc{%
2555 \bbl@ifunset{T####1}%
2556 {}%
2557 {\ProvideTextCommand##1{####1}{##2}%
2558 \bbl@tglobal##1%
2559 \expandafter
2560 \bbl@tglobal\csname####1\string##1\endcsname}}}%
2561 \def\bbl@sctest{%
2562 \bbl@xin@{\, \bbl@opt@strings,}{, \bbl@sc@label, \bbl@sc@fontenc,}}%
2563 \fi
2564 \ifx\bbl@opt@strings\@nnil % ie, no strings key -> defaults
2565 \else\ifx\bbl@opt@strings\relax % ie, strings=encoded
2566 \let\AfterBabelCommands\bbl@aftercmds
2567 \let\SetString\bbl@setstring
2568 \let\bbl@stringdef\bbl@encstring
2569 \else % ie, strings=value
2570 \bbl@sctest
2571 \ifin@
2572 \let\AfterBabelCommands\bbl@aftercmds
2573 \let\SetString\bbl@setstring
2574 \let\bbl@stringdef\bbl@provstring
2575 \fi\fi\fi
2576 \bbl@scswitch
2577 \ifx\bbl@G\@empty
2578 \def\SetString##1##2{%
2579 \bbl@error{Missing group for string \string##1}%
2580 {You must assign strings to some category, typically\\%
2581 captions or extras, but you set none}}%
2582 \fi
2583 \ifx\@empty#1%
2584 \bbl@usehooks{defaultcommands}}}%
2585 \else
2586 \@expandtwoargs
2587 \bbl@usehooks{encodedcommands}{\bbl@sc@charset}{\bbl@sc@fontenc}}%
2588 \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).

```

2589 \def\bbl@forlang#1##2{%
2590 \bbl@for#1\bbl@L{%
2591 \bbl@xin@{, #1,}{, \BabelLanguages,}%
2592 \ifin@#2\relax\fi}}

```

```

2593 \def\bbl@scswitch{%
2594   \bbl@forlang\bbl@tempa{%
2595     \ifx\bbl@G\@empty\else
2596       \ifx\SetString\@gobbletwo\else
2597         \edef\bbl@GL{\bbl@G\bbl@tempa}%
2598         \bbl@xin{,\bbl@GL,}{,\bbl@screset,}%
2599         \ifin@\else
2600           \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
2601           \xdef\bbl@screset{\bbl@screset,\bbl@GL}%
2602         \fi
2603       \fi
2604     \fi}}
2605 \AtEndOfPackage{%
2606   \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{#2}}}%
2607   \let\bbl@scswitch\relax}
2608 \@onlypreamble\EndBabelCommands
2609 \def\EndBabelCommands{%
2610   \bbl@usehooks{stopcommands}{}%
2611   \endgroup
2612   \endgroup
2613   \bbl@scafter}
2614 \let\bbl@endcommands\EndBabelCommands

```

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.

```

2615 \def\bbl@setstring#1#2{% eg, \prefacename{<string>}
2616   \bbl@forlang\bbl@tempa{%
2617     \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
2618     \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
2619     {\bbl@exp{%
2620       \global\bbbl@add\<\bbl@G\bbl@tempa>{\bbbl@scset\#1\<\bbl@LC>}}}%
2621     }%
2622     \def\BabelString{#2}%
2623     \bbl@usehooks{stringprocess}{}%
2624     \expandafter\bbl@stringdef
2625     \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.

```

2626 \ifx\bbl@opt@strings\relax
2627   \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
2628   \bbl@patchuclc
2629   \let\bbl@encoded\relax
2630   \def\bbl@encoded@uclc#1{%
2631     \@inmathwarn#1%
2632     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
2633       \expandafter\ifx\csname ?\string#1\endcsname\relax
2634         \TextSymbolUnavailable#1%
2635       \else
2636         \csname ?\string#1\endcsname
2637       \fi
2638     \else
2639       \csname\cf@encoding\string#1\endcsname

```

```

2640     \fi}
2641 \else
2642   \def\bbl@scset#1#2{\def#1{#2}}
2643 \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.

```

2644 <<*Macros local to BabelCommands>> ≡
2645 \def\SetStringLoop##1##2{%
2646   \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
2647   \count@\z@
2648   \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
2649     \advance\count@\@ne
2650     \toks@\expandafter{\bbl@tempa}%
2651     \bbl@exp{%
2652       \\\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
2653       \count@=\the\count@\relax}}}%
2654 <</Macros local to BabelCommands>>

```

**Delaying code** Now the definition of `\AfterBabelCommands` when it is activated.

```

2655 \def\bbl@aftercmds#1{%
2656   \toks@\expandafter{\bbl@scafter#1}%
2657   \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.

```

2658 <<*Macros local to BabelCommands>> ≡
2659 \newcommand\SetCase[3][]{%
2660   \bbl@patchuclc
2661   \bbl@forlang\bbl@tempa{%
2662     \expandafter\bbl@encstring
2663     \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
2664     \expandafter\bbl@encstring
2665     \csname\bbl@tempa @bbl@uc\endcsname{##2}%
2666     \expandafter\bbl@encstring
2667     \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
2668 <</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.

```

2669 <<*Macros local to BabelCommands>> ≡
2670 \newcommand\SetHyphenMap[1]{%
2671   \bbl@forlang\bbl@tempa{%
2672     \expandafter\bbl@stringdef
2673     \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
2674 <</Macros local to BabelCommands>>

```

There are 3 helper macros which do most of the work for you.

```

2675 \newcommand\BabelLower[2]{% one to one.
2676   \ifnum\lccode#1=#2\else
2677     \babel@savevariable{\lccode#1}%
2678     \lccode#1=#2\relax
2679   \fi}
2680 \newcommand\BabelLowerMM[4]{% many-to-many
2681   \@tempcnta=#1\relax

```

```

2682 \@tempcntb=#4\relax
2683 \def\bbl@tempa{%
2684   \ifnum\@tempcnta>#2\else
2685     \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
2686     \advance\@tempcnta#3\relax
2687     \advance\@tempcntb#3\relax
2688     \expandafter\bbl@tempa
2689   \fi}%
2690 \bbl@tempa}
2691 \newcommand\BabelLowerM0[4]{% many-to-one
2692 \@tempcnta=#1\relax
2693 \def\bbl@tempa{%
2694   \ifnum\@tempcnta>#2\else
2695     \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
2696     \advance\@tempcnta#3
2697     \expandafter\bbl@tempa
2698   \fi}%
2699 \bbl@tempa}

```

The following package options control the behavior of hyphenation mapping.

```

2700 <<{*More package options}>> ≡
2701 \DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
2702 \DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\@ne}
2703 \DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
2704 \DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@}
2705 \DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}
2706 <</More package options>>

```

Initial setup to provide a default behavior if hyphenmap is not set.

```

2707 \AtEndOfPackage{%
2708   \ifx\bbl@opt@hyphenmap\undefined
2709     \bbl@xin@{,}{\bbl@language@opts}%
2710     \chardef\bbl@opt@hyphenmap\ifin@4\else\@ne\fi
2711   \fi}

```

This section ends with a general tool for resetting the caption names with a unique interface. With the old way, which mixes the switcher and the string, we convert it to the new one, which separates these two steps.

```

2712 \newcommand\setlocalecaption{% TODO. Catch typos. What about ensure?
2713   \@ifstar\bbl@setcaption@s\bbl@setcaption@x}
2714 \def\bbl@setcaption@x#1#2#3{% language caption-name string
2715   \bbl@trim@def\bbl@tempa{#2}%
2716   \bbl@xin@{.template}{\bbl@tempa}%
2717   \ifin@
2718     \bbl@ini@captions@template{#3}{#1}%
2719   \else
2720     \edef\bbl@tempd{%
2721       \expandafter\expandafter\expandafter
2722       \strip@prefix\expandafter\meaning\csname captions#1\endcsname}%
2723     \bbl@xin@
2724       {\expandafter\string\csname #2name\endcsname}%
2725     {\bbl@tempd}%
2726     \ifin@ % Renew caption
2727       \bbl@xin@{\string\bbl@scset}{\bbl@tempd}%
2728     \ifin@
2729       \bbl@exp{%
2730         \\bbl@ifsamestring{\bbl@tempa}{\language}%
2731         {\bbl@scset\<#2name>\<#1#2name>}%
2732         {}}%

```



```

2733 \else % Old way converts to new way
2734 \bbl@ifunset{#1#2name}%
2735 {\bbl@exp{%
2736 \\\bbl@add\<captions#1>{\def\<#2name>{\<#1#2name>}}}%
2737 \\\bbl@ifsamestring{\bbl@tempa}{\language}%
2738 {\def\<#2name>{\<#1#2name>}}}%
2739 {}}}%
2740 {}%
2741 \fi
2742 \else
2743 \bbl@xin@{\string\bbl@scset}{\bbl@tempd}% New
2744 \ifin@ % New way
2745 \bbl@exp{%
2746 \\\bbl@add\<captions#1>{\\\bbl@scset\<#2name>\<#1#2name>}}%
2747 \\\bbl@ifsamestring{\bbl@tempa}{\language}%
2748 {\\\bbl@scset\<#2name>\<#1#2name>}}%
2749 {}}}%
2750 \else % Old way, but defined in the new way
2751 \bbl@exp{%
2752 \\\bbl@add\<captions#1>{\def\<#2name>{\<#1#2name>}}}%
2753 \\\bbl@ifsamestring{\bbl@tempa}{\language}%
2754 {\def\<#2name>{\<#1#2name>}}}%
2755 {}}}%
2756 \fi%
2757 \fi
2758 \@namedef{#1#2name}{#3}%
2759 \toks@ \expandafter{\bbl@captionslist}%
2760 \bbl@exp{\in@{\<#2name>}{\the\toks@}}%
2761 \ifin@ \else
2762 \bbl@exp{\\\bbl@add\\bbl@captionslist{\<#2name>}}%
2763 \bbl@to\global\bbl@captionslist
2764 \fi
2765 \fi}
2766 % \def\bbl@setcaption@#1#2#3{} % TODO. Not yet implemented

```

## 9.11 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.

```

2767 \bbl@trace{Macros related to glyphs}
2768 \def\set@low@box#1{\setbox\tw@ \hbox{,}\setbox\z@ \hbox{#1}%
2769 \dimen\z@ \ht\z@ \advance\dimen\z@ -\ht\tw@%
2770 \setbox\z@ \hbox{\lower\dimen\z@ \box\z@}\ht\z@ \ht\tw@ \dp\z@ \dp\tw@}

```

`\save@sfc@q` The macro `\save@sfc@q` is used to save and reset the current space factor.

```

2771 \def\save@sfc@q#1{\leavevmode
2772 \begingroup
2773 \edef\@SF{\spacefactor\the\spacefactor}#1\@SF
2774 \endgroup}

```

## 9.12 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`.

### 9.12.1 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.

```
2775 \ProvideTextCommand{\quotedblbase}{OT1}{%
2776   \save@sf@q{\set@low@box{\textquotedblright\}%
2777     \box\z@\kern-.04em\bb1@allowhyphens}}
```

Make sure that when an encoding other than OT1 or T1 is used this glyph can still be typeset.

```
2778 \ProvideTextCommandDefault{\quotedblbase}{%
2779   \UseTextSymbol{OT1}{\quotedblbase}}
```

`\quotesinglbase` We also need the single quote character at the baseline.

```
2780 \ProvideTextCommand{\quotesinglbase}{OT1}{%
2781   \save@sf@q{\set@low@box{\textquoteright\}%
2782     \box\z@\kern-.04em\bb1@allowhyphens}}
```

Make sure that when an encoding other than OT1 or T1 is used this glyph can still be typeset.

```
2783 \ProvideTextCommandDefault{\quotesinglbase}{%
2784   \UseTextSymbol{OT1}{\quotesinglbase}}
```

`\guillemetleft` `\guillemetright` The guillemet characters are not available in OT1 encoding. They are faked. (Wrong names with o preserved for compatibility.)

```
2785 \ProvideTextCommand{\guillemetleft}{OT1}{%
2786   \ifmmode
2787     \ll
2788   \else
2789     \save@sf@q{\nobreak
2790       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bb1@allowhyphens}%
2791     \fi}
2792 \ProvideTextCommand{\guillemetright}{OT1}{%
2793   \ifmmode
2794     \gg
2795   \else
2796     \save@sf@q{\nobreak
2797       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bb1@allowhyphens}%
2798     \fi}
2799 \ProvideTextCommand{\guillemotleft}{OT1}{%
2800   \ifmmode
2801     \ll
2802   \else
2803     \save@sf@q{\nobreak
2804       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bb1@allowhyphens}%
2805     \fi}
2806 \ProvideTextCommand{\guillemotright}{OT1}{%
2807   \ifmmode
2808     \gg
2809   \else
2810     \save@sf@q{\nobreak
2811       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bb1@allowhyphens}%
2812     \fi}
```

Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.

```
2813 \ProvideTextCommandDefault{\guillemetleft}{%
2814   \UseTextSymbol{OT1}{\guillemetleft}}
2815 \ProvideTextCommandDefault{\guillemetright}{%
2816   \UseTextSymbol{OT1}{\guillemetright}}
2817 \ProvideTextCommandDefault{\guillemotleft}{%
2818   \UseTextSymbol{OT1}{\guillemotleft}}
2819 \ProvideTextCommandDefault{\guillemotright}{%
2820   \UseTextSymbol{OT1}{\guillemotright}}
```

`\guilsinglleft` The single guillemets are not available in OT1 encoding. They are faked.

`\guilsinglright`

```

2821 \ProvideTextCommand{\guilsinglleft}{OT1}{%
2822   \ifmmode
2823     <%
2824   \else
2825     \save@sf@q{\nobreak
2826       \raise.2ex\hbox{$\scriptscriptstyle<$}\bbl@allowhyphens}%
2827   \fi}
2828 \ProvideTextCommand{\guilsinglright}{OT1}{%
2829   \ifmmode
2830     >%
2831   \else
2832     \save@sf@q{\nobreak
2833       \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
2834   \fi}

```

Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.

```

2835 \ProvideTextCommandDefault{\guilsinglleft}{%
2836   \UseTextSymbol{OT1}{\guilsinglleft}}
2837 \ProvideTextCommandDefault{\guilsinglright}{%
2838   \UseTextSymbol{OT1}{\guilsinglright}}

```

### 9.12.2 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.

```

2839 \DeclareTextCommand{\ij}{OT1}{%
2840   i\kern-0.02em\bbl@allowhyphens j}
2841 \DeclareTextCommand{\IJ}{OT1}{%
2842   I\kern-0.02em\bbl@allowhyphens J}
2843 \DeclareTextCommand{\ij}{T1}{\char188}
2844 \DeclareTextCommand{\IJ}{T1}{\char156}

```

Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.

```

2845 \ProvideTextCommandDefault{\ij}{%
2846   \UseTextSymbol{OT1}{\ij}}
2847 \ProvideTextCommandDefault{\IJ}{%
2848   \UseTextSymbol{OT1}{\IJ}}

```

`\dj` The croatian language needs the letters `\dj` and `\DJ`; they are available in the T1 encoding, 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).

```

2849 \def\crrtic@{\hrule height0.1ex width0.3em}
2850 \def\crttic@{\hrule height0.1ex width0.33em}
2851 \def\ddj@{%
2852   \setbox0\hbox{d}\dimen@=\ht0
2853   \advance\dimen@1ex
2854   \dimen@.45\dimen@
2855   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
2856   \advance\dimen@ii.5ex
2857   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\box{\crrtic@}}}}
2858 \def\DDJ@{%
2859   \setbox0\hbox{D}\dimen@=.55\ht0
2860   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
2861   \advance\dimen@ii.15ex % correction for the dash position
2862   \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
2863   \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@

```

```

2864 \leavevmode\rlap{\raise\dimen@hbox{\kern\dimen@ii\vbox{\crttic@}}}}
2865 %
2866 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
2867 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}

```

Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.

```

2868 \ProvideTextCommandDefault{\dj}{%
2869 \UseTextSymbol{OT1}{\dj}}
2870 \ProvideTextCommandDefault{\DJ}{%
2871 \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.

```

2872 \DeclareTextCommand{\SS}{OT1}{SS}
2873 \ProvideTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}

```

### 9.12.3 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 2874 \ProvideTextCommandDefault{\glq}{%
2875 \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}

```

The definition of \grq depends on the fontencoding. With T1 encoding no extra kerning is needed.

```

2876 \ProvideTextCommand{\grq}{T1}{%
2877 \textormath{\kern\z@\textquoteleft}{\mbox{\textquoteleft}}}
2878 \ProvideTextCommand{\grq}{TU}{%
2879 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
2880 \ProvideTextCommand{\grq}{OT1}{%
2881 \save@sf@q{\kern-.0125em
2882 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
2883 \kern.07em\relax}}
2884 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}{\grq}}

```

**\glqq** The ‘german’ double quotes.

```

\grqq 2885 \ProvideTextCommandDefault{\glqq}{%
2886 \textormath{\quotedblbase}{\mbox{\quotedblbase}}}

```

The definition of \grqq depends on the fontencoding. With T1 encoding no extra kerning is needed.

```

2887 \ProvideTextCommand{\grqq}{T1}{%
2888 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
2889 \ProvideTextCommand{\grqq}{TU}{%
2890 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
2891 \ProvideTextCommand{\grqq}{OT1}{%
2892 \save@sf@q{\kern-.07em
2893 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
2894 \kern.07em\relax}}
2895 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}{\grqq}}

```

**\flq** The ‘french’ single guillemets.

```

\frq 2896 \ProvideTextCommandDefault{\flq}{%
2897 \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
2898 \ProvideTextCommandDefault{\frq}{%
2899 \textormath{\guilsinglright}{\mbox{\guilsinglright}}}

```

```

\flqq The ‘french’ double guillemets.
\frqq
2900 \ProvideTextCommandDefault{\flqq}{%
2901   \textormath{\guillemetleft}{\mbox{\guillemetleft}}}
2902 \ProvideTextCommandDefault{\frqq}{%
2903   \textormath{\guillemetright}{\mbox{\guillemetright}}}

```

### 9.12.4 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 positioning, the default will be `\umlauthigh` (the normal positioning).

```

2904 \def\uumlauthigh{%
2905   \def\bbl@umlauta##1{\leavevmode\bgroup%
2906     \expandafter\accent\csname\fontencoding dqpos\endcsname
2907     ##1\bbl@allowhyphens\egroup}%
2908   \let\bbl@umlaute\bbl@umlauta}
2909 \def\uumlautlow{%
2910   \def\bbl@umlauta{\protect\lower@umlaut}}
2911 \def\uumlautelow{%
2912   \def\bbl@umlaute{\protect\lower@umlaut}}
2913 \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.

```

2914 \expandafter\ifx\csname U@D\endcsname\relax
2915   \csname newdimen\endcsname\U@D
2916 \fi

```

The following code fools  $\TeX$ ’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.

```

2917 \def\lower@umlaut#1{%
2918   \leavevmode\bgroup
2919   \U@D 1ex%
2920   {\setbox\z@\hbox{%
2921     \expandafter\char\csname\fontencoding dqpos\endcsname}%
2922     \dimen@ -.45ex\advance\dimen@\ht\z@
2923     \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
2924     \expandafter\accent\csname\fontencoding dqpos\endcsname
2925     \fontdimen5\font\U@D #1%
2926   \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).

```

2927 \AtBeginDocument{%

```

```

2928 \DeclareTextCompositeCommand{"}{OT1}{a}{\bbl@umlauta{a}}%
2929 \DeclareTextCompositeCommand{"}{OT1}{e}{\bbl@umlaute{e}}%
2930 \DeclareTextCompositeCommand{"}{OT1}{i}{\bbl@umlaute{i}}%
2931 \DeclareTextCompositeCommand{"}{OT1}{\i}{\bbl@umlaute{i}}}%
2932 \DeclareTextCompositeCommand{"}{OT1}{o}{\bbl@umlauta{o}}%
2933 \DeclareTextCompositeCommand{"}{OT1}{u}{\bbl@umlauta{u}}%
2934 \DeclareTextCompositeCommand{"}{OT1}{A}{\bbl@umlauta{A}}%
2935 \DeclareTextCompositeCommand{"}{OT1}{E}{\bbl@umlaute{E}}%
2936 \DeclareTextCompositeCommand{"}{OT1}{I}{\bbl@umlaute{I}}%
2937 \DeclareTextCompositeCommand{"}{OT1}{O}{\bbl@umlauta{O}}%
2938 \DeclareTextCompositeCommand{"}{OT1}{U}{\bbl@umlauta{U}}%

```

Finally, make sure the default hyphenrules are defined (even if empty). For internal use, another empty \language is defined. Currently used in Amharic.

```

2939 \ifx\l@english\@undefined
2940 \chardef\l@english\z@
2941 \fi
2942 % The following is used to cancel rules in ini files (see Amharic).
2943 \ifx\l@babelnohyphens\@undefined
2944 \newlanguage\l@babelnohyphens
2945 \fi

```

## 9.13 Layout

Layout is mainly intended to set bidi documents, but there is at least a tool useful in general.

```

2946 \bbl@trace{Bidi layout}
2947 \providecommand\IfBabelLayout[3]{#3}%
2948 \newcommand\BabelPatchSection[1]{%
2949 \@ifundefined{#1}{%
2950 \bbl@exp{\let\bbl@ss@#1>\<#1>}%
2951 \@namedef{#1}{%
2952 \ifstar\bbl@presec@#1}%
2953 {\@dblarg\bbl@presec@x{#1}}}%
2954 \def\bbl@presec@x#1[#2]#3{%
2955 \bbl@exp{%
2956 \\\select@language@x{\bbl@main@language}%
2957 \\\bbl@cs{sspre@#1}%
2958 \\\bbl@cs{ss@#1}%
2959 [\\foreignlanguage{\language}{\unexpanded{#2}}}%
2960 {\foreignlanguage{\language}{\unexpanded{#3}}}%
2961 \\\select@language@x{\language}}%
2962 \def\bbl@presec@#1#2{%
2963 \bbl@exp{%
2964 \\\select@language@x{\bbl@main@language}%
2965 \\\bbl@cs{sspre@#1}%
2966 \\\bbl@cs{ss@#1}*%
2967 {\foreignlanguage{\language}{\unexpanded{#2}}}%
2968 \\\select@language@x{\language}}%
2969 \IfBabelLayout{sectioning}%
2970 {\BabelPatchSection{part}%
2971 \BabelPatchSection{chapter}%
2972 \BabelPatchSection{section}%
2973 \BabelPatchSection{subsection}%
2974 \BabelPatchSection{subsubsection}%
2975 \BabelPatchSection{paragraph}%
2976 \BabelPatchSection{subparagraph}%
2977 \def\babel@toc#1{%
2978 \select@language@x{\bbl@main@language}}}%

```

```

2979 \IfBabelLayout{captions}%
2980 {\BabelPatchSection{caption}}{}

```

## 9.14 Load engine specific macros

```

2981 \bbl@trace{Input engine specific macros}
2982 \ifcase\bbl@engine
2983   \input txtbabel.def
2984 \or
2985   \input luababel.def
2986 \or
2987   \input xebabel.def
2988 \fi

```

## 9.15 Creating and modifying 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.

```

2989 \bbl@trace{Creating languages and reading ini files}
2990 \newcommand\babelprovide[2][]{%
2991   \let\bbl@savelangname\language
2992   \edef\bbl@savelocaleid{\the\localeid}%
2993   % Set name and locale id
2994   \edef\language{#2}%
2995   % \global\@namedef\bbl@lcname@#2}{#2}%
2996   \bbl@id@assign
2997   \let\bbl@KVP@captions\@nil
2998   \let\bbl@KVP@date\@nil
2999   \let\bbl@KVP@import\@nil
3000   \let\bbl@KVP@main\@nil
3001   \let\bbl@KVP@script\@nil
3002   \let\bbl@KVP@language\@nil
3003   \let\bbl@KVP@hyphenrules\@nil
3004   \let\bbl@KVP@mapfont\@nil
3005   \let\bbl@KVP@maparabic\@nil
3006   \let\bbl@KVP@mapdigits\@nil
3007   \let\bbl@KVP@intraspace\@nil
3008   \let\bbl@KVP@intrapenalty\@nil
3009   \let\bbl@KVP@onchar\@nil
3010   \let\bbl@KVP@transforms\@nil
3011   \let\bbl@KVP@alph\@nil
3012   \let\bbl@KVP@Alph\@nil
3013   \let\bbl@KVP@labels\@nil
3014   \bbl@csarg\let{KVP@labels*}\@nil
3015   \global\let\bbl@inidata\@empty
3016   \bbl@forkv{#1}{% TODO - error handling
3017     \in@{/{}}{##1}%
3018     \ifin@
3019       \bbl@renewinikey##1\@{##2}%
3020     \else
3021       \bbl@csarg\def{KVP@##1}{##2}%
3022     \fi}%
3023   % == init ==
3024   \ifx\bbl@screset\@undefined
3025     \bbl@ldfinit
3026   \fi
3027   % ==
3028   \let\bbl@lbfkflag\relax % \@empty = do setup linebreak

```

```

3029 \bbl@ifunset{date#2}%
3030 {\let\bbl@lbkflag\@empty}% new
3031 {\ifx\bbl@KVP@hyphenrules\@nil\else
3032 \let\bbl@lbkflag\@empty
3033 \fi
3034 \ifx\bbl@KVP@import\@nil\else
3035 \let\bbl@lbkflag\@empty
3036 \fi}%
3037 % == import, captions ==
3038 \ifx\bbl@KVP@import\@nil\else
3039 \bbl@exp{\bbl@ifblank{\bbl@KVP@import}}%
3040 {\ifx\bbl@initload\relax
3041 \begingroup
3042 \def\BabelBeforeIni##1##2{\gdef\bbl@KVP@import{##1}\endinput}%
3043 \bbl@input@texini{#2}%
3044 \endgroup
3045 \else
3046 \xdef\bbl@KVP@import{\bbl@initload}%
3047 \fi}%
3048 {}%
3049 \fi
3050 \ifx\bbl@KVP@captions\@nil
3051 \let\bbl@KVP@captions\bbl@KVP@import
3052 \fi
3053 % ==
3054 \ifx\bbl@KVP@transforms\@nil\else
3055 \bbl@replace\bbl@KVP@transforms{ }{,}%
3056 \fi
3057 % Load ini
3058 \bbl@ifunset{date#2}%
3059 {\bbl@provide@new{#2}}%
3060 {\bbl@ifblank{#1}%
3061 {}% With \bbl@load@basic below
3062 {\bbl@provide@renew{#2}}}%
3063 % Post tasks
3064 % -----
3065 % == ensure captions ==
3066 \ifx\bbl@KVP@captions\@nil\else
3067 \bbl@ifunset{\bbl@extracaps@#2}%
3068 {\bbl@exp{\bbl@babelensure[exclude=\today]{#2}}%
3069 {\toks@\expandafter\expandafter\expandafter
3070 {\csname bbl@extracaps@#2\endcsname}%
3071 \bbl@exp{\bbl@babelensure[exclude=\today,include=\the\toks@]{#2}}%
3072 \bbl@ifunset{\bbl@ensure@\language}%
3073 {\bbl@exp{%
3074 \\\DeclareRobustCommand\<bbl@ensure@\language>[1]{%
3075 \\\foreignlanguage{\language}%
3076 {####1}}}%
3077 {}%
3078 \bbl@exp{%
3079 \\\bbl@tglobal\<bbl@ensure@\language>%
3080 \\\bbl@tglobal\<bbl@ensure@\language\space>%
3081 \fi
3082 % ==
3083 % At this point all parameters are defined if 'import'. Now we
3084 % execute some code depending on them. But what about if nothing was
3085 % imported? We just set the basic parameters, but still loading the
3086 % whole ini file.
3087 \bbl@load@basic{#2}%

```



```

3088 % == script, language ==
3089 % Override the values from ini or defines them
3090 \ifx\bb1@KVP@script\@nil\else
3091   \bb1@csarg\edef\sname@#2{\bb1@KVP@script}%
3092 \fi
3093 \ifx\bb1@KVP@language\@nil\else
3094   \bb1@csarg\edef\lname@#2{\bb1@KVP@language}%
3095 \fi
3096 % == onchar ==
3097 \ifx\bb1@KVP@onchar\@nil\else
3098   \bb1@luahyphenate
3099   \directlua{
3100     if Babel.locale_mapped == nil then
3101       Babel.locale_mapped = true
3102       Babel.linebreaking.add_before(Babel.locale_map)
3103       Babel.loc_to_scr = {}
3104       Babel.chr_to_loc = Babel.chr_to_loc or {}
3105     end}%
3106   \bb1@xin@{ ids }{ \bb1@KVP@onchar\space}%
3107   \ifin@
3108     \ifx\bb1@starthyphens\@undefined % Needed if no explicit selection
3109       \AddBabelHook{babel-onchar}{beforestart}{\bb1@starthyphens}%
3110     \fi
3111     \bb1@exp{\bb1@add\bb1@starthyphens
3112       {\bb1@patterns@lua{\language}}}%
3113     % TODO - error/warning if no script
3114     \directlua{
3115       if Babel.script_blocks['\bb1@cl{sbc}'] then
3116         Babel.loc_to_scr[\the\localeid] =
3117           Babel.script_blocks['\bb1@cl{sbc}']
3118         Babel.locale_props[\the\localeid].lc = \the\localeid\space
3119         Babel.locale_props[\the\localeid].lg = \the\@nameuse{1\language}\space
3120       end
3121     }%
3122   \fi
3123   \bb1@xin@{ fonts }{ \bb1@KVP@onchar\space}%
3124   \ifin@
3125     \bb1@ifunset{bb1@lsys\language}{\bb1@provide@lsys\language}{}%
3126     \bb1@ifunset{bb1@wdir\language}{\bb1@provide@dirs\language}{}%
3127     \directlua{
3128       if Babel.script_blocks['\bb1@cl{sbc}'] then
3129         Babel.loc_to_scr[\the\localeid] =
3130           Babel.script_blocks['\bb1@cl{sbc}']
3131       end}%
3132   \ifx\bb1@mapselect\@undefined
3133     \AtBeginDocument{%
3134       \expandafter\bb1@add\csname selectfont \endcsname{\bb1@mapselect}%
3135       {\selectfont}}%
3136     \def\bb1@mapselect{%
3137       \let\bb1@mapselect\relax
3138       \edef\bb1@prefontid{\fontid\font}}%
3139     \def\bb1@mapdir##1{%
3140       {\def\language{##1}%
3141         \let\bb1@ifrestoring\@firstoftwo % To avoid font warning
3142         \bb1@switchfont
3143         \directlua{
3144           Babel.locale_props[\the\csname bb1@id@##1\endcsname]
3145             ['\bb1@prefontid'] = \fontid\font\space}}}%
3146   \fi

```

```

3147 \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\language}}}%
3148 \fi
3149 % TODO - catch non-valid values
3150 \fi
3151 % == mapfont ==
3152 % For bidi texts, to switch the font based on direction
3153 \ifx\bbl@KVP@mapfont\@nil\else
3154 \bbl@ifsamestring{\bbl@KVP@mapfont}{direction}}%
3155 {\bbl@error{Option '\bbl@KVP@mapfont' unknown for\%
3156 mapfont. Use 'direction'.%
3157 {See the manual for details.}}}%
3158 \bbl@ifunset{\bbl@lsys@\language}{\bbl@provide@lsys@\language}}}%
3159 \bbl@ifunset{\bbl@wdir@\language}{\bbl@provide@dirs@\language}}}%
3160 \ifx\bbl@mapselect\@undefined
3161 \AtBeginDocument{%
3162 \expandafter\bbl@add\csname selectfont \endcsname{\bbl@mapselect}}%
3163 {\selectfont}}%
3164 \def\bbl@mapselect{%
3165 \let\bbl@mapselect\relax
3166 \edef\bbl@prefontid{\fontid\font}}%
3167 \def\bbl@mapdir##1{%
3168 {\def\language{##1}%
3169 \let\bbl@ifrestoring\@firstoftwo % avoid font warning
3170 \bbl@switchfont
3171 \directlua{Babel.fontmap
3172 [\the\csname bbl@wdir@##1\endcsname]%
3173 [\bbl@prefontid]=\fontid\font}}}%
3174 \fi
3175 \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\language}}}%
3176 \fi
3177 % == Line breaking: intraspace, intrapenalty ==
3178 % For CJK, East Asian, Southeast Asian, if interspace in ini
3179 \ifx\bbl@KVP@intraspace\@nil\else % We can override the ini or set
3180 \bbl@csarg\edef{intsp@#2}{\bbl@KVP@intraspace}%
3181 \fi
3182 \bbl@provide@intraspace
3183 % == Line breaking: hyphenate.other.locale/.script==
3184 \ifx\bbl@lbkflag\@empty
3185 \bbl@ifunset{\bbl@hyotl@\language}}%
3186 {\bbl@csarg\bbl@replace{hyotl@\language}{ }{,}%
3187 \bbl@startcommands*\language}%
3188 \bbl@csarg\bbl@foreach{hyotl@\language}{%
3189 \ifcase\bbl@engine
3190 \ifnum##1<257
3191 \SetHyphenMap{\BabelLower{##1}{##1}}%
3192 \fi
3193 \else
3194 \SetHyphenMap{\BabelLower{##1}{##1}}%
3195 \fi}%
3196 \bbl@endcommands}%
3197 \bbl@ifunset{\bbl@hyots@\language}}%
3198 {\bbl@csarg\bbl@replace{hyots@\language}{ }{,}%
3199 \bbl@csarg\bbl@foreach{hyots@\language}{%
3200 \ifcase\bbl@engine
3201 \ifnum##1<257
3202 \global\lccode##1=##1\relax
3203 \fi
3204 \else
3205 \global\lccode##1=##1\relax

```

```

3206     \fi}}%
3207 \fi
3208 % == Counters: maparabic ==
3209 % Native digits, if provided in ini (TeX level, xe and lua)
3210 \ifcase\bbbl@engine\else
3211     \bbbl@ifunset{\bbbl@dgnat@\languagename}{}%
3212     {\expandafter\ifx\csname bbl@dgnat@\languagename\endcsname\@empty\else
3213         \expandafter\expandafter\expandafter
3214         \bbbl@setdigits\csname bbl@dgnat@\languagename\endcsname
3215         \ifx\bbbl@KVP@maparabic\@nil\else
3216             \ifx\bbbl@latinarabic\@undefined
3217                 \expandafter\let\expandafter\@arabic
3218                 \csname bbl@counter@\languagename\endcsname
3219             \else % ie, if layout=counters, which redefines \@arabic
3220                 \expandafter\let\expandafter\bbbl@latinarabic
3221                 \csname bbl@counter@\languagename\endcsname
3222             \fi
3223         \fi
3224     \fi}%
3225 \fi
3226 % == Counters: mapdigits ==
3227 % Native digits (lua level).
3228 \ifodd\bbbl@engine
3229     \ifx\bbbl@KVP@mapdigits\@nil\else
3230         \bbbl@ifunset{\bbbl@dgnat@\languagename}{}%
3231         {\RequirePackage{luatexbase}%
3232         \bbbl@activate@preotf
3233         \directlua{
3234             Babel = Babel or {} %%% -> presets in luababel
3235             Babel.digits_mapped = true
3236             Babel.digits = Babel.digits or {}
3237             Babel.digits[\the\localeid] =
3238                 table.pack(string.utfvalue('\bbbl@cl{dgnat}'))
3239             if not Babel.numbers then
3240                 function Babel.numbers(head)
3241                     local LOCALE = luatexbase.registernumber'\bbbl@attr@locale'
3242                     local GLYPH = node.id'glyph'
3243                     local inmath = false
3244                     for item in node.traverse(head) do
3245                         if not inmath and item.id == GLYPH then
3246                             local temp = node.get_attribute(item, LOCALE)
3247                             if Babel.digits[temp] then
3248                                 local chr = item.char
3249                                 if chr > 47 and chr < 58 then
3250                                     item.char = Babel.digits[temp][chr-47]
3251                                 end
3252                             end
3253                         elseif item.id == node.id'math' then
3254                             inmath = (item.subtype == 0)
3255                         end
3256                     end
3257                     return head
3258                 end
3259             end
3260         }}%
3261     \fi
3262 \fi
3263 % == Counters: alph, Alph ==
3264 % What if extras<lang> contains a \babel@save\@alph? It won't be

```

```

3265 % restored correctly when exiting the language, so we ignore
3266 % this change with the \bbl@alph@savd trick.
3267 \ifx\bbl@KVP@alph@nil\else
3268   \toks@{\expandafter\expandafter\expandafter{%
3269     \csname extras\language\endcsname}%
3270   \bbl@exp{%
3271     \def<extras\language>{%
3272       \let\\bbl@alph@savd\\@alph
3273       \the\toks@
3274       \let\\@alph\\bbl@alph@savd
3275       \\babel@savd\\@alph
3276       \let\\@alph<bbl@cntr@bbl@KVP@alph @\language>}}%
3277   \fi
3278 \ifx\bbl@KVP@Alph@nil\else
3279   \toks@{\expandafter\expandafter\expandafter{%
3280     \csname extras\language\endcsname}%
3281   \bbl@exp{%
3282     \def<extras\language>{%
3283       \let\\bbl@Alph@savd\\@Alph
3284       \the\toks@
3285       \let\\@Alph\\bbl@Alph@savd
3286       \\babel@savd\\@Alph
3287       \let\\@Alph<bbl@cntr@bbl@KVP@Alph @\language>}}%
3288   \fi
3289 % == require.babel in ini ==
3290 % To load or reload the babel-*.tex, if require.babel in ini
3291 \ifx\bbl@beforestart\relax\else % But not in doc aux or body
3292   \bbl@ifunset{bbl@rtex@\language}{}%
3293     {\expandafter\ifx\csname bbl@rtex@\language\endcsname\@empty\else
3294       \let\BabelBeforeIni@gobbletwo
3295       \chardef\atcatcode=\catcode`\@
3296       \catcode`\@=11\relax
3297       \bbl@input@texini{\bbl@cs{rtex@\language}}%
3298       \catcode`\@=\atcatcode
3299       \let\atcatcode\relax
3300     \fi}%
3301 \fi
3302 % == main ==
3303 \ifx\bbl@KVP@main@nil % Restore only if not 'main'
3304   \let\language\bbl@savelangname
3305   \chardef\localeid\bbl@savelocaleid\relax
3306 \fi}

```

Depending on whether or not the language exists, we define two macros.

```

3307 \def\bbl@provide@new#1{%
3308   \@namedef{date#1}{}% marks lang exists - required by \StartBabelCommands
3309   \@namedef{extras#1}{}%
3310   \@namedef{noextras#1}{}%
3311   \bbl@startcommands*{#1}{captions}%
3312   \ifx\bbl@KVP@captions@nil % and also if import, implicit
3313     \def\bbl@tempb##1{% elt for \bbl@captionslist
3314       \ifx##1\@empty\else
3315         \bbl@exp{%
3316           \\SetString\\##1{%
3317             \\bbl@nocaption{\bbl@stripslash##1}{#1\bbl@stripslash##1}}%
3318           \expandafter\bbl@tempb
3319         \fi}%
3320     \expandafter\bbl@tempb\bbl@captionslist\@empty
3321   \else

```

```

3322 \ifx\bbbl@initoload\relax
3323 \bbbl@read@ini{\bbbl@KVP@captions}2% % Here letters cat = 11
3324 \else
3325 \bbbl@read@ini{\bbbl@initoload}2% % Same
3326 \fi
3327 \fi
3328 \StartBabelCommands*{#1}{date}%
3329 \ifx\bbbl@KVP@import\@nil
3330 \bbbl@exp{%
3331 \\\SetString\\today{\bbbl@nocaption{today}{#1today}}}%
3332 \else
3333 \bbbl@savetoday
3334 \bbbl@savedate
3335 \fi
3336 \bbbl@endcommands
3337 \bbbl@load@basic{#1}%
3338 % == hyphenmins == (only if new)
3339 \bbbl@exp{%
3340 \gdef\<#1hyphenmins>{%
3341 {\bbbl@ifunset{\bbbl@lfthm@#1}{2}{\bbbl@cs{lfthm@#1}}}%
3342 {\bbbl@ifunset{\bbbl@rgthm@#1}{3}{\bbbl@cs{rgthm@#1}}}%
3343 % == hyphenrules ==
3344 \bbbl@provide@hyphens{#1}%
3345 % == frenchspacing == (only if new)
3346 \bbbl@ifunset{\bbbl@frspc@#1}{}%
3347 {\edef\bbbl@tempa{\bbbl@cl{frspc}}%
3348 \edef\bbbl@tempa{\expandafter\@car\bbbl@tempa\@nil}%
3349 \if u\bbbl@tempa % do nothing
3350 \else\if n\bbbl@tempa % non french
3351 \expandafter\bbbl@add\csname extras#1\endcsname{%
3352 \let\bbbl@elt\bbbl@fs@elt@i
3353 \bbbl@fs@chars}%
3354 \else\if y\bbbl@tempa % french
3355 \expandafter\bbbl@add\csname extras#1\endcsname{%
3356 \let\bbbl@elt\bbbl@fs@elt@ii
3357 \bbbl@fs@chars}%
3358 \fi\fi\fi}%
3359 %
3360 \ifx\bbbl@KVP@main\@nil\else
3361 \expandafter\main@language\expandafter{#1}%
3362 \fi}
3363 % A couple of macros used above, to avoid hashes #####...
3364 \def\bbbl@fs@elt@i#1#2#3{%
3365 \ifnum\sfcode`#1=#2\relax
3366 \babel@savevariable{\sfcode`#1}%
3367 \sfcode`#1=#3\relax
3368 \fi}%
3369 \def\bbbl@fs@elt@ii#1#2#3{%
3370 \ifnum\sfcode`#1=#3\relax
3371 \babel@savevariable{\sfcode`#1}%
3372 \sfcode`#1=#2\relax
3373 \fi}%
3374 %
3375 \def\bbbl@provide@renew#1{%
3376 \ifx\bbbl@KVP@captions\@nil\else
3377 \StartBabelCommands*{#1}{captions}%
3378 \bbbl@read@ini{\bbbl@KVP@captions}2% % Here all letters cat = 11
3379 \EndBabelCommands
3380 \fi

```

```

3381 \ifx\bb1@KVP@import\@nil\else
3382 \StartBabelCommands*{#1}{date}%
3383 \bb1@savetoday
3384 \bb1@savedate
3385 \EndBabelCommands
3386 \fi
3387 % == hyphenrules ==
3388 \ifx\bb1@lbfkflag\@empty
3389 \bb1@provide@hyphens{#1}%
3390 \fi}

```

Load the basic parameters (ids, typography, counters, and a few more), while captions and dates are left out. But it may happen some data has been loaded before automatically, so we first discard the saved values.

```

3391 \def\bb1@load@basic#1{%
3392 \bb1@ifunset{bb1@inidata@\languagename}{}%
3393 {\getlocaleproperty\bb1@tempa{\languagename}{identification/load.level}%
3394 \ifcase\bb1@tempa
3395 \bb1@csarg\let{lname@\languagename}\relax
3396 \fi}%
3397 \bb1@ifunset{bb1@lname@#1}%
3398 {\def\BabelBeforeIni##1##2{%
3399 \begingroup
3400 \let\bb1@ini@captions@aux\@gobbletwo
3401 \def\bb1@inidate ####1.####2.####3.####4\relax ####5####6}%
3402 \bb1@read@ini{##1}1%
3403 \ifx\bb1@initoload\relax\endinput\fi
3404 \endgroup}%
3405 \begingroup % boxed, to avoid extra spaces:
3406 \ifx\bb1@initoload\relax
3407 \bb1@input@texini{#1}%
3408 \else
3409 \setbox\z@\hbox{\BabelBeforeIni{\bb1@initoload}}}%
3410 \fi
3411 \endgroup}%
3412 {}}

```

The hyphenrules option is handled with an auxiliary macro.

```

3413 \def\bb1@provide@hyphens#1{%
3414 \let\bb1@tempa\relax
3415 \ifx\bb1@KVP@hyphenrules\@nil\else
3416 \bb1@replace\bb1@KVP@hyphenrules{ }{,}%
3417 \bb1@foreach\bb1@KVP@hyphenrules{%
3418 \ifx\bb1@tempa\relax % if not yet found
3419 \bb1@ifsamestring{##1}{+}%
3420 {\bb1@exp{\addlanguage\<l@##1>}}}%
3421 {}%
3422 \bb1@ifunset{l@##1}%
3423 {}%
3424 {\bb1@exp{\let\bb1@tempa\<l@##1>}}}%
3425 \fi}%
3426 \fi
3427 \ifx\bb1@tempa\relax % if no opt or no language in opt found
3428 \ifx\bb1@KVP@import\@nil
3429 \ifx\bb1@initoload\relax\else
3430 \bb1@exp{ % and hyphenrules is not empty
3431 \bb1@ifblank{\bb1@cs{hyphr@#1}}%
3432 {}%
3433 {\let\bb1@tempa\<l@bb1@cl{hyphr}>}}}%

```

```

3434 \fi
3435 \else % if importing
3436 \bbl@exp{% and hyphenrules is not empty
3437 \\\bbl@ifblank{\bbl@cs{hyphr@#1}}%
3438 }}%
3439 {\let\\bbl@tempa\<l@bbl@c1{hyphr}>}}%
3440 \fi
3441 \fi
3442 \bbl@ifunset{bbl@tempa}% ie, relax or undefined
3443 {\bbl@ifunset{l@#1}% no hyphenrules found - fallback
3444 {\bbl@exp{\\adddialect\<l@#1>\language}}%
3445 }}% so, l@<lang> is ok - nothing to do
3446 {\bbl@exp{\\adddialect\<l@#1>\bbl@tempa}}% found in opt list or ini

```

The reader of babel-...tex files. We reset temporarily some catcodes.

```

3447 \def\bbl@input@texini#1{%
3448 \bbl@bsphack
3449 \bbl@exp{%
3450 \catcode`\\%=14 \catcode`\\=0
3451 \catcode`\\{=1 \catcode`\\}=2
3452 \lowercase{\\InputIfFileExists{babel-#1.tex}{}}}%
3453 \catcode`\\%=the\catcode`\%relax
3454 \catcode`\\=the\catcode`\%relax
3455 \catcode`\\{=the\catcode`\{relax
3456 \catcode`\\}=the\catcode`\}relax}%
3457 \bbl@esphack}

```

The following macros read and store ini files (but don't process them). For each line, there are 3 possible actions: ignore if starts with ;, switch section if starts with [, and store otherwise. There are used in the first step of \bbl@read@ini.

```

3458 \def\bbl@inline#1\bbl@inline{%
3459 \@ifnextchar[\bbl@iniset{\@ifnextchar\bbl@iniskip\bbl@inistore}#1\@@}% ]
3460 \def\bbl@iniset[#1]#2\@@{\def\bbl@section{#1}}%
3461 \def\bbl@iniskip#1\@@{% if starts with ;
3462 \def\bbl@inistore#1=#2\@@{% full (default)
3463 \bbl@trim@def\bbl@tempa{#1}%
3464 \bbl@trim\toks@{#2}%
3465 \bbl@ifunset{bbl@KVP@\bbl@section/\bbl@tempa}%
3466 {\bbl@exp{%
3467 \\\g@addto@macro\\bbl@inidata{%
3468 \\\bbl@elt{\bbl@section}{\bbl@tempa}{\the\toks@}}}%
3469 }}%
3470 \def\bbl@inistore@min#1=#2\@@{% minimal (maybe set in \bbl@read@ini)
3471 \bbl@trim@def\bbl@tempa{#1}%
3472 \bbl@trim\toks@{#2}%
3473 \bbl@xin@{.identification.}{.\bbl@section.}%
3474 \ifin@
3475 \bbl@exp{\\g@addto@macro\\bbl@inidata{%
3476 \\\bbl@elt{identification}{\bbl@tempa}{\the\toks@}}}%
3477 \fi}%

```

Now, the 'main loop', which **must be executed inside a group**. At this point, \bbl@inidata may contain data declared in \babelprovide, with 'slashed' keys. There are 3 steps: first read the ini file and store it; then traverse the stored values, and process some groups if required (date, captions, labels, counters); finally, 'export' some values by defining global macros (identification, typography, characters, numbers). The second argument is 0 when called to read the minimal data for fonts; with \babelprovide it's either 1 or 2.

```

3478 \ifx\bbl@readstream\undefined
3479 \csname newread\endcsname\bbl@readstream

```

```

3480 \fi
3481 \def\bbl@read@ini#1#2{%
3482   \openin\bbl@readstream=babel-#1.ini
3483   \ifeof\bbl@readstream
3484     \bbl@error
3485     {There is no ini file for the requested language\\%
3486      (#1). Perhaps you misspelled it or your installation\\%
3487      is not complete.}%
3488     {Fix the name or reinstall babel.}%
3489   \else
3490     % Store ini data in \bbl@inidata
3491     \catcode\ [=12 \catcode\]=12 \catcode\==12 \catcode\&=12
3492     \catcode\;=12 \catcode\|=12 \catcode\%=14 \catcode\-=12
3493     \bbl@info{Importing
3494               \ifcase#2font and identification \or basic \fi
3495               data for \language\\%
3496               from babel-#1.ini. Reported}%
3497     \ifnum#2=\z@
3498       \global\let\bbl@inidata\@empty
3499       \let\bbl@inistore\bbl@inistore@min    % Remember it's local
3500     \fi
3501     \def\bbl@section{identification}%
3502     \bbl@exp{\ \bbl@inistore tag.ini=#1\ \@}%
3503     \bbl@inistore load.level=#2\ @@
3504     \loop
3505     \if T\ifeof\bbl@readstream F\fi T\relax % Trick, because inside \loop
3506       \endlinechar\m@ne
3507       \read\bbl@readstream to \bbl@line
3508       \endlinechar\^^M
3509       \ifx\bbl@line\@empty\else
3510         \expandafter\bbl@inline\bbl@line\bbl@inline
3511       \fi
3512     \repeat
3513     % Process stored data
3514     \bbl@csarg\xdef{lini@\language}{#1}%
3515     \let\bbl@savestrings\@empty
3516     \let\bbl@savetoday\@empty
3517     \let\bbl@savestate\@empty
3518     \def\bbl@elt##1##2##3{%
3519       \def\bbl@section{##1}%
3520       \in@{=date.}{##1}% Find a better place
3521       \ifin@
3522         \bbl@ini@calendar{##1}%
3523       \fi
3524       \global\bbl@csarg\let{bbl@KVP###1/##2}\relax
3525       \bbl@ifunset{bbl@inikv###1}{}%
3526       {\csname bbl@inikv###1\endcsname{##2}{##3}}}%
3527     \bbl@inidata
3528     % 'Export' data
3529     \bbl@ini@exports{#2}%
3530     \global\bbl@csarg\let{inidata@\language}\bbl@inidata
3531     \global\let\bbl@inidata\@empty
3532     \bbl@exp{\ \bbl@add@list\ \bbl@ini@loaded{\language}}%
3533     \bbl@to\global\bbl@ini@loaded
3534   \fi}

```

A somewhat hackish tool to handle calendar sections. To be improved.

```

3535 \def\bbl@ini@calendar#1{%
3536   \lowercase{\def\bbl@tempa{=#1=}}%

```



```

3537 \bbl@replace\bbl@tempa{=date.gregorian}{}%
3538 \bbl@replace\bbl@tempa{=date.}{}%
3539 \in@{.licr={#1=}}%
3540 \ifin@
3541 \ifcase\bbl@engine
3542 \bbl@replace\bbl@tempa{.licr={}}%
3543 \else
3544 \let\bbl@tempa\relax
3545 \fi
3546 \fi
3547 \ifx\bbl@tempa\relax\else
3548 \bbl@replace\bbl@tempa{=}{}%
3549 \bbl@exp{%
3550 \def<bbl@inikv@#1>###1####2{%
3551 \\\bbl@inidate###1...\relax{###2}{\bbl@tempa}}}%
3552 \fi}

```

A key with a slash in `\babelprovide` replaces the value in the ini file (which is ignored altogether). The mechanism is simple (but suboptimal): add the data to the ini one (at this point the ini file has not yet been read), and define a dummy macro. When the ini file is read, just skip the corresponding key and reset the macro (in `\bbl@inistore` above).

```

3553 \def\bbl@renewinikey#1/#2\@#3{%
3554 \edef\bbl@tempa{\zap@space #1 \@empty}% section
3555 \edef\bbl@tempb{\zap@space #2 \@empty}% key
3556 \bbl@trim\toks@{#3}% value
3557 \bbl@exp{%
3558 \global\let<bbl@KVP@>\bbl@tempa/\bbl@tempb>\\@empty % just a flag
3559 \\\g@addto@macro\\bbl@inidata{%
3560 \\\bbl@elt{\bbl@tempa}{\bbl@tempb}{\the\toks@}}}%

```

The previous assignments are local, so we need to export them. If the value is empty, we can provide a default value.

```

3561 \def\bbl@exportkey#1#2#3{%
3562 \bbl@ifunset{bbl@kv@#2}%
3563 {\bbl@csarg\gdef{#1@<language>}{#3}}%
3564 {\expandafter\ifx\csname bbl@kv@#2\endcsname\@empty
3565 \bbl@csarg\gdef{#1@<language>}{#3}}%
3566 \else
3567 \bbl@exp{\global\let<bbl@#1@<language>><bbl@kv@#2>}%
3568 \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@ini@exports` is called always (via `\bbl@inisec`), while `\bbl@after@ini` must be called explicitly after `\bbl@read@ini` if necessary.

```

3569 \def\bbl@iniwarning#1{%
3570 \bbl@ifunset{bbl@kv@identification.warning#1}{}%
3571 {\bbl@warning{%
3572 From babel-\bbl@cs{lini@<language>}.ini:\\%
3573 \bbl@cs{kv@identification.warning#1}\\%
3574 Reported }}}
3575 %
3576 \let\bbl@release@transforms\@empty
3577 %
3578 \def\bbl@ini@exports#1{%
3579 % Identification always exported
3580 \bbl@iniwarning}%
3581 \ifcase\bbl@engine
3582 \bbl@iniwarning{.pdflatex}%
3583 \or

```

```

3584 \bbl@iniwarning{.lualatex}%
3585 \or
3586 \bbl@iniwarning{.xelatex}%
3587 \fi%
3588 \bbl@exportkey{elname}{identification.name.english}{}%
3589 \bbl@exp{\bbl@exportkey{lname}{identification.name.opentype}%
3590 {\csname bbl@elname\language\endcsname}}%
3591 \bbl@exportkey{tbc}{identification.tag.bcp47}{}%
3592 \bbl@exportkey{lbc}{identification.language.tag.bcp47}{}%
3593 \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
3594 \bbl@exportkey{esname}{identification.script.name}{}%
3595 \bbl@exp{\bbl@exportkey{sname}{identification.script.name.opentype}%
3596 {\csname bbl@esname\language\endcsname}}%
3597 \bbl@exportkey{sbc}{identification.script.tag.bcp47}{}%
3598 \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}%
3599 % Also maps bcp47 -> language
3600 \ifbbl@bcptoname
3601 \bbl@csarg\xdef{bcp@map@bbl@cl{tbc}}{\language}%
3602 \fi
3603 % Finish here transforms, too
3604 \bbl@release@transforms\relax % \relax closes the last item.
3605 % Conditional
3606 \ifnum#1>\z@ % 0 = only info, 1, 2 = basic, (re)new
3607 \bbl@exportkey{lbrk}{typography.linebreaking}{h}%
3608 \bbl@exportkey{hyphr}{typography.hyphenrules}{}%
3609 \bbl@exportkey{lftm}{typography.lefthyphenmin}{2}%
3610 \bbl@exportkey{rgtm}{typography.righthyphenmin}{3}%
3611 \bbl@exportkey{prehc}{typography.prehyphenchar}{}%
3612 \bbl@exportkey{hyotl}{typography.hyphenate.other.locale}{}%
3613 \bbl@exportkey{hyots}{typography.hyphenate.other.script}{}%
3614 \bbl@exportkey{intsp}{typography.intraspaces}{}%
3615 \bbl@exportkey{chrng}{characters.ranges}{}%
3616 \bbl@exportkey{dgnat}{numbers.digits.native}{}%
3617 \bbl@exportkey{frspc}{typography.frenchspacing}{u}%
3618 \ifnum#1=\tw@ % only (re)new
3619 \bbl@exportkey{rqtex}{identification.require.babel}{}%
3620 \bbl@toglobal\bbl@savetoday
3621 \bbl@toglobal\bbl@savestate
3622 \bbl@savestrings
3623 \fi
3624 \fi}

```

A shared handler for key=val lines to be stored in \bbl@kv@<section>.<key>.

```

3625 \def\bbl@inikv#1#2{% key=value
3626 \toks@{#2}% This hides #'s from ini values
3627 \bbl@csarg\edef{kv@bbl@section.#1}{\the\toks@}}

```

By default, the following sections are just read. Actions are taken later.

```

3628 \let\bbl@inikv@identification\bbl@inikv
3629 \let\bbl@inikv@typography\bbl@inikv
3630 \let\bbl@inikv@characters\bbl@inikv
3631 \let\bbl@inikv@numbers\bbl@inikv

```

Additive numerals require an additional definition. When .1 is found, two macros are defined – the basic one, without .1 called by \localnumeral, and another one preserving the trailing .1 for the ‘units’.

```

3632 \def\bbl@inikv@counters#1#2{%
3633 \bbl@ifsamestring{#1}{digits}%
3634 {\bbl@error{The counter name 'digits' is reserved for mapping\%

```

```

3635             decimal digits}%
3636         {Use another name.}}%
3637     {}%
3638     \def\bbl@tempc{#1}%
3639     \bbl@trim@def{\bbl@tempb*}{#2}%
3640     \in@{.1$}{#1$}%
3641     \ifin@
3642         \bbl@replace\bbl@tempc{.1}{}%
3643         \bbl@csarg\protected@xdef{cntr@\bbl@tempc @\language}\{
3644             \noexpand\bbl@alphanumeric{\bbl@tempc}}%
3645     \fi
3646     \in@{.F.}{#1}%
3647     \ifin@else\in@{.S.}{#1}\fi
3648     \ifin@
3649         \bbl@csarg\protected@xdef{cntr@#1@\language}\{\bbl@tempb*}%
3650     \else
3651         \toks@{}% Required by \bbl@buildifcase, which returns \bbl@tempa
3652         \expandafter\bbl@buildifcase\bbl@tempb* \ \ % Space after \ \
3653         \bbl@csarg{\global\expandafter\let}{cntr@#1@\language}\bbl@tempa
3654     \fi}

```

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.

```

3655 \ifcase\bbl@engine
3656     \bbl@csarg\def{inikv@captions.licr}#1#2{%
3657         \bbl@ini@captions@aux{#1}{#2}}
3658 \else
3659     \def\bbl@inikv@captions#1#2{%
3660         \bbl@ini@captions@aux{#1}{#2}}
3661 \fi

```

The auxiliary macro for captions define \<caption>name.

```

3662 \def\bbl@ini@captions@template#1#2{% string language tempa=capt-name
3663     \bbl@replace\bbl@tempa{.template}{}%
3664     \def\bbl@toreplace{#1}{}%
3665     \bbl@replace\bbl@toreplace{[ ]}{\nobreakspace}}%
3666     \bbl@replace\bbl@toreplace{[ ]}{\csname}%
3667     \bbl@replace\bbl@toreplace{[ ]}{\csname the}%
3668     \bbl@replace\bbl@toreplace{[ ]}{\name\endcsname}}%
3669     \bbl@replace\bbl@toreplace{[ ]}{\endcsname}}%
3670     \bbl@xin@{,\bbl@tempa,}{,chapter,appendix,part,}%
3671     \ifin@
3672         \@nameuse{\bbl@patch\bbl@tempa}%
3673         \global\bbl@csarg\let{\bbl@tempa fmt@#2}\bbl@toreplace
3674     \fi
3675     \bbl@xin@{,\bbl@tempa,}{,figure,table,}%
3676     \ifin@
3677         \toks@\expandafter{\bbl@toreplace}%
3678         \bbl@exp{\gdef\<fnum@\bbl@tempa>{\the\toks@}}%
3679     \fi}
3680 \def\bbl@ini@captions@aux#1#2{%
3681     \bbl@trim@def\bbl@tempa{#1}%
3682     \bbl@xin@{.template}{\bbl@tempa}%
3683     \ifin@
3684         \bbl@ini@captions@template{#2}\language
3685     \else
3686         \bbl@ifblank{#2}%
3687         {\bbl@exp%

```

```

3688     \toks@{\bbbl@nocaption{\bbbl@tempa}{\language\bbbl@tempa name}}}%
3689     {\bbbl@trim\toks@{#2}}}%
3690     \bbbl@exp{%
3691         \bbbl@add\bbbl@savestrings{%
3692             \SetString\<\bbbl@tempa name>{\the\toks@}}}%
3693     \toks@\expandafter{\bbbl@captionslist}%
3694     \bbbl@exp{\in{\<\bbbl@tempa name>}{\the\toks@}}}%
3695     \ifin@
3696     \bbbl@exp{%
3697         \bbbl@add\<\bbbl@extracaps@\language\>{\<\bbbl@tempa name>}%
3698         \bbbl@tglobal\<\bbbl@extracaps@\language\>}%
3699     \fi
3700 \fi}

```

**Labels.** Captions must contain just strings, no format at all, so there is new group in ini files.

```

3701 \def\bbbl@list@the{%
3702     part,chapter,section,subsection,subsubsection,paragraph,%
3703     subparagraph,enumi,enumii,enumiii,enumiv,equation,figure,%
3704     table,page,footnote,mpfootnote,mpfn}
3705 \def\bbbl@map@cnt#1{% #1:roman,etc, // #2:enumi,etc
3706     \bbbl@ifunset{\bbbl@map@#1@\language}%
3707     {\@nameuse{#1}}}%
3708     {\@nameuse{\bbbl@map@#1@\language}}}%
3709 \def\bbbl@inikv@labels#1#2{%
3710     \in@{.map}{#1}%
3711     \ifin@
3712         \ifx\bbbl@KVP@labels\@nil\else
3713             \bbbl@xin@{ map }{\bbbl@KVP@labels\space}%
3714             \ifin@
3715                 \def\bbbl@tempc{#1}%
3716                 \bbbl@replace\bbbl@tempc{.map}{}%
3717                 \in@{,#2,}{,arabic,roman,Roman,alph,Alph,fnsymbol,}%
3718                 \bbbl@exp{%
3719                     \gdef\<\bbbl@map@\bbbl@tempc @\language\>%
3720                     {\ifin@<#2>\else\\localecounter{#2}\fi}}}%
3721                 \bbbl@foreach\bbbl@list@the{%
3722                     \bbbl@ifunset{the##1}{}%
3723                     {\bbbl@exp{\let\bbbl@tempd\<the##1>}%
3724                     \bbbl@exp{%
3725                         \bbbl@sreplace\<the##1>%
3726                         {\<\bbbl@tempc>{##1}}{\bbbl@map@cnt{\bbbl@tempc}{##1}}}%
3727                         \bbbl@sreplace\<the##1>%
3728                         {\<\@empty @\bbbl@tempc>\<c##1>}{\bbbl@map@cnt{\bbbl@tempc}{##1}}}%
3729                     \expandafter\ifx\csname the##1\endcsname\bbbl@tempd\else
3730                     \toks@\expandafter\expandafter\expandafter{%
3731                         \csname the##1\endcsname}%
3732                     \expandafter\xdef\csname the##1\endcsname{\the\toks@}}}%
3733                     \fi}}}%
3734     \fi
3735     \fi
3736 %
3737 \else
3738 %
3739 % The following code is still under study. You can test it and make
3740 % suggestions. Eg, enumerate.2 = ([enumi]).([enumii]). It's
3741 % language dependent.
3742 \in@{enumerate.}{#1}%
3743 \ifin@
3744     \def\bbbl@tempa{#1}%

```

```

3745 \bbl@replace\bbl@tempa{enumerate.}{}%
3746 \def\bbl@toreplace{#2}%
3747 \bbl@replace\bbl@toreplace{[ ]}{\nobreakspace{}}%
3748 \bbl@replace\bbl@toreplace{[ ]}{\csname the}%
3749 \bbl@replace\bbl@toreplace{ ]}{\endcsname{}}}%
3750 \toks@ \expandafter{\bbl@toreplace}%
3751 \bbl@exp{%
3752   \\ \bbl@add\<extras\language\>%
3753   \\ \babel@save\<labelenum\romannumeral\bbl@tempa>%
3754   \def\<labelenum\romannumeral\bbl@tempa>{\the\toks@}%
3755   \\ \bbl@toglobal\<extras\language\>%
3756   \fi
3757 \fi}

```

To show correctly some captions in a few languages, we need to patch some internal macros, because the order is hardcoded. For example, in Japanese the chapter number is surrounded by two string, while in Hungarian is placed after. These replacement works in many classes, but not all. Actually, the following lines are somewhat tentative.

```

3758 \def\bbl@chapttype{chapter}
3759 \ifx\@makechapterhead\undefined
3760 \let\bbl@patchchapter\relax
3761 \else\ifx\thechapter\undefined
3762 \let\bbl@patchchapter\relax
3763 \else\ifx\ps@headings\undefined
3764 \let\bbl@patchchapter\relax
3765 \else
3766 \def\bbl@patchchapter{%
3767   \global\let\bbl@patchchapter\relax
3768   \bbl@add\appendix{\def\bbl@chapttype{appendix}}% Not harmful, I hope
3769   \bbl@toglobal\appendix
3770   \bbl@sreplace\ps@headings
3771   {\@chapapp\ thechapter}%
3772   {\bbl@chapterformat}%
3773   \bbl@toglobal\ps@headings
3774   \bbl@sreplace\chaptermark
3775   {\@chapapp\ thechapter}%
3776   {\bbl@chapterformat}%
3777   \bbl@toglobal\chaptermark
3778   \bbl@sreplace\@makechapterhead
3779   {\@chapapp\space\thechapter}%
3780   {\bbl@chapterformat}%
3781   \bbl@toglobal\@makechapterhead
3782   \gdef\bbl@chapterformat{%
3783     \bbl@ifunset{\bbl@bbl@chapttype fmt@\language}%
3784     {\@chapapp\space\thechapter}
3785     {\@nameuse{\bbl@bbl@chapttype fmt@\language}}}}
3786 \let\bbl@patchappendix\bbl@patchchapter
3787 \fi\fi\fi
3788 \ifx\@part\undefined
3789 \let\bbl@patchpart\relax
3790 \else
3791 \def\bbl@patchpart{%
3792   \global\let\bbl@patchpart\relax
3793   \bbl@sreplace\@part
3794   {\partname\nobreakspace\thepart}%
3795   {\bbl@partformat}%
3796   \bbl@toglobal\@part
3797   \gdef\bbl@partformat{%
3798     \bbl@ifunset{\bbl@partfmt@\language}%

```

```

3799      {\partname\nobreakspace\thepart}
3800      {\@nameuse{bbl@partfmt@\languagename}}}}
3801 \fi

Date. TODO. Document

3802 % Arguments are _not_ protected.
3803 \let\bbl@calendar\@empty
3804 \DeclareRobustCommand\localedate[1][\bbl@localedate{#1}]
3805 \def\bbl@localedate#1#2#3#4{%
3806   \begingroup
3807     \ifx\@empty#1\@empty\else
3808       \let\bbl@ld@calendar\@empty
3809       \let\bbl@ld@variant\@empty
3810       \edef\bbl@tempa{\zap@space#1 \@empty}%
3811       \def\bbl@tempb##1=##2\@{\@namedef{bbl@ld@##1}{##2}}%
3812       \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
3813       \edef\bbl@calendar{%
3814         \bbl@ld@calendar
3815         \ifx\bbl@ld@variant\@empty\else
3816           .\bbl@ld@variant
3817         \fi}%
3818       \bbl@replace\bbl@calendar{gregorian}{}%
3819     \fi
3820     \bbl@cased
3821     {\@nameuse{bbl@date@\languagename @\bbl@calendar}{#2}{#3}{#4}}%
3822   \endgroup}
3823 % eg: 1=months, 2=wide, 3=1, 4=dummy, 5=value, 6=calendar
3824 \def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
3825   \bbl@trim@def\bbl@tempa{#1.#2}%
3826   \bbl@ifsamestring{\bbl@tempa}{months.wide}%      to savedate
3827   {\bbl@trim@def\bbl@tempa{#3}%
3828     \bbl@trim\toks@{#5}%
3829     \@temptokena\expandafter{\bbl@savestate}%
3830     \bbl@exp{% Reverse order - in ini last wins
3831       \def\\bbl@savestate{%
3832         \\SetString\<month\romannumeral\bbl@tempa#6name>{\the\toks@}%
3833         \the\@temptokena}}}%
3834   {\bbl@ifsamestring{\bbl@tempa}{date.long}%      defined now
3835     {\lowercase{\def\bbl@tempb{#6}}%
3836       \bbl@trim@def\bbl@toreplace{#5}%
3837       \bbl@TG@@date
3838       \bbl@ifunset{bbl@date@\languagename @}%
3839       {\global\bbl@csarg\let{date@\languagename @}\bbl@toreplace
3840         % TODO. Move to a better place.
3841         \bbl@exp{%
3842           \gdef\<\languagename date>{\protect\<\languagename date >}%
3843           \gdef\<\languagename date >####1####2####3{%
3844             \\bbl@usedategroupttrue
3845             \<bbl@ensure@\languagename>{%
3846               \\localedate{####1}{####2}{####3}}}%
3847             \\bbl@add\\bbl@savestate{%
3848               \\SetString\\today{%
3849                 \<\languagename date>%
3850                 {\the\year}{\the\month}{\the\day}}}}}%
3851           {}}%
3852       \ifx\bbl@tempb\@empty\else
3853         \global\bbl@csarg\let{date@\languagename @}\bbl@tempb\bbl@toreplace
3854       \fi}%
3855     {}}}}

```

**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.

```

3856 \let\bbl@calendar\@empty
3857 \newcommand\BabelDateSpace{\nobreakspace}
3858 \newcommand\BabelDateDot{.\@} % TODO. \let instead of repeating
3859 \newcommand\BabelDated[1]{\number#1}
3860 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
3861 \newcommand\BabelDateM[1]{\number#1}
3862 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}
3863 \newcommand\BabelDateMMMM[1]{%
3864   \csname month\romannumeral#1\bbl@calendar name\endcsname}%
3865 \newcommand\BabelDatey[1]{\number#1}%
3866 \newcommand\BabelDateyy[1]{%
3867   \ifnum#1<10 0\number#1 %
3868   \else\ifnum#1<100 \number#1 %
3869   \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
3870   \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
3871   \else
3872     \bbl@error
3873     {Currently two-digit years are restricted to the\
3874      range 0-9999.}%
3875     {There is little you can do. Sorry.}%
3876   \fi\fi\fi\fi}
3877 \newcommand\BabelDateyyyy[1]{\number#1} % FIXME - add leading 0
3878 \def\bbl@replace@finish@iii#1{%
3879   \bbl@exp{\def\#1####1####2####3{\the\toks@}}
3880 \def\bbl@TG@date{%
3881   \bbl@replace\bbl@toreplace{[ ]}{\BabelDateSpace{}}%
3882   \bbl@replace\bbl@toreplace{[.]}{\BabelDateDot{}}%
3883   \bbl@replace\bbl@toreplace{[d]}{\BabelDated{####3}}%
3884   \bbl@replace\bbl@toreplace{[dd]}{\BabelDatedd{####3}}%
3885   \bbl@replace\bbl@toreplace{[M]}{\BabelDateM{####2}}%
3886   \bbl@replace\bbl@toreplace{[MM]}{\BabelDateMM{####2}}%
3887   \bbl@replace\bbl@toreplace{[MMMM]}{\BabelDateMMMM{####2}}%
3888   \bbl@replace\bbl@toreplace{[y]}{\BabelDatey{####1}}%
3889   \bbl@replace\bbl@toreplace{[yy]}{\BabelDateyy{####1}}%
3890   \bbl@replace\bbl@toreplace{[yyyy]}{\BabelDateyyyy{####1}}%
3891   \bbl@replace\bbl@toreplace{[y]}{\bbl@datecctr[####1]}%
3892   \bbl@replace\bbl@toreplace{[m]}{\bbl@datecctr[####2]}%
3893   \bbl@replace\bbl@toreplace{[d]}{\bbl@datecctr[####3]}%
3894 % Note after \bbl@replace \toks@ contains the resulting string.
3895 % TODO - Using this implicit behavior doesn't seem a good idea.
3896   \bbl@replace@finish@iii\bbl@toreplace}
3897 \def\bbl@datecctr{\expandafter\bbl@xdatecctr\expandafter}
3898 \def\bbl@xdatecctr[#1|#2]{\localenumeral{#2}{#1}}

```

### Transforms.

```

3899 \let\bbl@release@transforms\@empty
3900 \@namedef{bbl@inikv@transforms.prehyphenation}{%
3901   \bbl@transforms\babelprehyphenation}
3902 \@namedef{bbl@inikv@transforms.posthyphenation}{%
3903   \bbl@transforms\babelposthyphenation}
3904 \def\bbl@transforms@aux#1#2#3,#4\relax{#1{#2}{#3}{#4}}
3905 \begingroup % TODO - to a lua file
3906   \catcode`\%=12
3907   \catcode`\&=14
3908   \gdef\bbl@transforms#1#2#3{&%
3909     \ifx\bbl@KVP@transforms\@nil\else

```

```

3910 \directlua{
3911     str = [==[#2]==]
3912     str = str:gsub('%.%d+%.%d+$', '')
3913     tex.print([[ \def\string\babeltempa{}} .. str .. [{}]])
3914 }&%
3915 \bbl@xin@{,\babeltempa,}{,\bbl@KVP@transforms,}&%
3916 \ifin@
3917     \in@{.0$}{#2$}&%
3918     \ifin@
3919         \bbl@add\bbl@release@transforms{&%
3920             \relax\bbl@transforms@aux#1{french}{#3}}&%
3921     \else
3922         \bbl@add\bbl@release@transforms{, {#3}}&%
3923     \fi
3924 \fi
3925 \fi}
3926 \endgroup

```

Language and Script values to be used when defining a font or setting the direction are set with the following macros.

```

3927 \def\bbl@provide@lsys#1{%
3928     \bbl@ifunset{bbl@lname@#1}%
3929     {\bbl@load@info{#1}}%
3930     }%
3931     \bbl@csarg\let{lsys@#1}\empty
3932     \bbl@ifunset{bbl@sname@#1}{\bbl@csarg\gdef{sname@#1}{Default}}{}%
3933     \bbl@ifunset{bbl@sotf@#1}{\bbl@csarg\gdef{sotf@#1}{DFLT}}{}%
3934     \bbl@csarg\bbl@add@list{lsys@#1}{Script=\bbl@cs{sname@#1}}%
3935     \bbl@ifunset{bbl@lname@#1}{}%
3936     {\bbl@csarg\bbl@add@list{lsys@#1}{Language=\bbl@cs{lname@#1}}}%
3937 \ifcase\bbl@engine\or\or
3938     \bbl@ifunset{bbl@prehc@#1}{}%
3939     {\bbl@exp{\bbl@ifblank{\bbl@cs{prehc@#1}}}%
3940     }%
3941     {\ifx\bbl@xenoxyph\undefined
3942         \let\bbl@xenoxyph\bbl@xenoxyph@d
3943         \ifx\AtBeginDocument\@notprerr
3944             \expandafter\@secondoftwo % to execute right now
3945             \fi
3946         \AtBeginDocument{%
3947             \expandafter\bbl@add
3948             \csname selectfont \endcsname{\bbl@xenoxyph}%
3949             \expandafter\selectlanguage\expandafter{\language}%
3950             \expandafter\bbl@toglobal\csname selectfont \endcsname}%
3951         \fi}}%
3952 \fi
3953 \bbl@csarg\bbl@toglobal{lsys@#1}}
3954 \def\bbl@xenoxyph@d{%
3955     \bbl@ifset{bbl@prehc@language}%
3956     {\ifnum\hyphenchar\font=\defaultshyphenchar
3957         \iffontchar\font\bbl@c1{prehc}\relax
3958         \hyphenchar\font\bbl@c1{prehc}\relax
3959     \else\iffontchar\font"200B
3960         \hyphenchar\font"200B
3961     \else
3962         \bbl@warning
3963         {Neither 0 nor ZERO WIDTH SPACE are available\\%
3964         in the current font, and therefore the hyphen\\%
3965         will be printed. Try changing the fontspec's\\%

```





```

4014 \def\\bbl@tempa####1{%
4015 \<ifcase>####1\space\the\toks@\\<else>\\@ctrerr\<fi>}}%
4016 \else
4017 \toks@\expandafter{\the\toks@\\or #1}%
4018 \expandafter\bbl@buildifcase
4019 \fi}

```

The code for additive counters is somewhat tricky and it's based on the fact the arguments just before \\@@ collects digits which have been left 'unused' in previous arguments, the first of them being the number of digits in the number to be converted. This explains the reverse set 76543210. Digits above 10000 are not handled yet. When the key contains the subkey .F., the number after is treated as an special case, for a fixed form (see babel-he.ini, for example).

```

4020 \newcommand\localenumeral[2]{\bbl@cs{cntr@#1@\\language}{#2}}
4021 \def\bbl@localecntr#1#2{\localenumeral{#2}{#1}}
4022 \newcommand\localecounter[2]{%
4023 \expandafter\bbl@localecntr
4024 \expandafter{\number\csname c@#2\endcsname}{#1}}
4025 \def\bbl@alphanumeric#1#2{%
4026 \expandafter\bbl@alphanumeric@i\number#2 76543210\\@@{#1}}
4027 \def\bbl@alphanumeric@i#1#2#3#4#5#6#7#8\\@@#9{%
4028 \ifcase\car#8\\nil\\or % Currenty <10000, but prepared for bigger
4029 \bbl@alphanumeric@ii{#9}000000#1\\or
4030 \bbl@alphanumeric@ii{#9}00000#1#2\\or
4031 \bbl@alphanumeric@ii{#9}0000#1#2#3\\or
4032 \bbl@alphanumeric@ii{#9}000#1#2#3#4\\else
4033 \bbl@alphnum@invalid{>9999}%
4034 \fi}
4035 \def\bbl@alphanumeric@ii#1#2#3#4#5#6#7#8{%
4036 \bbl@ifunset{\bbl@cntr@#1.F.\number#5#6#7#8@\\language}%
4037 {\bbl@cs{cntr@#1.4@\\language}{#5}
4038 \bbl@cs{cntr@#1.3@\\language}{#6}
4039 \bbl@cs{cntr@#1.2@\\language}{#7}
4040 \bbl@cs{cntr@#1.1@\\language}{#8}
4041 \ifnum#6#7#8>\z@ % TODO. An ad hoc rule for Greek. Ugly.
4042 \bbl@ifunset{\bbl@cntr@#1.S.321@\\language}{}}%
4043 {\bbl@cs{cntr@#1.S.321@\\language}}}%
4044 \fi}%
4045 {\bbl@cs{cntr@#1.F.\number#5#6#7#8@\\language}}}%
4046 \def\bbl@alphnum@invalid#1{%
4047 \bbl@error{Alphabetic numeral too large (#1)}%
4048 {Currently this is the limit.}}

```

The information in the identification section can be useful, so the following macro just exposes it with a user command.

```

4049 \newcommand\localeinfo[1]{%
4050 \bbl@ifunset{\bbl@csname bbl@info@#1\endcsname @\\language}%
4051 {\bbl@error{I've found no info for the current locale.\\%
4052 The corresponding ini file has not been loaded\\%
4053 Perhaps it doesn't exist}%
4054 {See the manual for details.}}%
4055 {\bbl@cs{\csname bbl@info@#1\endcsname @\\language}}}%
4056 % \@namedef{\bbl@info@name.locale}{lcname}
4057 \@namedef{\bbl@info@tag.ini}{lini}
4058 \@namedef{\bbl@info@name.english}{elname}
4059 \@namedef{\bbl@info@name.opentype}{lname}
4060 \@namedef{\bbl@info@tag.bcp47}{tbc}
4061 \@namedef{\bbl@info@language.tag.bcp47}{lbc}
4062 \@namedef{\bbl@info@tag.opentype}{lotf}
4063 \@namedef{\bbl@info@script.name}{esname}

```

```

4064 \@namedef{bbl@info@script.name.opentype}{sname}
4065 \@namedef{bbl@info@script.tag.bcp47}{sbcp}
4066 \@namedef{bbl@info@script.tag.opentype}{sotf}
4067 \let\bbl@ensureinfo\@gobble
4068 \newcommand\BabelEnsureInfo{%
4069   \ifx\InputIfFileExists\undefined\else
4070     \def\bbl@ensureinfo##1{%
4071       \bbl@ifunset{bbl@lname@##1}{\bbl@load@info{##1}}{}}%
4072   \fi
4073   \bbl@foreach\bbl@loaded{%
4074     \def\language{##1}%
4075     \bbl@ensureinfo{##1}}}%

```

More general, but non-expandable, is `\getlocaleproperty`. To inspect every possible loaded ini, we define `\LocaleForEach`, where `\bbl@ini@loaded` is a comma-separated list of locales, built by `\bbl@read@ini`.

```

4076 \newcommand\getlocaleproperty{%
4077   \@ifstar\bbl@getproperty@s\bbl@getproperty@x}
4078 \def\bbl@getproperty@s#1#2#3{%
4079   \let#1\relax
4080   \def\bbl@elt##1##2##3{%
4081     \bbl@ifsamestring{##1/##2}{#3}%
4082     {\providecommand#1{##3}%
4083     \def\bbl@elt####1####2####3{}}}%
4084   {}}%
4085   \bbl@cs{inidata@#2}}%
4086 \def\bbl@getproperty@x#1#2#3{%
4087   \bbl@getproperty@s{#1}{#2}{#3}%
4088   \ifx#1\relax
4089     \bbl@error
4090     {Unknown key for locale '#2':\%
4091     #3\%
4092     \string#1 will be set to \relax}%
4093     {Perhaps you misspelled it.}%
4094   \fi}
4095 \let\bbl@ini@loaded\@empty
4096 \newcommand\LocaleForEach{\bbl@foreach\bbl@ini@loaded}

```

## 10 Adjusting the Babel bahavior

A generic high level interface is provided to adjust some global and general settings.

```

4097 \newcommand\babeladjust[1]{% TODO. Error handling.
4098   \bbl@forkv{#1}{%
4099     \bbl@ifunset{bbl@ADJ@##1@##2}%
4100     {\bbl@cs{ADJ@##1}{##2}}%
4101     {\bbl@cs{ADJ@##1@##2}}}
4102 %
4103 \def\bbl@adjust@lua#1#2{%
4104   \ifvmode
4105     \ifnum\currentgrouplevel=\z@
4106       \directlua{ Babel.#2 }%
4107       \expandafter\expandafter\expandafter\@gobble
4108     \fi
4109   \fi
4110   {\bbl@error % The error is gobbled if everything went ok.
4111     {Currently, #1 related features can be adjusted only\%
4112     in the main vertical list.}%
4113     {Maybe things change in the future, but this is what it is.}}}

```

```

4114 \@namedef{bbl@ADJ@bidi.mirroring@on}{%
4115   \bbl@adjust@lua{bidi}{mirroring_enabled=true}}
4116 \@namedef{bbl@ADJ@bidi.mirroring@off}{%
4117   \bbl@adjust@lua{bidi}{mirroring_enabled=false}}
4118 \@namedef{bbl@ADJ@bidi.text@on}{%
4119   \bbl@adjust@lua{bidi}{bidi_enabled=true}}
4120 \@namedef{bbl@ADJ@bidi.text@off}{%
4121   \bbl@adjust@lua{bidi}{bidi_enabled=false}}
4122 \@namedef{bbl@ADJ@bidi.mapdigits@on}{%
4123   \bbl@adjust@lua{bidi}{digits_mapped=true}}
4124 \@namedef{bbl@ADJ@bidi.mapdigits@off}{%
4125   \bbl@adjust@lua{bidi}{digits_mapped=false}}
4126 %
4127 \@namedef{bbl@ADJ@linebreak.sea@on}{%
4128   \bbl@adjust@lua{linebreak}{sea_enabled=true}}
4129 \@namedef{bbl@ADJ@linebreak.sea@off}{%
4130   \bbl@adjust@lua{linebreak}{sea_enabled=false}}
4131 \@namedef{bbl@ADJ@linebreak.cjk@on}{%
4132   \bbl@adjust@lua{linebreak}{cjk_enabled=true}}
4133 \@namedef{bbl@ADJ@linebreak.cjk@off}{%
4134   \bbl@adjust@lua{linebreak}{cjk_enabled=false}}
4135 %
4136 \def\bbl@adjust@layout#1{%
4137   \ifvmode
4138     #1%
4139   \expandafter\@gobble
4140   \fi
4141   {\bbl@error   % The error is gobbled if everything went ok.
4142     {Currently, layout related features can be adjusted only\\%
4143       in vertical mode.}%
4144     {Maybe things change in the future, but this is what it is.}}}
4145 \@namedef{bbl@ADJ@layout.tabular@on}{%
4146   \bbl@adjust@layout{\let\@tabular\bbl@NL@@tabular}}
4147 \@namedef{bbl@ADJ@layout.tabular@off}{%
4148   \bbl@adjust@layout{\let\@tabular\bbl@OL@@tabular}}
4149 \@namedef{bbl@ADJ@layout.lists@on}{%
4150   \bbl@adjust@layout{\let\list\bbl@NL@list}}
4151 \@namedef{bbl@ADJ@layout.lists@off}{%
4152   \bbl@adjust@layout{\let\list\bbl@OL@list}}
4153 \@namedef{bbl@ADJ@hyphenation.extra@on}{%
4154   \bbl@activateposthyphen}
4155 %
4156 \@namedef{bbl@ADJ@autoload.bcp47@on}{%
4157   \bbl@bcpallowedtrue}
4158 \@namedef{bbl@ADJ@autoload.bcp47@off}{%
4159   \bbl@bcpallowedfalse}
4160 \@namedef{bbl@ADJ@autoload.bcp47.prefix#1{%
4161   \def\bbl@bcp@prefix{#1}}
4162 \def\bbl@bcp@prefix{bcp47-}
4163 \@namedef{bbl@ADJ@autoload.options#1{%
4164   \def\bbl@autoload@options{#1}}
4165 \let\bbl@autoload@bcptoptions\@empty
4166 \@namedef{bbl@ADJ@autoload.bcp47.options#1{%
4167   \def\bbl@autoload@bcptoptions{#1}}
4168 \newif\ifbbl@bcptoname
4169 \@namedef{bbl@ADJ@bcp47.toname@on}{%
4170   \bbl@bcptonametrue}
4171   \BabelEnsureInfo}
4172 \@namedef{bbl@ADJ@bcp47.toname@off}{%

```

```

4173 \bbl@bcptonamefalse}
4174% TODO: use babel name, override
4175%
4176% As the final task, load the code for lua.
4177%
4178 \ifx\directlua\@undefined\else
4179 \ifx\bbl@luapatterns\@undefined
4180 \input luababel.def
4181 \fi
4182 \fi
4183 </core>

A proxy file for switch.def

4184 <*kernel>
4185 \let\bbl@onlyswitch\@empty
4186 \input babel.def
4187 \let\bbl@onlyswitch\@undefined
4188 </kernel>
4189 <*patterns>

```

## 11 Loading hyphenation patterns

The following code is meant to be read by  $\text{\texttt{initex}}$  because it should instruct  $\text{\texttt{T\TeX}}$  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.

To make sure that  $\text{\texttt{L\TeX}}$  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.

```

4190 <<Make sure ProvidesFile is defined>>
4191 \ProvidesFile{hyphen.cfg}[<<date>>] <<version>> Babel hyphens]
4192 \xdef\bbl@format{\jobname}
4193 \def\bbl@version{<<version>>}
4194 \def\bbl@date{<<date>>}
4195 \ifx\AtBeginDocument\@undefined
4196 \def\@empty{}
4197 \let\orig@dump\dump
4198 \def\dump{%
4199 \ifx\@ztryfc\@undefined
4200 \else
4201 \toks0=\expandafter{\@preamblecmds}%
4202 \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
4203 \def\@begindocumenthook{}%
4204 \fi
4205 \let\dump\orig@dump\let\orig@dump\@undefined\dump}
4206 \fi
4207 <<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.

```

4208 \def\process@line#1#2 #3 #4 {%
4209 \ifx=#1%
4210 \process@synonym{#2}%

```

```

4211 \else
4212 \process@language{#1#2}{#3}{#4}%
4213 \fi
4214 \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.

```

4215 \toks@{}
4216 \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.

```

4217 \def\process@synonym#1{%
4218 \ifnum\last@language=\m@ne
4219 \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
4220 \else
4221 \expandafter\chardef\csname l@#1\endcsname\last@language
4222 \wlog{\string\l@#1=\string\language\the\last@language}%
4223 \expandafter\let\csname #1hyphenmins\expandafter\endcsname
4224 \csname\language\hyphenmins\endcsname
4225 \let\bbl@elt\relax
4226 \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}{}}}%
4227 \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` or `\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.

```

4228 \def\process@language#1#2#3{%
4229 \expandafter\addlanguage\csname l@#1\endcsname
4230 \expandafter\language\csname l@#1\endcsname
4231 \edef\language{#1}%
4232 \bbl@hook@everylanguage{#1}%

```

```

4233 % > luatex
4234 \bbl@get@enc#1::\@@@
4235 \begingroup
4236   \lefthyphenmin\m@ne
4237   \bbl@hook@loadpatterns{#2}%
4238   % > luatex
4239   \ifnum\lefthyphenmin=\m@ne
4240   \else
4241     \expandafter\xdef\csname #1hyphenmins\endcsname{%
4242       \the\lefthyphenmin\the\righthyphenmin}%
4243   \fi
4244 \endgroup
4245 \def\bbl@tempa{#3}%
4246 \ifx\bbl@tempa\@empty\else
4247   \bbl@hook@loadexceptions{#3}%
4248   % > luatex
4249 \fi
4250 \let\bbl@elt\relax
4251 \edef\bbl@languages{%
4252   \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
4253 \ifnum\the\language=\z@
4254   \expandafter\ifx\csname #1hyphenmins\endcsname\relax
4255     \set@hyphenmins\tw@\thr@@\relax
4256   \else
4257     \expandafter\expandafter\expandafter\set@hyphenmins
4258     \csname #1hyphenmins\endcsname
4259   \fi
4260   \the\toks@
4261   \toks@{}%
4262 \fi}

```

\bbl@get@enc The macro \bbl@get@enc extracts the font encoding from the language name and stores it in  
\bbl@hyph@enc \bbl@hyph@enc. It uses delimited arguments to achieve this.

```

4263 \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. loadkernel currently loads nothing, but define some basic macros instead.

```

4264 \def\bbl@hook@everylanguage#1{}
4265 \def\bbl@hook@loadpatterns#1{\input #1\relax}
4266 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
4267 \def\bbl@hook@loadkernel#1{%
4268   \def\addlanguage{\csname newlanguage\endcsname}%
4269   \def\adddialect##1##2{%
4270     \global\chardef##1##2\relax
4271     \wlog{\string##1 = a dialect from \string\language##2}}%
4272   \def\iflanguage##1{%
4273     \expandafter\ifx\csname l@##1\endcsname\relax
4274       \@nolanerr{##1}%
4275     \else
4276       \ifnum\csname l@##1\endcsname=\language
4277         \expandafter\expandafter\expandafter\@firstoftwo
4278       \else
4279         \expandafter\expandafter\expandafter\@secondoftwo
4280       \fi
4281     \fi}%
4282   \def\providehyphenmins##1##2{%
4283     \expandafter\ifx\csname ##1hyphenmins\endcsname\relax

```

```

4284 \namedef{##1hyphenmins}{##2}%
4285 \fi}%
4286 \def\set@hyphenmins##1##2{%
4287 \lefthyphenmin##1\relax
4288 \righthyphenmin##2\relax}%
4289 \def\selectlanguage{%
4290 \errhelp{Selecting a language requires a package supporting it}%
4291 \errmessage{Not loaded}}%
4292 \let\foreignlanguage\selectlanguage
4293 \let\otherlanguage\selectlanguage
4294 \expandafter\let\csname otherlanguage*\endcsname\selectlanguage
4295 \def\bbl@usehooks##1##2{% TODO. Temporary!!
4296 \def\setlocale{%
4297 \errhelp{Find an armchair, sit down and wait}%
4298 \errmessage{Not yet available}}%
4299 \let\uselocale\setlocale
4300 \let\locale\setlocale
4301 \let\selectlocale\setlocale
4302 \let\localename\setlocale
4303 \let\textlocale\setlocale
4304 \let\textlanguage\setlocale
4305 \let\languagegettext\setlocale}
4306 \begingroup
4307 \def\AddBabelHook#1#2{%
4308 \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
4309 \def\next{\toks1}%
4310 \else
4311 \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname####1}%
4312 \fi
4313 \next}
4314 \ifx\directlua\undefined
4315 \ifx\XeTeXinputencoding\undefined\else
4316 \input xebabel.def
4317 \fi
4318 \else
4319 \input luababel.def
4320 \fi
4321 \openin1 = babel-\bbl@format.cfg
4322 \ifeof1
4323 \else
4324 \input babel-\bbl@format.cfg\relax
4325 \fi
4326 \closein1
4327 \endgroup
4328 \bbl@hook@loadkernel{switch.def}

```

\readconfigfile The configuration file can now be opened for reading.

```

4329 \openin1 = language.dat

```

See if the file exists, if not, use the default hyphenation file hyphen.tex. The user will be informed about this.

```

4330 \def\language{english}%
4331 \ifeof1
4332 \message{I couldn't find the file language.dat,\space
4333 I will try the file hyphen.tex}
4334 \input hyphen.tex\relax
4335 \chardef\l@english\z@
4336 \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`.

```
4337 \last@language\m@ne
```

We now read lines from the file until the end is found. 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.

```
4338 \loop
4339 \endlinechar\m@ne
4340 \read1 to \bbl@line
4341 \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.

```
4342 \if T\ifeof1F\fi T\relax
4343 \ifx\bbl@line\@empty\else
4344 \edef\bbl@line{\bbl@line\space\space\space}%
4345 \expandafter\process@line\bbl@line\relax
4346 \fi
4347 \repeat
```

Check for the end of the file. We must reverse the test for `\ifeof` without `\else`. Then reactivate the default patterns, and close the configuration file.

```
4348 \begingroup
4349 \def\bbl@elt#1#2#3#4{%
4350 \global\language=#2\relax
4351 \gdef\language#1}%
4352 \def\bbl@elt##1##2##3##4{}}%
4353 \bbl@languages
4354 \endgroup
4355 \fi
4356 \closein1
```

We add a message about the fact that babel is loaded in the format and with which language patterns to the `\everyjob` register.

```
4357 \if/\the\toks@\else
4358 \errhelp{language.dat loads no language, only synonyms}
4359 \errmessage{Orphan language synonym}
4360 \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.

```
4361 \let\bbl@line\@undefined
4362 \let\process@line\@undefined
4363 \let\process@synonym\@undefined
4364 \let\process@language\@undefined
4365 \let\bbl@get@enc\@undefined
4366 \let\bbl@hyph@enc\@undefined
4367 \let\bbl@tempa\@undefined
4368 \let\bbl@hook@loadkernel\@undefined
4369 \let\bbl@hook@everylanguage\@undefined
4370 \let\bbl@hook@loadpatterns\@undefined
4371 \let\bbl@hook@loadexceptions\@undefined
4372 </patterns>
```

Here the code for `iniTeX` ends.

## 12 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].

```
4373 <<*More package options>> ≡
4374 \chardef\bbl@bidimode\z@
4375 \DeclareOption{bidi=default}{\chardef\bbl@bidimode=\@ne}
4376 \DeclareOption{bidi=basic}{\chardef\bbl@bidimode=101 }
4377 \DeclareOption{bidi=basic-r}{\chardef\bbl@bidimode=102 }
4378 \DeclareOption{bidi=bidi}{\chardef\bbl@bidimode=201 }
4379 \DeclareOption{bidi=bidi-r}{\chardef\bbl@bidimode=202 }
4380 \DeclareOption{bidi=bidi-l}{\chardef\bbl@bidimode=203 }
4381 <</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`.

At the time of this writing, fontspec shows a warning about there are languages not available, which some people think refers to babel, even if there is nothing wrong. Here is hack to patch fontspec to avoid the misleading message, which is replaced by a more explanatory one.

```
4382 <<*Font selection>> ≡
4383 \bbl@trace{Font handling with fontspec}
4384 \ifx\ExplSyntaxOn\@undefined\else
4385   \ExplSyntaxOn
4386   \catcode\ =10
4387   \def\bbl@loadfontspec{%
4388     \usepackage{fontspec}%
4389     \expandafter
4390     \def\csname msg-text->~fontspec/language-not-exist\endcsname##1##2##3##4{%
4391       Font '\l_fontspec_fontname_tl' is using the\\%
4392       default features for language '##1'.\\%
4393       That's usually fine, because many languages\\%
4394       require no specific features, but if the output is\\%
4395       not as expected, consider selecting another font.}
4396     \expandafter
4397     \def\csname msg-text->~fontspec/no-script\endcsname##1##2##3##4{%
4398       Font '\l_fontspec_fontname_tl' is using the\\%
4399       default features for script '##2'.\\%
4400       That's not always wrong, but if the output is\\%
4401       not as expected, consider selecting another font.}}
4402   \ExplSyntaxOff
4403 \fi
4404 \@onlypreamble\babelfont
4405 \newcommand\babelfont[2][{}]{% 1=langs/scripts 2=fam
4406   \bbl@foreach{#1}{%
4407     \expandafter\ifx\csname date##1\endcsname\relax
4408       \IfFileExists{babel-##1.tex}%
4409       {\babelprovide{##1}}}%
4410   }%
4411   \fi}%
4412 \edef\bbl@tempa{#1}%
4413 \def\bbl@tempb{#2}% Used by \bbl@bblfont
4414 \ifx\fontspec\@undefined
4415   \bbl@loadfontspec
4416 \fi
4417 \EnableBabelHook{babel-fontspec}% Just calls \bbl@switchfont
4418 \bbl@bblfont}
4419 \newcommand\bbl@bblfont[2][{}]{% 1=features 2=fontname, @font=rm|sf|tt
4420   \bbl@ifunset{\bbl@tempb family}%
```

```

4421 {\bbl@providedefam{\bbl@tempb}}%
4422 {\bbl@exp{%
4423   \\bbl@sreplace\<\bbl@tempb family >%
4424   {\@nameuse{\bbl@tempb default}}{\<\bbl@tempb default>}}}%
4425 % For the default font, just in case:
4426 \bbl@ifunset{\bbl@lsys@\language}\bbl@provide@lsys{\language}}{}%
4427 \expandafter\bbl@ifblank\expandafter{\bbl@tempa}%
4428 {\bbl@csarg\edef{\bbl@tempb dflt@}{<{\#1}{\#2}}% save bbl@rmdflt@
4429 \bbl@exp{%
4430   \let\<\bbl@\bbl@tempb dflt@\language>\<\bbl@\bbl@tempb dflt@>%
4431   \\bbl@font@set\<\bbl@\bbl@tempb dflt@\language>%
4432   \<\bbl@tempb default>\<\bbl@tempb family>}}}%
4433 {\bbl@foreach\bbl@tempa{% ie bbl@rmdflt@lang / *scrt
4434   \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:

```

4435 \def\bbl@providedefam#1{%
4436   \bbl@exp{%
4437     \\newcommand\<#1default>{}% Just define it
4438     \\bbl@add@list\\bbl@font@fams{#1}%
4439     \\DeclareRobustCommand\<#1family>{%
4440       \\not@math@alphabet\<#1family>\relax
4441       \\fontfamily\<#1default>\\selectfont}%
4442     \\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.

```

4443 \def\bbl@nostdfont#1{%
4444   \bbl@ifunset{\bbl@WFF@\f@family}%
4445   {\bbl@csarg\gdef{\bbl@WFF@\f@family}}{}% Flag, to avoid dupl warns
4446   \bbl@infowarn{The current font is not a babel standard family:\\%
4447     #1%
4448     \fontname\font\\%
4449     There is nothing intrinsically wrong with this warning, and\\%
4450     you can ignore it altogether if you do not need these\\%
4451     families. But if they are used in the document, you should be\\%
4452     aware 'babel' will no set Script and Language for them, so\\%
4453     you may consider defining a new family with \string\babelfont.\\%
4454     See the manual for further details about \string\babelfont.\\%
4455     Reported}}
4456   {}}}%
4457 \gdef\bbl@switchfont{%
4458   \bbl@ifunset{\bbl@lsys@\language}\bbl@provide@lsys{\language}}{}%
4459   \bbl@exp{% eg Arabic -> arabic
4460     \lowercase{\edef\\bbl@tempa{\bbl@cl{sname}}}}}%
4461   \bbl@foreach\bbl@font@fams{%
4462     \bbl@ifunset{\bbl@##1dflt@\language}% (1) language?
4463     {\bbl@ifunset{\bbl@##1dflt@*\bbl@tempa}% (2) from script?
4464       {\bbl@ifunset{\bbl@##1dflt@}% 2=F - (3) from generic?
4465         {}}% 123=F - nothing!
4466         {\bbl@exp{% 3=T - from generic
4467           \global\let\<\bbl@##1dflt@\language>%
4468           \<\bbl@##1dflt@>}}}%
4469         {\bbl@exp{% 2=T - from script
4470           \global\let\<\bbl@##1dflt@\language>%
4471           \<\bbl@##1dflt@*\bbl@tempa>}}}%
4472       {}}% 1=T - language, already defined
4473   \def\bbl@tempa{\bbl@nostdfont}}}%
4474   \bbl@foreach\bbl@font@fams{% don't gather with prev for

```

```

4475 \bbl@ifunset{\bbl@##1dflt@\languagename}%
4476 {\bbl@cs{famrst@##1}%
4477 \global\bbl@csarg\let{famrst@##1}\relax}%
4478 {\bbl@exp{% order is relevant. TODO: but sometimes wrong!
4479 \\\bbl@add\\\originalTeX{%
4480 \\\bbl@font@rst{\bbl@cl{##1dflt}}}%
4481 \<##1default>\<##1family>{##1}}%
4482 \\\bbl@font@set{\bbl@##1dflt@\languagename}% the main part!
4483 \<##1default>\<##1family>}}}%
4484 \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.

```

4485 \ifx\fbfamily\undefined\else % if latex
4486 \ifcase\bbl@engine % if pdftex
4487 \let\bbl@ckeckstdfonts\relax
4488 \else
4489 \def\bbl@ckeckstdfonts{%
4490 \begingroup
4491 \global\let\bbl@ckeckstdfonts\relax
4492 \let\bbl@tempa\@empty
4493 \bbl@foreach\bbl@font@fams{%
4494 \bbl@ifunset{\bbl@##1dflt@}%
4495 {\@nameuse{##1family}%
4496 \bbl@csarg\gdef{WFF@\fbfamily}}}% Flag
4497 \bbl@exp{\bbl@add\\\bbl@tempa{* \<##1family>= \fbfamily\\}%
4498 \space\space\fontname\font\\}%
4499 \bbl@csarg\xdef{##1dflt@}{\fbfamily}%
4500 \expandafter\xdef\csname ##1default\endcsname{\fbfamily}%
4501 }%
4502 \ifx\bbl@tempa\@empty\else
4503 \bbl@infowarn{The following font families will use the default\\%
4504 settings for all or some languages:\\%
4505 \bbl@tempa
4506 There is nothing intrinsically wrong with it, but\\%
4507 'babel' will no set Script and Language, which could\\%
4508 be relevant in some languages. If your document uses\\%
4509 these families, consider redefining them with \string\babelfont.\\%
4510 Reported}%
4511 \fi
4512 \endgroup}
4513 \fi
4514 \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.

```

4515 \def\bbl@font@set#1#2#3{% eg \bbl@rmdflt@lang \rmdefault \rmfamily
4516 \bbl@xin@{<>}{#1}%
4517 \ifin@
4518 \bbl@exp{\bbl@fontspec@set\\#1\expandafter@gobbletwo#1\\#3}%
4519 \fi
4520 \bbl@exp{%
4521 \def\\#2{#1}% eg, \rmdefault{\bbl@rmdflt@lang}
4522 \\\bbl@ifsamestring{#2}{\fbfamily}%
4523 {\#3%
4524 \\\bbl@ifsamestring{\fbseries}{\bfdefault}{\bfseries}}}%
4525 \let\\bbl@tempa\relax}%

```

```

4526     {}}}
4527 %      TODO - next should be global?, but even local does its job. I'm
4528 %      still not sure -- must investigate:
4529 \def\bbl@fontspec@set#1#2#3#4{% eg \bbl@rmdflt@lang fnt-opt fnt-nme \xxfamily
4530 \let\bbl@tempe\bbl@mapselect
4531 \let\bbl@mapselect\relax
4532 \let\bbl@temp@fam#4%      eg, '\rmfamily', to be restored below
4533 \let#4\@empty      %      Make sure \renewfontfamily is valid
4534 \bbl@exp{%
4535   \let\\bbl@temp@pfam\<\bbl@stripslash#4\space>% eg, '\rmfamily '
4536   \<keys_if_exist:nnF>{fontspec-opentype}{Script/\bbl@cl{sname}}}%
4537   {\newfontscript{\bbl@cl{sname}}{\bbl@cl{sotf}}}%
4538   \<keys_if_exist:nnF>{fontspec-opentype}{Language/\bbl@cl{lname}}}%
4539   {\newfontlanguage{\bbl@cl{lname}}{\bbl@cl{lotf}}}%
4540   \\renewfontfamily\\#4%
4541   [\bbl@cs{lsys@language},#2]}{#3}% ie \bbl@exp{.}{#3}
4542 \begingroup
4543   #4%
4544   \xdef#1{\f@family}%      eg, \bbl@rmdflt@lang{FreeSerif(0)}
4545 \endgroup
4546 \let#4\bbl@temp@fam
4547 \bbl@exp{\let\<\bbl@stripslash#4\space>\bbl@temp@pfam
4548 \let\bbl@mapselect\bbl@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.

```

4549 \def\bbl@font@rst#1#2#3#4{%
4550 \bbl@csarg\def{famrst@#4}{\bbl@font@set{#1}#2#3}%

```

The default font families. They are eurocentric, but the list can be expanded easily with \babelfont.

```

4551 \def\bbl@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 :-).

```

4552 \newcommand\babelFSstore[2][{%
4553 \bbl@ifblank{#1}%
4554 {\bbl@csarg\def{sname@#2}{Latin}}%
4555 {\bbl@csarg\def{sname@#2}{#1}}%
4556 \bbl@provide@dirs{#2}%
4557 \bbl@csarg\ifnum{wdir@#2}>\z@
4558 \let\bbl@beforeforeign\leavevmode
4559 \EnableBabelHook{babel-bidi}%
4560 \fi
4561 \bbl@foreach{#2}{%
4562 \bbl@FSstore{##1}{rm}\rmdefault\bbl@save@rmdefault
4563 \bbl@FSstore{##1}{sf}\sfdefault\bbl@save@sfdefault
4564 \bbl@FSstore{##1}{tt}\ttdefault\bbl@save@ttdefault}}
4565 \def\bbl@FSstore#1#2#3#4{%
4566 \bbl@csarg\edef{#2default#1}{#3}%
4567 \expandafter\addto\csname extras#1\endcsname{%
4568 \let#4#3%
4569 \ifx#3\f@family
4570 \edef#3{\csname bbl@#2default#1\endcsname}%
4571 \fontfamily{#3}\selectfont
4572 \else
4573 \edef#3{\csname bbl@#2default#1\endcsname}%
4574 \fi}%
4575 \expandafter\addto\csname noextras#1\endcsname{%

```

```

4576 \ifx#3\f@family
4577 \fontfamily{#4}\selectfont
4578 \fi
4579 \let#3#4}}
4580 \let\bbbl@langfeatures\@empty
4581 \def\babelFSfeatures{% make sure \fontspec is redefined once
4582 \let\bbbl@ori@fontspec\fontspec
4583 \renewcommand\fontspec[1][{%
4584 \bbbl@ori@fontspec[\bbbl@langfeatures##1]}
4585 \let\babelFSfeatures\bbbl@FSfeatures
4586 \babelFSfeatures}
4587 \def\bbbl@FSfeatures#1#2{%
4588 \expandafter\addto\csname extras#1\endcsname{%
4589 \babel@save\bbbl@langfeatures
4590 \edef\bbbl@langfeatures{#2,}}
4591 <</Font selection>>

```

## 13 Hooks for XeTeX and LuaTeX

### 13.1 XeTeX

Unfortunately, the current encoding cannot be retrieved and therefore it is reset always to utf8, which seems a sensible default.

```

4592 <<(*Footnote changes)>> ≡
4593 \bbbl@trace{Bidi footnotes}
4594 \ifnum\bbbl@bidimode>\z@
4595 \def\bbbl@footnote#1#2#3{%
4596 \ifnextchar[%
4597 {\bbbl@footnote@o{#1}{#2}{#3}}%
4598 {\bbbl@footnote@x{#1}{#2}{#3}}}
4599 \long\def\bbbl@footnote@x#1#2#3#4{%
4600 \bgroup
4601 \select@language@x{\bbbl@main@language}%
4602 \bbbl@fn@footnote{#2#1{\ignorespaces#4}#3}%
4603 \egroup}
4604 \long\def\bbbl@footnote@o#1#2#3[#4]#5{%
4605 \bgroup
4606 \select@language@x{\bbbl@main@language}%
4607 \bbbl@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}%
4608 \egroup}
4609 \def\bbbl@footnotetext#1#2#3{%
4610 \ifnextchar[%
4611 {\bbbl@footnotetext@o{#1}{#2}{#3}}%
4612 {\bbbl@footnotetext@x{#1}{#2}{#3}}}
4613 \long\def\bbbl@footnotetext@x#1#2#3#4{%
4614 \bgroup
4615 \select@language@x{\bbbl@main@language}%
4616 \bbbl@fn@footnotetext{#2#1{\ignorespaces#4}#3}%
4617 \egroup}
4618 \long\def\bbbl@footnotetext@o#1#2#3[#4]#5{%
4619 \bgroup
4620 \select@language@x{\bbbl@main@language}%
4621 \bbbl@fn@footnotetext[#4]{#2#1{\ignorespaces#5}#3}%
4622 \egroup}
4623 \def\BabelFootnote#1#2#3#4{%
4624 \ifx\bbbl@fn@footnote\@undefined
4625 \let\bbbl@fn@footnote\footnote
4626 \fi

```

```

4627 \ifx\bb1@fn@footnotetext\@undefined
4628 \let\bb1@fn@footnotetext\footnotetext
4629 \fi
4630 \bb1@ifblank{#2}%
4631 {\def#1{\bb1@footnote{\@firstofone}{#3}{#4}}
4632 \@namedef{\bb1@stripslash#1text}%
4633 {\bb1@footnotetext{\@firstofone}{#3}{#4}}}%
4634 {\def#1{\bb1@exp{\bb1@footnote{\foreignlanguage{#2}}}{#3}{#4}}%
4635 \@namedef{\bb1@stripslash#1text}%
4636 {\bb1@exp{\bb1@footnotetext{\foreignlanguage{#2}}}{#3}{#4}}}%
4637 \fi
4638 <</Footnote changes>>

```

Now, the code.

```

4639 (*xetex)
4640 \def\BabelStringsDefault{unicode}
4641 \let\xebbl@stop\relax
4642 \AddBabelHook{xetex}{encodedcommands}{%
4643 \def\bb1@tempa{#1}%
4644 \ifx\bb1@tempa\@empty
4645 \XeTeXinputencoding"bytes"%
4646 \else
4647 \XeTeXinputencoding"#1"%
4648 \fi
4649 \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
4650 \AddBabelHook{xetex}{stopcommands}{%
4651 \xebbl@stop
4652 \let\xebbl@stop\relax}
4653 \def\bb1@intraspace#1 #2 #3@@{%
4654 \bb1@csarg\gdef\xeisp@\languageName}%
4655 {\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}}
4656 \def\bb1@intrapenalty#1@@{%
4657 \bb1@csarg\gdef\xeipn@\languageName}%
4658 {\XeTeXlinebreakpenalty #1\relax}}
4659 \def\bb1@provide@intraspace{%
4660 \bb1@xin@\bb1@cl{lnbrk}}{s}%
4661 \ifin@else\bb1@xin@\bb1@cl{lnbrk}}{c}\fi
4662 \ifin@
4663 \bb1@ifunset{\bb1@intsp@\languageName}{}%
4664 {\expandafter\ifx\csname \bb1@intsp@\languageName\endcsname\@empty\else
4665 \ifx\bb1@KVP@intraspace\@nil
4666 \bb1@exp{%
4667 \bb1@intraspace\bb1@cl{intsp}\@@}%
4668 \fi
4669 \ifx\bb1@KVP@intrapenalty\@nil
4670 \bb1@intrapenalty0\@@
4671 \fi
4672 \fi
4673 \ifx\bb1@KVP@intraspace\@nil\else % We may override the ini
4674 \expandafter\bb1@intraspace\bb1@KVP@intraspace\@@
4675 \fi
4676 \ifx\bb1@KVP@intrapenalty\@nil\else
4677 \expandafter\bb1@intrapenalty\bb1@KVP@intrapenalty\@@
4678 \fi
4679 \bb1@exp{%
4680 \bb1@add\<extras\languageName>{%
4681 \XeTeXlinebreaklocale "\bb1@cl{tbcpr}"%
4682 \<\bb1@xeisp@\languageName>%
4683 \<\bb1@xeipn@\languageName>}%

```

```

4684      \\\bbl@toglobal\<extras\language>%
4685      \\\bbl@add\<noextras\language>{%
4686      \XeTeXlinebreaklocale "en"%
4687      \\\bbl@toglobal\<noextras\language>}}%
4688      \ifx\bbl@ispace\@undefined
4689      \gdef\bbl@ispace{\bbl@cl{xisp}}%
4690      \ifx\AtBeginDocument\@notprerr
4691      \expandafter\@secondoftwo % to execute right now
4692      \fi
4693      \AtBeginDocument{%
4694      \expandafter\bbl@add
4695      \csname selectfont \endcsname{\bbl@ispace}%
4696      \expandafter\bbl@toglobal\csname selectfont \endcsname}%
4697      \fi}%
4698      \fi}
4699      \ifx\DisableBabelHook\@undefined\endinput\fi
4700      \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
4701      \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ckeckstdfonts}
4702      \DisableBabelHook{babel-fontspec}
4703      <<Font selection>>
4704      \input txtbabel.def
4705      </xetex>

```

## 13.2 Layout

*In progress.*

Note elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titles, and geometry.

\bbl@startskip and \bbl@endskip are available to package authors. Thanks to the T<sub>E</sub>X expansion mechanism the following constructs are valid: \adim\bbl@startskip, \advance\bbl@startskip\adim, \bbl@startskip\adim.

Consider txtbabel as a shorthand for *tex-xet babel*, which is the bidi model in both pdf<sub>TEX</sub> and xet<sub>EX</sub>.

```

4706 <*texxet>
4707 \providecommand\bbl@provide@intraspace{}
4708 \bbl@trace{Redefinitions for bidi layout}
4709 \def\bbl@sspre@caption{%
4710   \bbl@exp{\everyhbox{\\\bbl@texmdir\bbl@cs{wdir}\bbl@main@language}}}}
4711 \ifx\bbl@opt@layout\@nnil\endinput\fi % No layout
4712 \def\bbl@startskip{\ifcase\bbl@thepardir\leftskip\else\rightskip\fi}
4713 \def\bbl@endskip{\ifcase\bbl@thepardir\rightskip\else\leftskip\fi}
4714 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
4715   \def\@hangfrom#1{%
4716     \setbox\@tempboxa\hbox{#1}%
4717     \hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
4718     \noindent\box\@tempboxa}
4719   \def\raggedright{%
4720     \let\@centercr
4721     \bbl@startskip\z@skip
4722     \@rightskip\@flushglue
4723     \bbl@endskip\@rightskip
4724     \parindent\z@
4725     \parfillskip\bbl@startskip}
4726   \def\raggedleft{%
4727     \let\@centercr
4728     \bbl@startskip\@flushglue
4729     \bbl@endskip\z@skip
4730     \parindent\z@
4731     \parfillskip\bbl@endskip}

```



```

4732 \fi
4733 \IfBabelLayout{lists}
4734   {\bbl@sreplace\list
4735     {\totalleftmargin\leftmargin}{\totalleftmargin\bbl@listleftmargin}%
4736     \def\bbl@listleftmargin{%
4737       \ifcase\bbl@thepardir\leftmargin\else\rightmargin\fi}%
4738     \ifcase\bbl@engine
4739       \def\labelenumii{}\theenumii{}\pdfTeX doesn't reverse ()
4740       \def\p@enumii{\p@enumii}\theenumii}%
4741     \fi
4742     \bbl@sreplace\@verbatim
4743       {\leftskip\totalleftmargin}%
4744       {\bbl@startskip\textwidth
4745         \advance\bbl@startskip-\linewidth}%
4746     \bbl@sreplace\@verbatim
4747       {\rightskip\z@skip}%
4748       {\bbl@endskip\z@skip}}%
4749   {}
4750 \IfBabelLayout{contents}
4751   {\bbl@sreplace\@dottedtocline{\leftskip}{\bbl@startskip}%
4752     \bbl@sreplace\@dottedtocline{\rightskip}{\bbl@endskip}}
4753   {}
4754 \IfBabelLayout{columns}
4755   {\bbl@sreplace\@outputdblcol{\hb@xt@\textwidth}{\bbl@outputbox}%
4756     \def\bbl@outputbox#1{%
4757       \hb@xt@\textwidth{%
4758         \hskip\columnwidth
4759         \hfil
4760         {\normalcolor\vrule \@width\columnseprule}%
4761         \hfil
4762         \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
4763         \hskip-\textwidth
4764         \hb@xt@\columnwidth{\box\@outputbox \hss}%
4765         \hskip\columnsep
4766         \hskip\columnwidth}}}%
4767   {}
4768 \langle\langle Footnote changes \rangle\rangle
4769 \IfBabelLayout{footnotes}%
4770   {\BabelFootnote\footnote\language\language{}{}}%
4771   \BabelFootnote\localfootnote\language\language{}{}}%
4772   \BabelFootnote\mainfootnote{}{}{}%
4773   {}

```

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.

```

4774 \IfBabelLayout{counters}%
4775   {\let\bbl@latinarabic=\@arabic
4776     \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
4777     \let\bbl@asciroman=\@roman
4778     \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
4779     \let\bbl@asciiRoman=\@Roman
4780     \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}}%
4781 \langle\langle texet \rangle\rangle

```

### 13.3 LuaTeX

The 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, `\bbl@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`). FIX - This isn’t true anymore. 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.

This files is read at three places: (1) when `plain.def`, `babel.sty` starts, to read the list of available languages from `language.dat` (for the base option); (2) at `hyphen.cfg`, to modify some macros; (3) in the middle of `plain.def` and `babel.sty`, by `babel.def`, with the commands and other definitions for `luatex` (eg, `\babelpatterns`).

```

4782 (*luatex)
4783 \ifx\AddBabelHook\undefined % When plain.def, babel.sty starts
4784 \bbl@trace{Read language.dat}
4785 \ifx\bbl@readstream\undefined
4786   \csname newread\endcsname\bbl@readstream
4787 \fi
4788 \begingroup
4789   \toks@{}
4790   \count@z@ % 0=start, 1=0th, 2=normal
4791   \def\bbl@process@line#1#2 #3 #4 {%
4792     \ifx=#1%
4793       \bbl@process@synonym{#2}%
4794     \else
4795       \bbl@process@language{#1#2}{#3}{#4}%
4796     \fi
4797     \ignorespaces}
4798   \def\bbl@manylang{%
4799     \ifnum\bbl@last>\@ne
4800       \bbl@info{Non-standard hyphenation setup}%
4801     \fi
4802     \let\bbl@manylang\relax}
4803   \def\bbl@process@language#1#2#3{%
4804     \ifcase\count@
4805       \ifundefined{zth#1}{\count@\tw@}{\count@\@ne}%
4806     \or
4807       \count@\tw@
4808     \fi
4809     \ifnum\count@=\tw@

```

```

4810 \expandafter\addlanguage\csname l@#1\endcsname
4811 \language\allocationnumber
4812 \chardef\bbl@last\allocationnumber
4813 \bbl@manylang
4814 \let\bbl@elt\relax
4815 \xdef\bbl@languages{%
4816   \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
4817 \fi
4818 \the\toks@
4819 \toks@{}}
4820 \def\bbl@process@synonym@aux#1#2{%
4821   \global\expandafter\chardef\csname l@#1\endcsname#2\relax
4822   \let\bbl@elt\relax
4823   \xdef\bbl@languages{%
4824     \bbl@languages\bbl@elt{#1}{#2}{}}}%
4825 \def\bbl@process@synonym#1{%
4826   \ifcase\count@
4827     \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
4828   \or
4829     \ifundefined{zth@#1}{\bbl@process@synonym@aux{#1}{0}}}%
4830   \else
4831     \bbl@process@synonym@aux{#1}{\the\bbl@last}%
4832   \fi}
4833 \ifx\bbl@languages@undefined % Just a (sensible?) guess
4834   \chardef\l@english\z@
4835   \chardef\l@USenglish\z@
4836   \chardef\bbl@last\z@
4837   \global\@namedef{bbl@hyphendata@0}{\hyphen.tex}{}
4838   \gdef\bbl@languages{%
4839     \bbl@elt{english}{0}{\hyphen.tex}{}%
4840     \bbl@elt{USenglish}{0}{}}%
4841 \else
4842   \global\let\bbl@languages@format\bbl@languages
4843   \def\bbl@elt#1#2#3#4{% Remove all except language 0
4844     \ifnum#2>\z@\else
4845       \noexpand\bbl@elt{#1}{#2}{#3}{#4}%
4846     \fi}%
4847   \xdef\bbl@languages{\bbl@languages}%
4848 \fi
4849 \def\bbl@elt#1#2#3#4{\@namedef{zth@#1}{}} % Define flags
4850 \bbl@languages
4851 \openin\bbl@readstream=language.dat
4852 \ifeof\bbl@readstream
4853   \bbl@warning{I couldn't find language.dat. No additional\\%
4854     patterns loaded. Reported}%
4855 \else
4856   \loop
4857     \endlinechar\m@ne
4858     \read\bbl@readstream to \bbl@line
4859     \endlinechar\^^M
4860     \if T\ifeof\bbl@readstream F\fi T\relax
4861     \ifx\bbl@line\empty\else
4862       \edef\bbl@line{\bbl@line\space\space\space}%
4863       \expandafter\bbl@process@line\bbl@line\relax
4864     \fi
4865   \repeat
4866 \fi
4867 \endgroup
4868 \bbl@trace{Macros for reading patterns files}

```

```

4869 \def\bbl@get@enc#1:#2:#3\@@@{\def\bbl@hyph@enc{#2}}
4870 \ifx\babelcatcodetablenum\undefined
4871 \ifx\newcatcodetable\undefined
4872 \def\babelcatcodetablenum{5211}
4873 \def\bbl@pattcodes{\numexpr\babelcatcodetablenum+1\relax}
4874 \else
4875 \newcatcodetable\babelcatcodetablenum
4876 \newcatcodetable\bbl@pattcodes
4877 \fi
4878 \else
4879 \def\bbl@pattcodes{\numexpr\babelcatcodetablenum+1\relax}
4880 \fi
4881 \def\bbl@luapatterns#1#2{%
4882 \bbl@get@enc#1::\@@@
4883 \setbox\z@\hbox\bgroup
4884 \begingroup
4885 \savecatcodetable\babelcatcodetablenum\relax
4886 \initcatcodetable\bbl@pattcodes\relax
4887 \catcodetable\bbl@pattcodes\relax
4888 \catcode`\#=6 \catcode`\$=3 \catcode`\&=4 \catcode`\^=7
4889 \catcode`\_ =8 \catcode`\{=1 \catcode`\}=2 \catcode`\~=13
4890 \catcode`\@=11 \catcode`\^^I=10 \catcode`\^^J=12
4891 \catcode`\<=12 \catcode`\>=12 \catcode`\*=12 \catcode`\.=12
4892 \catcode`\-=12 \catcode`\/=12 \catcode`\[=12 \catcode`\]=12
4893 \catcode`\`=12 \catcode`\'=12 \catcode`\`=12
4894 \input #1\relax
4895 \catcodetable\babelcatcodetablenum\relax
4896 \endgroup
4897 \def\bbl@tempa{#2}%
4898 \ifx\bbl@tempa\empty\else
4899 \input #2\relax
4900 \fi
4901 \egroup}%
4902 \def\bbl@patterns@lua#1{%
4903 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
4904 \csname l@#1\endcsname
4905 \edef\bbl@tempa{#1}%
4906 \else
4907 \csname l@#1:\f@encoding\endcsname
4908 \edef\bbl@tempa{#1:\f@encoding}%
4909 \fi\relax
4910 \@namedef{lu@texhyphen@loaded@the\language}}}% Temp
4911 \@ifundefined{bbl@hyphendata@the\language}%
4912 {\def\bbl@elt##1##2##3##4{%
4913 \ifnum##2=\csname l@bbl@tempa\endcsname % #2=spanish, dutch:OT1...
4914 \def\bbl@tempb{##3}%
4915 \ifx\bbl@tempb\empty\else % if not a synonymous
4916 \def\bbl@tempc{##3}{##4}}%
4917 \fi
4918 \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4919 \fi}%
4920 \bbl@languages
4921 \@ifundefined{bbl@hyphendata@the\language}%
4922 {\bbl@info{No hyphenation patterns were set for\%
4923 language '\bbl@tempa'. Reported}}%
4924 {\expandafter\expandafter\expandafter\bbl@luapatterns
4925 \csname bbl@hyphendata@the\language\endcsname}}}%
4926 \endinput\fi
4927 % Here ends \ifx\AddBabelHook\undefined

```

```

4928 % A few lines are only read by hyphen.cfg
4929 \ifx\DisableBabelHook\undefined
4930 \AddBabelHook{luatex}{everylanguage}{%
4931   \def\process@language##1##2##3{%
4932     \def\process@line####1####2 ####3 ####4 {}}}
4933 \AddBabelHook{luatex}{loadpatterns}{%
4934   \input #1\relax
4935   \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
4936     {{#1}}}}
4937 \AddBabelHook{luatex}{loadexceptions}{%
4938   \input #1\relax
4939   \def\bbl@tempb##1##2{{##1}{##2}}%
4940   \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
4941     {\expandafter\expandafter\expandafter\bbl@tempb
4942       \csname bbl@hyphendata@the\language\endcsname}}
4943 \endinput\fi
4944 % Here stops reading code for hyphen.cfg
4945 % The following is read the 2nd time it's loaded
4946 \begingroup % TODO - to a lua file
4947 \catcode`\%=12
4948 \catcode`\'=12
4949 \catcode`\ "=12
4950 \catcode`\:=12
4951 \directlua{
4952   Babel = Babel or {}
4953   function Babel.bytes(line)
4954     return line:gsub(".",
4955       function (chr) return unicode.utf8.char(string.byte(chr)) end)
4956   end
4957   function Babel.begin_process_input()
4958     if luatexbase and luatexbase.add_to_callback then
4959       luatexbase.add_to_callback('process_input_buffer',
4960         Babel.bytes, 'Babel.bytes')
4961     else
4962       Babel.callback = callback.find('process_input_buffer')
4963       callback.register('process_input_buffer', Babel.bytes)
4964     end
4965   end
4966   function Babel.end_process_input ()
4967     if luatexbase and luatexbase.remove_from_callback then
4968       luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
4969     else
4970       callback.register('process_input_buffer', Babel.callback)
4971     end
4972   end
4973   function Babel.addpatterns(pp, lg)
4974     local lg = lang.new(lg)
4975     local pats = lang.patterns(lg) or ''
4976     lang.clear_patterns(lg)
4977     for p in pp:gmatch('[^s]+') do
4978       ss = ''
4979       for i in string.utfcharacters(p:gsub('%d', '')) do
4980         ss = ss .. '%d?' .. i
4981       end
4982       ss = ss:gsub('^%%d%?%', '%%.') .. '%d?'
4983       ss = ss:gsub('%.%%d%?$', '%%.')
4984       pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
4985       if n == 0 then
4986         tex.sprint(

```

```

4987         [[\string\csname\space bbl@info\endcsname{New pattern: }]
4988         .. p .. [{}]])
4989         pats = pats .. ' ' .. p
4990     else
4991         tex.sprint(
4992             [[\string\csname\space bbl@info\endcsname{Renew pattern: }]
4993             .. p .. [{}]])
4994     end
4995 end
4996 lang.patterns(lg, pats)
4997 end
4998 }
4999 \endgroup
5000 \ifx\newattribute\@undefined\else
5001   \newattribute\bbl@attr@locale
5002   \directlua{ Babel.attr_locale = luatexbase.registernumber'bbl@attr@locale'}
5003   \AddBabelHook{luatex}{beforeextras}{%
5004     \setattribute\bbl@attr@locale\localeid}
5005 \fi
5006 \def\BabelStringsDefault{unicode}
5007 \let\luabbl@stop\relax
5008 \AddBabelHook{luatex}{encodedcommands}{%
5009   \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
5010   \ifx\bbl@tempa\bbl@tempb\else
5011     \directlua{Babel.begin_process_input()}%
5012     \def\luabbl@stop{%
5013       \directlua{Babel.end_process_input()}}%
5014   \fi}%
5015 \AddBabelHook{luatex}{stopcommands}{%
5016   \luabbl@stop
5017   \let\luabbl@stop\relax}
5018 \AddBabelHook{luatex}{patterns}{%
5019   \@ifundefined{bbl@hyphendata@the\language}%
5020   {\def\bbl@elt##1##2##3##4{%
5021     \ifnum##2=\csname l@##2\endcsname % #2=spanish, dutch:OT1...
5022     \def\bbl@tempb{##3}%
5023     \ifx\bbl@tempb\@empty\else % if not a synonymous
5024       \def\bbl@tempc{##3}{##4}}%
5025     \fi
5026     \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
5027     \fi}%
5028   \bbl@languages
5029   \@ifundefined{bbl@hyphendata@the\language}%
5030   {\bbl@info{No hyphenation patterns were set for\%
5031     language '#2'. Reported}}%
5032   {\expandafter\expandafter\expandafter\bbl@luapatterns
5033     \csname bbl@hyphendata@the\language\endcsname}}}%
5034   \@ifundefined{bbl@patterns@}{}%
5035   \begingroup
5036     \bbl@xin@{, \number\language,}{, \bbl@pttnlist}%
5037     \ifin\else
5038       \ifx\bbl@patterns@\@empty\else
5039         \directlua{ Babel.addpatterns(
5040           [[\bbl@patterns@]], \number\language) }%
5041       \fi
5042       \@ifundefined{bbl@patterns@#1}%
5043       \@empty
5044       {\directlua{ Babel.addpatterns(
5045         [[\space\csname bbl@patterns@#1\endcsname]],

```

```

5046         \number\language) } }%
5047     \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
5048     \fi
5049 \endgroup}%
5050 \bbl@exp{%
5051     \bbl@ifunset{\bbl@prehc@\languagename}{ }%
5052     { \bbl@ifblank{\bbl@cs{\prehc@\languagename}}{ }%
5053     {\prehyphenchar=\bbl@c1{\prehc}\relax}}}%

```

**\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.

```

5054 \@onlypreamble\babelpatterns
5055 \AtEndOfPackage{%
5056     \newcommand\babelpatterns[2][\@empty]{%
5057         \ifx\bbl@patterns@\relax
5058             \let\bbl@patterns@\@empty
5059         \fi
5060         \ifx\bbl@pttnlist\@empty\else
5061             \bbl@warning{%
5062                 You must not intermingle \string\selectlanguage\space and\%
5063                 \string\babelpatterns\space or some patterns will not\%
5064                 be taken into account. Reported}%
5065             \fi
5066             \ifx\@empty#1%
5067                 \protected@edef\bbl@patterns@{\bbl@patterns@\space#2}%
5068             \else
5069                 \edef\bbl@tempb{\zap@space#1 \@empty}%
5070                 \bbl@for\bbl@tempa\bbl@tempb{%
5071                     \bbl@fixname\bbl@tempa
5072                     \bbl@iflanguage\bbl@tempa{%
5073                         \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
5074                             \ifundefined{\bbl@patterns@\bbl@tempa}%
5075                                 \@empty
5076                                 {\csname \bbl@patterns@\bbl@tempa\endcsname\space}%
5077                             #2}}}%
5078                 \fi}}

```

## 13.4 Southeast Asian scripts

First, some general code for line breaking, used by `\babelposthyphenation`. 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. See Unicode UAX 14.

```

5079% TODO - to a lua file
5080 \directlua{
5081     Babel = Babel or {}
5082     Babel.linebreaking = Babel.linebreaking or {}
5083     Babel.linebreaking.before = {}
5084     Babel.linebreaking.after = {}
5085     Babel.locale = {} % Free to use, indexed with \localeid
5086     function Babel.linebreaking.add_before(func)
5087         tex.print([[noexpand\csname \bbl@luahyphenate\endcsname]])
5088         table.insert(Babel.linebreaking.before, func)
5089     end
5090     function Babel.linebreaking.add_after(func)
5091         tex.print([[noexpand\csname \bbl@luahyphenate\endcsname]])
5092         table.insert(Babel.linebreaking.after, func)

```

```

5093 end
5094 }
5095 \def\bbl@intraspace#1 #2 #3\@{
5096 \directlua{
5097   Babel = Babel or {}
5098   Babel.intraspaces = Babel.intraspaces or {}
5099   Babel.intraspaces['\csname bbl@sbc@language\endcsname'] = %
5100     {b = #1, p = #2, m = #3}
5101   Babel.locale_props[\the\localeid].intraspace = %
5102     {b = #1, p = #2, m = #3}
5103 }}
5104 \def\bbl@intrapenalty#1\@{
5105 \directlua{
5106   Babel = Babel or {}
5107   Babel.intrapenalties = Babel.intrapenalties or {}
5108   Babel.intrapenalties['\csname bbl@sbc@language\endcsname'] = #1
5109   Babel.locale_props[\the\localeid].intrapenalty = #1
5110 }}
5111 \begingroup
5112 \catcode`\%=12
5113 \catcode`\^=14
5114 \catcode`\'=12
5115 \catcode`\~=12
5116 \gdef\bbl@seaintraspace{^
5117 \let\bbl@seaintraspace\relax
5118 \directlua{
5119   Babel = Babel or {}
5120   Babel.sea_enabled = true
5121   Babel.sea_ranges = Babel.sea_ranges or {}
5122   function Babel.set_chranges (script, chrng)
5123     local c = 0
5124     for s, e in string.gmatch(chrng..' ', '(.-%.%.(-)%s') do
5125       Babel.sea_ranges[script..c]={tonumber(s,16), tonumber(e,16)}
5126       c = c + 1
5127     end
5128   end
5129   function Babel.sea_disc_to_space (head)
5130     local sea_ranges = Babel.sea_ranges
5131     local last_char = nil
5132     local quad = 655360 ^% 10 pt = 655360 = 10 * 65536
5133     for item in node.traverse(head) do
5134       local i = item.id
5135       if i == node.id'glyph' then
5136         last_char = item
5137       elseif i == 7 and item.subtype == 3 and last_char
5138         and last_char.char > 0x0C99 then
5139         quad = font.getfont(last_char.font).size
5140         for lg, rg in pairs(sea_ranges) do
5141           if last_char.char > rg[1] and last_char.char < rg[2] then
5142             lg = lg:sub(1, 4) ^% Remove trailing number of, eg, Cyr11
5143             local intraspace = Babel.intraspaces[lg]
5144             local intrapenalty = Babel.intrapenalties[lg]
5145             local n
5146             if intrapenalty ~= 0 then
5147               n = node.new(14, 0) ^% penalty
5148               n.penalty = intrapenalty
5149               node.insert_before(head, item, n)
5150             end
5151             n = node.new(12, 13) ^% (glue, spaceskip)

```



```

5152         node.setglue(n, intraspace.b * quad,
5153                     intraspace.p * quad,
5154                     intraspace.m * quad)
5155         node.insert_before(head, item, n)
5156         node.remove(head, item)
5157     end
5158 end
5159 end
5160 end
5161 end
5162 }^^
5163 \bbl@luahyphenate}
5164 \catcode`\%=14
5165 \gdef\bbl@cjkintraspac{%
5166     \let\bbl@cjkintraspac\relax
5167     \directlua{
5168         Babel = Babel or {}
5169         require('babel-data-cjk.lua')
5170         Babel.cjk_enabled = true
5171         function Babel.cjk_linebreak(head)
5172             local GLYPH = node.id'glyph'
5173             local last_char = nil
5174             local quad = 655360      % 10 pt = 655360 = 10 * 65536
5175             local last_class = nil
5176             local last_lang = nil
5177
5178             for item in node.traverse(head) do
5179                 if item.id == GLYPH then
5180
5181                     local lang = item.lang
5182
5183                     local LOCALE = node.get_attribute(item,
5184                 luatexbase.registernumber'bbl@attr@locale')
5185                     local props = Babel.locale_props[LOCALE]
5186
5187                     local class = Babel.cjk_class[item.char].c
5188
5189                     if class == 'cp' then class = 'cl' end % ]) as CL
5190                     if class == 'id' then class = 'I' end
5191
5192                     local br = 0
5193                     if class and last_class and Babel.cjk_breaks[last_class][class] then
5194                         br = Babel.cjk_breaks[last_class][class]
5195                     end
5196
5197                     if br == 1 and props.linebreak == 'c' and
5198                         lang ~= \the\l@nohyphenation\space and
5199                         last_lang ~= \the\l@nohyphenation then
5200                         local intrapenalty = props.intrapenalty
5201                         if intrapenalty ~= 0 then
5202                             local n = node.new(14, 0)      % penalty
5203                             n.penalty = intrapenalty
5204                             node.insert_before(head, item, n)
5205                         end
5206                         local intraspace = props.intraspace
5207                         local n = node.new(12, 13)          % (glue, spaceskip)
5208                         node.setglue(n, intraspace.b * quad,
5209                                     intraspace.p * quad,
5210                                     intraspace.m * quad)

```

```

5211         node.insert_before(head, item, n)
5212     end
5213
5214     if font.getfont(item.font) then
5215         quad = font.getfont(item.font).size
5216     end
5217     last_class = class
5218     last_lang = lang
5219 else % if penalty, glue or anything else
5220     last_class = nil
5221 end
5222 end
5223 lang.hyphenate(head)
5224 end
5225 }%
5226 \bbl@luahyphenate}
5227 \gdef\bbl@luahyphenate{%
5228 \let\bbl@luahyphenate\relax
5229 \directlua{
5230     luatexbase.add_to_callback('hyphenate',
5231     function (head, tail)
5232         if Babel.linebreaking.before then
5233             for k, func in ipairs(Babel.linebreaking.before) do
5234                 func(head)
5235             end
5236         end
5237         if Babel.cjk_enabled then
5238             Babel.cjk_linebreak(head)
5239         end
5240         lang.hyphenate(head)
5241         if Babel.linebreaking.after then
5242             for k, func in ipairs(Babel.linebreaking.after) do
5243                 func(head)
5244             end
5245         end
5246         if Babel.sea_enabled then
5247             Babel.sea_disc_to_space(head)
5248         end
5249     end,
5250     'Babel.hyphenate')
5251 }
5252 }
5253 \endgroup
5254 \def\bbl@provide@intraspace{%
5255 \bbl@ifunset{\bbl@intsp@{language}}{%
5256     {\expandafter\ifx\csname bbl@intsp@{language}\endcsname\@empty\else
5257         \bbl@xin@{\bbl@cl{lnbrk}}{c}%
5258         \ifin@ % cjk
5259         \bbl@cjkintraspace
5260         \directlua{
5261             Babel = Babel or {}
5262             Babel.locale_props = Babel.locale_props or {}
5263             Babel.locale_props[\the\localeid].linebreak = 'c'
5264         }%
5265         \bbl@exp{\bbl@intraspace\bbl@cl{intsp}}{\@}%
5266         \ifx\bbl@KVP@intrapenalty\@nil
5267             \bbl@intrapenalty0\@
5268         \fi
5269     \else % sea

```

```

5270      \bbl@seaintraspace
5271      \bbl@exp{\bbl@intraspace\bbl@cl{intsp}\bbl@cl{intsp}\bbl@cl{intsp}\bbl@cl{intsp}}
5272      \directlua{
5273          Babel = Babel or {}
5274          Babel.sea_ranges = Babel.sea_ranges or {}
5275          Babel.set_chranges('\bbl@cl{sbcp}',
5276                          '\bbl@cl{chrng}')
5277      }%
5278      \ifx\bbl@KVP@intrapenalty\@nil
5279      \bbl@intrapenalty0\@@
5280      \fi
5281      \fi
5282      \fi
5283      \ifx\bbl@KVP@intrapenalty\@nil\else
5284      \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@@
5285      \fi}}

```

### 13.5 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.

```

5286 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
5287 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@cckstdfont}
5288 \DisableBabelHook{babel-fontspec}
5289 <<Font selection>>

```

### 13.6 Automatic fonts and ids switching

After defining the blocks for a number of scripts (must be extended and very likely fine tuned), we define a short function which just traverse the node list to carry out the replacements. The table `loc_to_scr` gets the locale from a script range (note the locale is the key, and that there is an intermediate table built on the fly for optimization). This locale is then used to get the `\language` and the `\localeid` as stored in `locale_props`, as well as the font (as requested). In the latter table a key starting with `/` maps the font from the global one (the key) to the local one (the value). Maths are skipped and discretionaries are handled in a special way.

```

5290 % TODO - to a lua file
5291 \directlua{
5292 Babel.script_blocks = {
5293   ['dflt'] = {},
5294   ['Arab'] = {{0x0600, 0x06FF}, {0x08A0, 0x08FF}, {0x0750, 0x077F},
5295               {0xFE70, 0xFEFF}, {0xFB50, 0xFDFF}, {0x1EE00, 0x1EEFF}},
5296   ['Armn'] = {{0x0530, 0x058F}},
5297   ['Beng'] = {{0x0980, 0x09FF}},
5298   ['Cher'] = {{0x13A0, 0x13FF}, {0xAB70, 0xABBF}},
5299   ['Copt'] = {{0x03E2, 0x03EF}, {0x2C80, 0x2CFF}, {0x102E0, 0x102FF}},
5300   ['Cyr1'] = {{0x0400, 0x04FF}, {0x0500, 0x052F}, {0x1C80, 0x1C8F},
5301               {0x2DE0, 0x2DFF}, {0xA640, 0xA69F}},
5302   ['Deva'] = {{0x0900, 0x097F}, {0xA8E0, 0xA8FF}},
5303   ['Ethi'] = {{0x1200, 0x137F}, {0x1380, 0x139F}, {0x2D80, 0x2DDF},
5304               {0xAB00, 0xAB2F}},
5305   ['Geor'] = {{0x10A0, 0x10FF}, {0x2D00, 0x2D2F}},
5306   % Don't follow strictly Unicode, which places some Coptic letters in

```

```

5307 % the 'Greek and Coptic' block
5308 ['Grek'] = {{0x0370, 0x03E1}, {0x03F0, 0x03FF}, {0x1F00, 0x1FFF}},
5309 ['Hans'] = {{0x2E80, 0x2EFF}, {0x3000, 0x303F}, {0x31C0, 0x31EF},
5310             {0x3300, 0x33FF}, {0x3400, 0x4DBF}, {0x4E00, 0x9FFF},
5311             {0xF900, 0xFAFF}, {0xFE30, 0xFE4F}, {0xFF00, 0xFFEF},
5312             {0x20000, 0x2A6DF}, {0x2A700, 0x2B73F},
5313             {0x2B740, 0x2B81F}, {0x2B820, 0x2CEAF},
5314             {0x2CEB0, 0x2EBEF}, {0x2F800, 0x2FA1F}},
5315 ['Hebr'] = {{0x0590, 0x05FF}},
5316 ['Jpan'] = {{0x3000, 0x303F}, {0x3040, 0x309F}, {0x30A0, 0x30FF},
5317             {0x4E00, 0x9FAF}, {0xFF00, 0xFFEF}},
5318 ['Khmr'] = {{0x1780, 0x17FF}, {0x19E0, 0x19FF}},
5319 ['Knda'] = {{0x0C80, 0x0CFF}},
5320 ['Kore'] = {{0x1100, 0x11FF}, {0x3000, 0x303F}, {0x3130, 0x318F},
5321             {0x4E00, 0x9FAF}, {0xA960, 0xA97F}, {0xAC00, 0xD7AF},
5322             {0xD7B0, 0xD7FF}, {0xFF00, 0xFFEF}},
5323 ['Lao'] = {{0x0E80, 0x0EFF}},
5324 ['Latn'] = {{0x0000, 0x007F}, {0x0080, 0x00FF}, {0x0100, 0x017F},
5325             {0x0180, 0x024F}, {0x1E00, 0x1EFF}, {0x2C60, 0x2C7F},
5326             {0xA720, 0xA7FF}, {0xAB30, 0xAB6F}},
5327 ['Mahj'] = {{0x11150, 0x1117F}},
5328 ['Mlym'] = {{0x0D00, 0x0D7F}},
5329 ['Mymr'] = {{0x1000, 0x109F}, {0xAA60, 0xAA7F}, {0xA9E0, 0xA9FF}},
5330 ['Orya'] = {{0x0B00, 0x0B7F}},
5331 ['Sinh'] = {{0x0D80, 0x0DFF}, {0x111E0, 0x111FF}},
5332 ['Syrc'] = {{0x0700, 0x074F}, {0x0860, 0x086F}},
5333 ['Taml'] = {{0x0B80, 0x0BFF}},
5334 ['Telu'] = {{0x0C00, 0x0C7F}},
5335 ['Tfng'] = {{0x2D30, 0x2D7F}},
5336 ['Thai'] = {{0x0E00, 0x0E7F}},
5337 ['Tibt'] = {{0x0F00, 0x0FFF}},
5338 ['Vaii'] = {{0xA500, 0xA63F}},
5339 ['Yiii'] = {{0xA000, 0xA48F}, {0xA490, 0xA4CF}}
5340 }
5341
5342 Babel.script_blocks.Cyrs = Babel.script_blocks.Cyrl
5343 Babel.script_blocks.Hant = Babel.script_blocks.Hans
5344 Babel.script_blocks.Kana = Babel.script_blocks.Jpan
5345
5346 function Babel.locale_map(head)
5347   if not Babel.locale_mapped then return head end
5348
5349   local LOCALE = luatexbase.registernumber'bbl@attr@locale'
5350   local GLYPH = node.id('glyph')
5351   local inmath = false
5352   local toloc_save
5353   for item in node.traverse(head) do
5354     local toloc
5355     if not inmath and item.id == GLYPH then
5356       % Optimization: build a table with the chars found
5357       if Babel.chr_to_loc[item.char] then
5358         toloc = Babel.chr_to_loc[item.char]
5359       else
5360         for lc, maps in pairs(Babel.loc_to_scr) do
5361           for _, rg in pairs(maps) do
5362             if item.char >= rg[1] and item.char <= rg[2] then
5363               Babel.chr_to_loc[item.char] = lc
5364               toloc = lc
5365               break

```

```

5366         end
5367     end
5368 end
5369 end
5370 % Now, take action, but treat composite chars in a different
5371 % fashion, because they 'inherit' the previous locale. Not yet
5372 % optimized.
5373 if not toloc and
5374     (item.char >= 0x0300 and item.char <= 0x036F) or
5375     (item.char >= 0x1AB0 and item.char <= 0x1AFF) or
5376     (item.char >= 0x1DC0 and item.char <= 0x1DFF) then
5377     toloc = toloc_save
5378 end
5379 if toloc and toloc > -1 then
5380     if Babel.locale_props[toloc].lg then
5381         item.lang = Babel.locale_props[toloc].lg
5382         node.set_attribute(item, LOCALE, toloc)
5383     end
5384     if Babel.locale_props[toloc]['/'..item.font] then
5385         item.font = Babel.locale_props[toloc]['/'..item.font]
5386     end
5387     toloc_save = toloc
5388 end
5389 elseif not inmath and item.id == 7 then
5390     item.replace = item.replace and Babel.locale_map(item.replace)
5391     item.pre      = item.pre and Babel.locale_map(item.pre)
5392     item.post     = item.post and Babel.locale_map(item.post)
5393 elseif item.id == node.id'math' then
5394     inmath = (item.subtype == 0)
5395 end
5396 end
5397 return head
5398 end
5399 }

```

The code for `\babelcharproperty` is straightforward. Just note the modified lua table can be different.

```

5400 \newcommand\babelcharproperty[1]{%
5401   \count@=#1\relax
5402   \ifvmode
5403     \expandafter\bbl@chprop
5404   \else
5405     \bbl@error{\string\babelcharproperty\space can be used only in\\%
5406               vertical mode (preamble or between paragraphs)}%
5407     {See the manual for futher info}%
5408   \fi}
5409 \newcommand\bbl@chprop[3][\the\count@]{%
5410   \@tempcnta=#1\relax
5411   \bbl@ifunset{\bbl@chprop@#2}%
5412   {\bbl@error{No property named '#2'. Allowed values are\\%
5413             direction (bc), mirror (bmg), and linebreak (lb)}%
5414     {See the manual for futher info}}%
5415   }%
5416   \loop
5417     \bbl@cs{\chprop@#2}{#3}%
5418   \ifnum\count@<\@tempcnta
5419     \advance\count@\@ne
5420   \repeat}
5421 \def\bbl@chprop@direction#1{%

```

```

5422 \directlua{
5423   Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
5424   Babel.characters[\the\count@]['d'] = '#1'
5425 }}
5426 \let\bbl@chprop@bc\bbl@chprop@direction
5427 \def\bbl@chprop@mirror#1{%
5428   \directlua{
5429     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
5430     Babel.characters[\the\count@]['m'] = '\number#1'
5431   }}
5432 \let\bbl@chprop@bmg\bbl@chprop@mirror
5433 \def\bbl@chprop@linebreak#1{%
5434   \directlua{
5435     Babel.cjk_characters[\the\count@] = Babel.cjk_characters[\the\count@] or {}
5436     Babel.cjk_characters[\the\count@]['c'] = '#1'
5437   }}
5438 \let\bbl@chprop@lb\bbl@chprop@linebreak
5439 \def\bbl@chprop@locale#1{%
5440   \directlua{
5441     Babel.chr_to_loc = Babel.chr_to_loc or {}
5442     Babel.chr_to_loc[\the\count@] =
5443       \bbl@ifblank{#1}{-1000}{\the\bbl@cs{id@#1}}\space
5444   }}

```

Post-handling hyphenation patterns for non-standard rules, like ff to ff-f. There are still some issues with speed (not very slow, but still slow).

After declaring the table containing the patterns with their replacements, we define some auxiliary functions: `str_to_nodes` converts the string returned by a function to a node list, taking the node at base as a model (font, language, etc.); `fetch_word` fetches a series of glyphs and discretionaries, which pattern is matched against (if there is a match, it is called again before trying other patterns, and this is very likely the main bottleneck).

`post_hyphenate_replace` is the callback applied after `lang.hyphenate`. This means the automatic hyphenation points are known. As empty captures return a byte position (as explained in the `luatex` manual), we must convert it to a utf8 position. With `first`, the last byte can be the leading byte in a utf8 sequence, so we just remove it and add 1 to the resulting length. With `last` we must take into account the capture position points to the next character. Here `word_head` points to the starting node of the text to be matched.

```

5445 \begingroup % TODO - to a lua file
5446 \catcode`\~ = 12
5447 \catcode`\# = 12
5448 \catcode`\% = 12
5449 \catcode`\& = 14
5450 \directlua{
5451   Babel.linebreaking.replacements = {}
5452   Babel.linebreaking.replacements[0] = {} && pre
5453   Babel.linebreaking.replacements[1] = {} && post
5454
5455   && Discretionaries contain strings as nodes
5456   function Babel.str_to_nodes(fn, matches, base)
5457     local n, head, last
5458     if fn == nil then return nil end
5459     for s in string.utfvalues(fn(matches)) do
5460       if base.id == 7 then
5461         base = base.replace
5462       end
5463       n = node.copy(base)
5464       n.char = s
5465       if not head then
5466         head = n

```

```

5467     else
5468         last.next = n
5469     end
5470     last = n
5471 end
5472 return head
5473 end
5474
5475 Babel.fetch_subtext = {}
5476
5477 %% Merging both functions doesn't seem feasible, because there are too
5478 %% many differences.
5479 Babel.fetch_subtext[0] = function(head)
5480     local word_string = ''
5481     local word_nodes = {}
5482     local lang
5483     local item = head
5484     local inmath = false
5485
5486     while item do
5487
5488         if item.id == 11 then
5489             inmath = (item.subtype == 0)
5490         end
5491
5492         if inmath then
5493             %% pass
5494
5495         elseif item.id == 29 then
5496             local locale = node.get_attribute(item, Babel.attr_locale)
5497
5498             if lang == locale or lang == nil then
5499                 if (item.char ~= 124) then %% ie, not | = space
5500                     lang = lang or locale
5501                     word_string = word_string .. unicode.utf8.char(item.char)
5502                     word_nodes[#word_nodes+1] = item
5503                 end
5504             else
5505                 break
5506             end
5507
5508         elseif item.id == 12 and item.subtype == 13 then
5509             word_string = word_string .. '|'
5510             word_nodes[#word_nodes+1] = item
5511
5512             %% Ignore leading unrecognized nodes, too.
5513         elseif word_string ~= '' then
5514             word_string = word_string .. Babel.us_char
5515             word_nodes[#word_nodes+1] = item %% Will be ignored
5516         end
5517
5518         item = item.next
5519     end
5520
5521     %% Here and above we remove some trailing chars but not the
5522     %% corresponding nodes. But they aren't accessed.
5523     if word_string:sub(-1) == '|' then
5524         word_string = word_string:sub(1,-2)
5525     end

```

```

5526     word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
5527     return word_string, word_nodes, item, lang
5528 end
5529
5530 Babel.fetch_subtext[1] = function(head)
5531     local word_string = ''
5532     local word_nodes = {}
5533     local lang
5534     local item = head
5535     local inmath = false
5536
5537     while item do
5538
5539         if item.id == 11 then
5540             inmath = (item.subtype == 0)
5541             end
5542
5543         if inmath then
5544             &% pass
5545
5546         elseif item.id == 29 then
5547             if item.lang == lang or lang == nil then
5548                 if (item.char ~= 124) and (item.char ~= 61) then &% not =, not |
5549                     lang = lang or item.lang
5550                     word_string = word_string .. unicode.utf8.char(item.char)
5551                     word_nodes[#word_nodes+1] = item
5552                 end
5553             else
5554                 break
5555             end
5556
5557         elseif item.id == 7 and item.subtype == 2 then
5558             word_string = word_string .. '='
5559             word_nodes[#word_nodes+1] = item
5560
5561         elseif item.id == 7 and item.subtype == 3 then
5562             word_string = word_string .. '|'
5563             word_nodes[#word_nodes+1] = item
5564
5565             &% (1) Go to next word if nothing was found, and (2) implicitly
5566             &% remove leading USs.
5567             elseif word_string == '' then
5568                 &% pass
5569
5570             &% This is the responsible for splitting by words.
5571             elseif (item.id == 12 and item.subtype == 13) then
5572                 break
5573
5574             else
5575                 word_string = word_string .. Babel.us_char
5576                 word_nodes[#word_nodes+1] = item &% Will be ignored
5577             end
5578
5579             item = item.next
5580         end
5581
5582     word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
5583     return word_string, word_nodes, item, lang
5584 end

```



```

5585
5586 function Babel.pre_hyphenate_replace(head)
5587     Babel.hyphenate_replace(head, 0)
5588 end
5589
5590 function Babel.post_hyphenate_replace(head)
5591     Babel.hyphenate_replace(head, 1)
5592 end
5593
5594 function Babel.debug_hyph(w, wn, sc, first, last, last_match)
5595     local ss = ''
5596     for pp = 1, 40 do
5597         if wn[pp] then
5598             if wn[pp].id == 29 then
5599                 ss = ss .. unicode.utf8.char(wn[pp].char)
5600             else
5601                 ss = ss .. '{' .. wn[pp].id .. '}'
5602             end
5603         end
5604     end
5605     print('nod', ss)
5606     print('lst_m',
5607         string.rep(' ', unicode.utf8.len(
5608             string.sub(w, 1, last_match))-1) .. '>')
5609     print('str', w)
5610     print('sc', string.rep(' ', sc-1) .. '^')
5611     if first == last then
5612         print('f=l', string.rep(' ', first-1) .. '!')
5613     else
5614         print('f/l', string.rep(' ', first-1) .. '[' ..
5615             string.rep(' ', last-first-1) .. ']')
5616     end
5617 end
5618
5619 Babel.us_char = string.char(31)
5620
5621 function Babel.hyphenate_replace(head, mode)
5622     local u = unicode.utf8
5623     local lbkr = Babel.linebreaking.replacements[mode]
5624
5625     local word_head = head
5626
5627     while true do    %% for each subtext block
5628
5629         local w, w_nodes, nw, lang = Babel.fetch_subtext[mode](word_head)
5630
5631         if Babel.debug then
5632             print()
5633             print((mode == 0) and '@@@<' or '@@@>', w)
5634         end
5635
5636         if nw == nil and w == '' then break end
5637
5638         if not lang then goto next end
5639         if not lbkr[lang] then goto next end
5640
5641         %% For each saved (pre|post)hyphenation. TODO. Reconsider how
5642         %% loops are nested.
5643         for k=1, #lbkr[lang] do

```

```

5644     local p = lbkr[lang][k].pattern
5645     local r = lbkr[lang][k].replace
5646
5647     if Babel.debug then
5648         print('*****', p, mode)
5649     end
5650
5651     %% This variable is set in some cases below to the first *byte*
5652     %% after the match, either as found by u.match (faster) or the
5653     %% computed position based on sc if w has changed.
5654     local last_match = 0
5655
5656     %% For every match.
5657     while true do
5658         if Babel.debug then
5659             print('====')
5660         end
5661         local new  %% used when inserting and removing nodes
5662         local refetch = false
5663
5664         local matches = { u.match(w, p, last_match) }
5665         if #matches < 2 then break end
5666
5667         %% Get and remove empty captures (with ())'s, which return a
5668         %% number with the position), and keep actual captures
5669         %% (from (...)), if any, in matches.
5670         local first = table.remove(matches, 1)
5671         local last  = table.remove(matches, #matches)
5672         %% Non re-fetched substrings may contain \31, which separates
5673         %% substrings.
5674         if string.find(w:sub(first, last-1), Babel.us_char) then break end
5675
5676         local save_last = last  %% with A()BC()D, points to D
5677
5678         %% Fix offsets, from bytes to unicode. Explained above.
5679         first = u.len(w:sub(1, first-1)) + 1
5680         last  = u.len(w:sub(1, last-1))  %% now last points to C
5681
5682         %% This loop stores in n small table the nodes
5683         %% corresponding to the pattern. Used by 'data' to provide a
5684         %% predictable behavior with 'insert' (now w_nodes is modified on
5685         %% the fly), and also access to 'remove'd nodes.
5686         local sc = first-1          %% Used below, too
5687         local data_nodes = {}
5688
5689         for q = 1, last-first+1 do
5690             data_nodes[q] = w_nodes[sc+q]
5691         end
5692
5693         %% This loop traverses the matched substring and takes the
5694         %% corresponding action stored in the replacement list.
5695         %% sc = the position in substr nodes / string
5696         %% rc = the replacement table index
5697         local rc = 0
5698
5699         while rc < last-first+1 do %% for each replacement
5700             if Babel.debug then
5701                 print('.....', rc + 1)
5702             end

```

```

5703         sc = sc + 1
5704         rc = rc + 1
5705
5706         if Babel.debug then
5707             Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
5708         end
5709
5710         local crep = r[rc]
5711         local item = w_nodes[sc]
5712         local item_base = item
5713         local placeholder = Babel.us_char
5714         local d
5715
5716         if crep and crep.data then
5717             item_base = data_nodes[crep.data]
5718         end
5719
5720         if crep and next(crep) == nil then &% = {}
5721             last_match = save_last    &% Optimization
5722             goto next
5723
5724         elseif crep == nil then &% = remove
5725             node.remove(head, item)
5726             table.remove(w_nodes, sc)
5727             w = u.sub(w, 1, sc-1) .. u.sub(w, sc+1)
5728             sc = sc - 1    &% Nothing has been inserted.
5729             last_match = utf8.offset(w, sc+1)
5730             goto next
5731
5732         elseif crep and crep.string then
5733             local str = crep.string(matches)
5734             if str == '' then    &% Gather with nil
5735                 node.remove(head, item)
5736                 table.remove(w_nodes, sc)
5737                 w = u.sub(w, 1, sc-1) .. u.sub(w, sc+1)
5738                 sc = sc - 1    &% Nothing has been inserted.
5739             else
5740                 local loop_first = true
5741                 for s in string.utfvalues(str) do
5742                     d = node.copy(item_base)
5743                     d.char = s
5744                     if loop_first then
5745                         loop_first = false
5746                         head, new = node.insert_before(head, item, d)
5747                         if sc == 1 then
5748                             word_head = head
5749                         end
5750                         w_nodes[sc] = d
5751                         w = u.sub(w, 1, sc-1) .. u.char(s) .. u.sub(w, sc+1)
5752                     else
5753                         sc = sc + 1
5754                         head, new = node.insert_before(head, item, d)
5755                         table.insert(w_nodes, sc, new)
5756                         w = u.sub(w, 1, sc-1) .. u.char(s) .. u.sub(w, sc)
5757                     end
5758                 end
5759                 if Babel.debug then
5760                     print('.....', 'str')
5761                     Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
5762                 end

```

```

5762         end %% for
5763         node.remove(head, item)
5764     end %% if ''
5765     last_match = utf8.offset(w, sc+1)
5766     goto next
5767
5768 elseif mode == 1 and crep and (crep.pre or crep.no or crep.post) then
5769     d = node.new(7, 0)    %% (disc, discretionary)
5770     d.pre    = Babel.str_to_nodes(crep.pre, matches, item_base)
5771     d.post   = Babel.str_to_nodes(crep.post, matches, item_base)
5772     d.replace = Babel.str_to_nodes(crep.no, matches, item_base)
5773     d.attr = item_base.attr
5774     if crep.pre == nil then %% TeXbook p96
5775         d.penalty = crep.penalty or tex.hyphenpenalty
5776     else
5777         d.penalty = crep.penalty or tex.exhyphenpenalty
5778     end
5779     placeholder = '|'
5780     head, new = node.insert_before(head, item, d)
5781
5782 elseif mode == 0 and crep and (crep.pre or crep.no or crep.post) then
5783     %% ERROR
5784
5785 elseif crep and crep.penalty then
5786     d = node.new(14, 0)    %% (penalty, userpenalty)
5787     d.attr = item_base.attr
5788     d.penalty = crep.penalty
5789     head, new = node.insert_before(head, item, d)
5790
5791 elseif crep and crep.space then
5792     %% 655360 = 10 pt = 10 * 65536 sp
5793     d = node.new(12, 13)    %% (glue, spaceskip)
5794     local quad = font.getfont(item_base.font).size or 655360
5795     node.setglue(d, crep.space[1] * quad,
5796                  crep.space[2] * quad,
5797                  crep.space[3] * quad)
5798     if mode == 0 then
5799         placeholder = '|'
5800     end
5801     head, new = node.insert_before(head, item, d)
5802
5803 elseif crep and crep.spacefactor then
5804     d = node.new(12, 13)    %% (glue, spaceskip)
5805     local base_font = font.getfont(item_base.font)
5806     node.setglue(d,
5807                  crep.spacefactor[1] * base_font.parameters['space'],
5808                  crep.spacefactor[2] * base_font.parameters['space_stretch'],
5809                  crep.spacefactor[3] * base_font.parameters['space_shrink'])
5810     if mode == 0 then
5811         placeholder = '|'
5812     end
5813     head, new = node.insert_before(head, item, d)
5814
5815 elseif mode == 0 and crep and crep.space then
5816     %% ERROR
5817
5818 end %% ie replacement cases
5819
5820 %% Shared by disc, space and penalty.

```

```

5821         if sc == 1 then
5822             word_head = head
5823         end
5824         if crep.insert then
5825             w = u.sub(w, 1, sc-1) .. placeholder .. u.sub(w, sc)
5826             table.insert(w_nodes, sc, new)
5827             last = last + 1
5828         else
5829             w_nodes[sc] = d
5830             node.remove(head, item)
5831             w = u.sub(w, 1, sc-1) .. placeholder .. u.sub(w, sc+1)
5832         end
5833
5834         last_match = utf8.offset(w, sc+1)
5835
5836         ::next::
5837
5838     end    %% for each replacement
5839
5840     if Babel.debug then
5841         print('.....', '/')
5842         Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
5843     end
5844
5845     end    %% for match
5846
5847     end    %% for patterns
5848
5849     ::next::
5850     word_head = nw
5851 end    %% for substring
5852 return head
5853 end
5854
5855 %% This table stores capture maps, numbered consecutively
5856 Babel.capture_maps = {}
5857
5858 %% The following functions belong to the next macro
5859 function Babel.capture_func(key, cap)
5860     local ret = "[" .. cap:gsub('{{[0-9]}}', "]]..m[%1]..[(") .. "]"
5861     ret = ret:gsub('{{[0-9]}}([^[^]+)|(.-)}', Babel.capture_func_map)
5862     ret = ret:gsub("%[%[%]%%%.%", '')
5863     ret = ret:gsub("%.%[%[%]%%%", '')
5864     return key .. [[=function(m) return ]] .. ret .. [[ end]]
5865 end
5866
5867 function Babel.capt_map(from, mapno)
5868     return Babel.capture_maps[mapno][from] or from
5869 end
5870
5871 %% Handle the {n|abc|ABC} syntax in captures
5872 function Babel.capture_func_map(capno, from, to)
5873     local froms = {}
5874     for s in string.utfcharacters(from) do
5875         table.insert(froms, s)
5876     end
5877     local cnt = 1
5878     table.insert(Babel.capture_maps, {})
5879     local mlen = table.getn(Babel.capture_maps)

```

```

5880   for s in string.utfcharacters(to) do
5881     Babel.capture_maps[mlen][from[cnt]] = s
5882     cnt = cnt + 1
5883   end
5884   return "]]..Babel.capt_map(m[" .. capno .. "], " ..
5885     (mlen) .. ").. " .. "["
5886 end
5887 }

```

Now the  $\TeX$  high level interface, which requires the function defined above for converting strings to functions returning a string. These functions handle the  $\{n\}$  syntax. For example,  $\text{pre}=\{1\}\{1\}$  becomes  $\text{function}(m)$  `return m[1]..m[1]..'-' end`, where  $m$  are the matches returned after applying the pattern. With a mapped capture the functions are similar to  $\text{function}(m)$  `return Babel.capt_map(m[1],1)` end, where the last argument identifies the mapping to be applied to  $m[1]$ . The way it is carried out is somewhat tricky, but the effect is not dissimilar to `lua load - save the code as string in a TeX macro, and expand this macro at the appropriate place`. As `\directlua` does not take into account the current catcode of `@`, we just avoid this character in macro names (which explains the internal group, too).

```

5888 \catcode\# = 6
5889 \gdef\babelposthyphenation#1#2#3{&%
5890   \bbl@activateposthyphen
5891   \begingroup
5892     \def\babeltempa{\bbl@add@list\babeltempb}&%
5893     \let\babeltempb\@empty
5894     \def\bbl@tempa{#3}&% TODO. Ugly trick to preserve {}:
5895     \bbl@replace\bbl@tempa{,}{ ,}&%
5896     \expandafter\bbl@foreach\expandafter{\bbl@tempa}{&%
5897       \bbl@ifsamestring{##1}{remove}&%
5898       {\bbl@add@list\babeltempb{nil}}&%
5899       {\directlua{
5900         local rep = {[##1]}
5901         rep = rep:gsub('^%s*(insert)%s*', 'insert = true, ')
5902         rep = rep:gsub('^(no)%s*=%s*([^\s,]*)', Babel.capture_func)
5903         rep = rep:gsub('^(pre)%s*=%s*([^\s,]*)', Babel.capture_func)
5904         rep = rep:gsub('^(post)%s*=%s*([^\s,]*)', Babel.capture_func)
5905         rep = rep:gsub('(string)%s*=%s*([^\s,]*)', Babel.capture_func)
5906         tex.print([[\\string\babeltempa{}}] .. rep .. [[}}]])
5907       }}&%
5908     \directlua{
5909       local lbkr = Babel.linebreaking.replacements[1]
5910       local u = unicode.utf8
5911       &% Convert pattern:
5912       local patt = string.gsub([=[#2]=], '%s', '')
5913       if not u.find(patt, '()', nil, true) then
5914         patt = '()' .. patt .. '()'
5915       end
5916       patt = string.gsub(patt, '%(%)^', '^()')
5917       patt = string.gsub(patt, '%$(%)', '()$')
5918       patt = u.gsub(patt, '{(.)}',
5919         function (n)
5920           return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
5921         end)
5922       lbkr[\\the\\csname l@#1\\endcsname] = lbkr[\\the\\csname l@#1\\endcsname] or {}
5923       table.insert(lbkr[\\the\\csname l@#1\\endcsname],
5924         { pattern = patt, replace = { \babeltempb } })
5925     }&%
5926   \endgroup}
5927 % TODO. Copy paste pattern.
5928 \gdef\babelprehyphenation#1#2#3{&%

```

```

5929 \bbl@activateprehyphen
5930 \begingroup
5931 \def\babeltempa{\bbl@add@list\babeltempb}&%
5932 \let\babeltempb\@empty
5933 \def\bbl@tempa{#3}&% TODO. Ugly trick to preserve {}:
5934 \bbl@replace\bbl@tempa{,}{ ,}&%
5935 \expandafter\bbl@foreach\expandafter{\bbl@tempa}{&%
5936 \bbl@ifsamestring{##1}{remove}&%
5937 {\bbl@add@list\babeltempb{nil}}&%
5938 {\directlua{
5939 local rep = [[##1]]
5940 rep = rep:gsub('^s*(insert)s*',',', 'insert = true, ')
5941 rep = rep:gsub('(string)s*=%s*([^\s,]*)', Babel.capture_func)
5942 rep = rep:gsub(' (space)s*=%s*([%d%.]+)%s+([%d%.]+)%s+([%d%.]+)',
5943 'space = {' .. '%2, %3, %4' .. '}')
5944 rep = rep:gsub(' (spacefactor)s*=%s*([%d%.]+)%s+([%d%.]+)%s+([%d%.]+)',
5945 'spacefactor = {' .. '%2, %3, %4' .. '}')
5946 tex.print([[string\babeltempa{}} .. rep .. [{}]])
5947 }}&%
5948 \directlua{
5949 local lbkr = Babel.linebreaking.replacements[0]
5950 local u = unicode.utf8
5951 &% Convert pattern:
5952 local patt = string.gsub(==[#2]==, '%s', '')
5953 if not u.find(patt, '()', nil, true) then
5954 patt = '()' .. patt .. '()'
5955 end
5956 &% patt = string.gsub(patt, '%(%)^', '^()')
5957 &% patt = string.gsub(patt, '([^\%])%$%(%)', '%1()$')
5958 patt = u.gsub(patt, '{(.)}',
5959 function (n)
5960 return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
5961 end)
5962 lbkr[\the\csname bbl@id@@#1\endcsname] = lbkr[\the\csname bbl@id@@#1\endcsname] or {}
5963 table.insert(lbkr[\the\csname bbl@id@@#1\endcsname],
5964 { pattern = patt, replace = { \babeltempb } })
5965 }&%
5966 \endgroup}
5967 \endgroup
5968 \def\bbl@activateposthyphen{%
5969 \let\bbl@activateposthyphen\relax
5970 \directlua{
5971 Babel.linebreaking.add_after(Babel.post_hyphenate_replace)
5972 }}
5973 \def\bbl@activateprehyphen{%
5974 \let\bbl@activateprehyphen\relax
5975 \directlua{
5976 Babel.linebreaking.add_before(Babel.pre_hyphenate_replace)
5977 }}

```

## 13.7 Layout

Unlike xetex, luatex requires only minimal changes for right-to-left layouts, particularly in monolingual documents (the engine itself reverses boxes – including column order or headings –, margins, etc.) with `bidi=basic`, without having to patch almost any macro where text direction is relevant.

`\@hangfrom` is useful in many contexts and it is redefined always with the layout option.

There are, however, a number of issues when the text direction is not the same as the box direction (as set by `\bodydir`), and when `\parbox` and `\hangindent` are involved. Fortunately, latest releases

of luatex simplify a lot the solution with \shapemode.

With the issue #15 I realized commands are best patched, instead of redefined. With a few lines, a modification could be applied to several classes and packages. Now, tabular seems to work (at least in simple cases) with array, tabularx, hhline, colortbl, longtable, booktabs, etc. However, dcolumn still fails.

```

5978 \bbl@trace{Redefinitions for bidi layout}
5979 \ifx\@eqnnum\undefined\else
5980   \ifx\bbl@attr@dir\undefined\else
5981     \edef\@eqnnum{%
5982       \unexpanded{\ifcase\bbl@attr@dir\else\bbl@textdir\@ne\fi}%
5983       \unexpanded\expandafter{\@eqnnum}}
5984   \fi
5985 \fi
5986 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
5987 \ifnum\bbl@bidimode>\z@
5988   \def\bbl@nextfake#1{% non-local changes, use always inside a group!
5989     \bbl@exp{%
5990       \mathdir\the\bodydir
5991       #1%           Once entered in math, set boxes to restore values
5992       \<ifmmode>%
5993         \everyvbox{%
5994           \the\everyvbox
5995           \bodydir\the\bodydir
5996           \mathdir\the\mathdir
5997           \everyhbox{\the\everyhbox}%
5998           \everyvbox{\the\everyvbox}}%
5999         \everyhbox{%
6000           \the\everyhbox
6001           \bodydir\the\bodydir
6002           \mathdir\the\mathdir
6003           \everyhbox{\the\everyhbox}%
6004           \everyvbox{\the\everyvbox}}%
6005       \<fi>}}%
6006   \def\@hangfrom#1{%
6007     \setbox\@tempboxa\hbox{#1}%
6008     \hangindent\wd\@tempboxa
6009     \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
6010       \shapemode\@ne
6011     \fi
6012     \noindent\box\@tempboxa}
6013 \fi
6014 \IfBabelLayout{tabular}
6015   {\let\bbl@OL@tabular\@tabular
6016     \bbl@replace\@tabular{$}\bbl@nextfake$}%
6017   \let\bbl@NL@tabular\@tabular
6018   \AtBeginDocument{%
6019     \ifx\bbl@NL@tabular\@tabular\else
6020       \bbl@replace\@tabular{$}\bbl@nextfake$}%
6021     \let\bbl@NL@tabular\@tabular
6022     \fi}}
6023 {}
6024 \IfBabelLayout{lists}
6025   {\let\bbl@OL@list\list
6026     \bbl@sreplace\list{\parshape}\bbl@listparshape}%
6027   \let\bbl@NL@list\list
6028   \def\bbl@listparshape#1#2#3{%
6029     \parshape #1 #2 #3 %
6030     \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else

```



```

6031     \shapemode\tw@
6032     \fi}}
6033 {}
6034 \IfBabelLayout{graphics}
6035 {\let\bbl@pictresetdir\relax
6036  \def\bbl@pictsetdir#1{%
6037   \ifcase\bbl@thetextdir
6038   \let\bbl@pictresetdir\relax
6039   \else
6040   \ifcase#1\bodydir TLT % Remember this sets the inner boxes
6041   \or\textdir TLT
6042   \else\bodydir TLT \textdir TLT
6043   \fi
6044   % \ (text|par)dir required in pgf:
6045   \def\bbl@pictresetdir{\bodydir TRT\pardir TRT\textdir TRT\relax}%
6046   \fi}%
6047 \ifx\AddToHook\@undefined\else
6048   \AddToHook{env/picture/begin}{\bbl@pictsetdir\tw@}%
6049   \directlua{
6050     Babel.get_picture_dir = true
6051     Babel.picture_has_bidi = 0
6052     function Babel.picture_dir (head)
6053       if not Babel.get_picture_dir then return head end
6054       for item in node.traverse(head) do
6055         if item.id == node.id'glyph' then
6056           local itemchar = item.char
6057           % TODO. Copy paste pattern from Babel.bidi (-r)
6058           local chardata = Babel.characters[itemchar]
6059           local dir = chardata and chardata.d or nil
6060           if not dir then
6061             for nn, et in ipairs(Babel.ranges) do
6062               if itemchar < et[1] then
6063                 break
6064               elseif itemchar <= et[2] then
6065                 dir = et[3]
6066                 break
6067             end
6068           end
6069           if dir and (dir == 'al' or dir == 'r') then
6070             Babel.picture_has_bidi = 1
6071           end
6072         end
6073       end
6074       return head
6075     end
6076     luatexbase.add_to_callback("hpack_filter", Babel.picture_dir,
6077       "Babel.picture_dir")
6078   }%
6079 \AtBeginDocument{%
6080   \long\def\put(#1,#2)#3{%
6081     \@killglue
6082     % Try:
6083     \ifx\bbl@pictresetdir\relax
6084       \def\bbl@tempc{0}%
6085     \else
6086       \directlua{
6087         Babel.get_picture_dir = true
6088         Babel.picture_has_bidi = 0
6089

```

```

6090      }%
6091      \setbox\z@\hb@xt@\z@{%
6092        \@defaultunitsset\@tempdimc{#1}\unitlength
6093        \kern\@tempdimc
6094        #3\hss}%
6095      \edef\bbl@tempc{\directlua{tex.print(Babel.picture_has_bidi)}}%
6096    \fi
6097    % Do:
6098    \@defaultunitsset\@tempdimc{#2}\unitlength
6099    \raise\@tempdimc\hb@xt@\z@{%
6100      \@defaultunitsset\@tempdimc{#1}\unitlength
6101      \kern\@tempdimc
6102      {\ifnum\bbl@tempc>\z@\bbl@pictresetdir\fi#3}\hss}%
6103    \ignorespaces}%
6104    \MakeRobust\put}%
6105  \fi
6106  \AtBeginDocument
6107    {\ifx\tikz@atbegin@node\@undefined\else
6108      \ifx\AddToHook\@undefined\else % TODO. Still tentative.
6109        \AddToHook{env/pgfpicture/begin}{\bbl@pictsetdir\@ne}%
6110        \bbl@add\pgfinterruptpicture{\bbl@pictresetdir}%
6111      \fi
6112      \let\bbl@OL@pgfpicture\pgfpicture
6113      \bbl@sreplace\pgfpicture{\pgfpicturetrue}%
6114      {\bbl@pictsetdir\z@\pgfpicturetrue}%
6115      \bbl@add\pgfsys@beginpicture{\bbl@pictsetdir\z@}%
6116      \bbl@add\tikz@atbegin@node{\bbl@pictresetdir}%
6117      \bbl@sreplace\tikz{\begingroup}%
6118      {\begingroup\bbl@pictsetdir\tw@}%
6119    \fi
6120    \ifx\AddToHook\@undefined\else
6121      \AddToHook{env/tcolorbox/begin}{\bbl@pictsetdir\@ne}%
6122    \fi
6123  }}
6124 {}

```

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`.

```

6125 \IfBabelLayout{counters}%
6126 {\let\bbl@OL@@textsuperscript\textsuperscript
6127  \bbl@sreplace\textsuperscript{\m@th}{\m@th\mathdir\pagedir}%
6128  \let\bbl@latinarabic=\@arabic
6129  \let\bbl@OL@@arabic\@arabic
6130  \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
6131  \@ifpackagewith{babel}{bidi=default}%
6132    {\let\bbl@asciroman=\@roman
6133     \let\bbl@OL@@roman\@roman
6134     \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
6135     \let\bbl@asciiRoman=\@Roman
6136     \let\bbl@OL@@roman\@Roman
6137     \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}%
6138     \let\bbl@OL@labelenumii\labelenumii
6139     \def\labelenumii{}\theenumii}%
6140     \let\bbl@OL@p@enumiii\p@enumiii
6141     \def\p@enumiii{\p@enumii}\theenumii{}}{}%
6142 <<Footnote changes>>
6143 \IfBabelLayout{footnotes}%
6144 {\let\bbl@OL@footnote\footnote

```

```

6145 \BabelFootnote\footnote\language\{}{}%
6146 \BabelFootnote\localfootnote\language\{}{}%
6147 \BabelFootnote\mainfootnote\{}{}{}%
6148 {}

```

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.

```

6149 \IfBabelLayout{extras}%
6150 {\let\bbl@OL@underline\underline
6151 \bbl@sreplace\underline{$\@@underline}{\bbl@nextfake$\@@underline}%
6152 \let\bbl@OL@LaTeX2e\LaTeX2e
6153 \DeclareRobustCommand{\LaTeXe}{\mbox{\m@th
6154 \if b\expandafter\@car\@series\@nil\boldmath\fi
6155 \babelsublr}%
6156 \LaTeX\kern.15em\bbl@nextfake$_{\textstyle\varepsilon}$}}%
6157 {}
6158 \end{luatex}

```

### 13.8 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.

```

6159 \begin{basic-r}
6160 \Babel = Babel or {}
6161

```

```

6162 Babel.bidi_enabled = true
6163
6164 require('babel-data-bidi.lua')
6165
6166 local characters = Babel.characters
6167 local ranges = Babel.ranges
6168
6169 local DIR = node.id("dir")
6170
6171 local function dir_mark(head, from, to, outer)
6172   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
6173   local d = node.new(DIR)
6174   d.dir = '+' .. dir
6175   node.insert_before(head, from, d)
6176   d = node.new(DIR)
6177   d.dir = '-' .. dir
6178   node.insert_after(head, to, d)
6179 end
6180
6181 function Babel.bidi(head, ispar)
6182   local first_n, last_n      -- first and last char with nums
6183   local last_es              -- an auxiliary 'last' used with nums
6184   local first_d, last_d      -- first and last char in L/R block
6185   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):

```

6186   local strong = ('TRT' == tex.pardir) and 'r' or 'l'
6187   local strong_lr = (strong == 'l') and 'l' or 'r'
6188   local outer = strong
6189
6190   local new_dir = false
6191   local first_dir = false
6192   local inmath = false
6193
6194   local last_lr
6195
6196   local type_n = ''
6197
6198   for item in node.traverse(head) do
6199
6200     -- three cases: glyph, dir, otherwise
6201     if item.id == node.id'glyph'
6202       or (item.id == 7 and item.subtype == 2) then
6203
6204       local itemchar
6205       if item.id == 7 and item.subtype == 2 then
6206         itemchar = item.replace.char
6207       else
6208         itemchar = item.char
6209       end
6210       local chardata = characters[itemchar]
6211       dir = chardata and chardata.d or nil
6212       if not dir then
6213         for nn, et in ipairs(ranges) do
6214           if itemchar < et[1] then
6215             break
6216           elseif itemchar <= et[2] then

```

```

6217         dir = et[3]
6218         break
6219     end
6220 end
6221 end
6222 dir = dir or 'l'
6223 if inmath then dir = ('TRT' == tex.mathdir) and 'r' or 'l' end

```

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).

```

6224     if new_dir then
6225         attr_dir = 0
6226         for at in node.traverse(item.attr) do
6227             if at.number == luatexbase.registernumber'bbl@attr@dir' then
6228                 attr_dir = at.value % 3
6229             end
6230         end
6231         if attr_dir == 1 then
6232             strong = 'r'
6233         elseif attr_dir == 2 then
6234             strong = 'al'
6235         else
6236             strong = 'l'
6237         end
6238         strong_lr = (strong == 'l') and 'l' or 'r'
6239         outer = strong_lr
6240         new_dir = false
6241     end
6242
6243     if dir == 'nsm' then dir = strong end -- W1

```

**Numbers.** The dual <al>/<r> system for R is somewhat cumbersome.

```

6244     dir_real = dir -- We need dir_real to set strong below
6245     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:

```

6246     if strong == 'al' then
6247         if dir == 'en' then dir = 'an' end -- W2
6248         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
6249         strong_lr = 'r' -- W3
6250     end

```

Once finished the basic setup for glyphs, consider the two other cases: dir node and the rest.

```

6251     elseif item.id == node.id'dir' and not inmath then
6252         new_dir = true
6253         dir = nil
6254     elseif item.id == node.id'math' then
6255         inmath = (item.subtype == 0)
6256     else
6257         dir = nil -- Not a char
6258     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>.

```

6259   if dir == 'en' or dir == 'an' or dir == 'et' then
6260       if dir ~= 'et' then
6261           type_n = dir
6262       end
6263       first_n = first_n or item
6264       last_n = last_es or item
6265       last_es = nil
6266   elseif dir == 'es' and last_n then -- W3+W6
6267       last_es = item
6268   elseif dir == 'cs' then             -- it's right - do nothing
6269   elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
6270       if strong_lr == 'r' and type_n ~= '' then
6271           dir_mark(head, first_n, last_n, 'r')
6272       elseif strong_lr == 'l' and first_d and type_n == 'an' then
6273           dir_mark(head, first_n, last_n, 'r')
6274           dir_mark(head, first_d, last_d, outer)
6275           first_d, last_d = nil, nil
6276       elseif strong_lr == 'l' and type_n ~= '' then
6277           last_d = last_n
6278       end
6279       type_n = ''
6280       first_n, last_n = nil, nil
6281   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:

```

6282   if dir == 'l' or dir == 'r' then
6283       if dir ~= outer then
6284           first_d = first_d or item
6285           last_d = item
6286       elseif first_d and dir ~= strong_lr then
6287           dir_mark(head, first_d, last_d, outer)
6288           first_d, last_d = nil, nil
6289       end
6290   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>, resptly, 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.

```

6291   if dir and not last_lr and dir ~= 'l' and outer == 'r' then
6292       item.char = characters[item.char] and
6293           characters[item.char].m or item.char
6294   elseif (dir or new_dir) and last_lr ~= item then
6295       local mir = outer .. strong_lr .. (dir or outer)
6296       if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
6297           for ch in node.traverse(node.next(last_lr)) do
6298               if ch == item then break end
6299               if ch.id == node.id'glyph' and characters[ch.char] then
6300                   ch.char = characters[ch.char].m or ch.char
6301               end
6302           end
6303       end
6304   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).

```

6305   if dir == 'l' or dir == 'r' then
6306       last_lr = item
6307       strong = dir_real          -- Don't search back - best save now
6308       strong_lr = (strong == 'l') and 'l' or 'r'
6309   elseif new_dir then
6310       last_lr = nil
6311   end
6312 end

```

Mirror the last chars if they are no directed. And make sure any open block is closed, too.

```

6313   if last_lr and outer == 'r' then
6314       for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
6315           if characters[ch.char] then
6316               ch.char = characters[ch.char].m or ch.char
6317           end
6318       end
6319   end
6320   if first_n then
6321       dir_mark(head, first_n, last_n, outer)
6322   end
6323   if first_d then
6324       dir_mark(head, first_d, last_d, outer)
6325   end

```

In boxes, the dir node could be added before the original head, so the actual head is the previous node.

```

6326   return node.prev(head) or head
6327 end
6328 </basic-r>

```

And here the Lua code for bidi=basic:

```

6329 <*basic>
6330 Babel = Babel or {}
6331
6332 -- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
6333
6334 Babel.fontmap = Babel.fontmap or {}
6335 Babel.fontmap[0] = {}          -- l
6336 Babel.fontmap[1] = {}          -- r
6337 Babel.fontmap[2] = {}          -- al/an
6338
6339 Babel.bidi_enabled = true
6340 Babel.mirroring_enabled = true
6341
6342 require('babel-data-bidi.lua')
6343
6344 local characters = Babel.characters
6345 local ranges = Babel.ranges
6346
6347 local DIR = node.id('dir')
6348 local GLYPH = node.id('glyph')
6349
6350 local function insert_implicit(head, state, outer)
6351     local new_state = state
6352     if state.sim and state.eim and state.sim ~= state.eim then
6353         dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
6354         local d = node.new(DIR)

```

```

6355     d.dir = '+' .. dir
6356     node.insert_before(head, state.sim, d)
6357     local d = node.new(DIR)
6358     d.dir = '-' .. dir
6359     node.insert_after(head, state.eim, d)
6360 end
6361 new_state.sim, new_state.eim = nil, nil
6362 return head, new_state
6363 end
6364
6365 local function insert_numeric(head, state)
6366     local new
6367     local new_state = state
6368     if state.san and state.ean and state.san ~= state.ean then
6369         local d = node.new(DIR)
6370         d.dir = '+TLT'
6371         _, new = node.insert_before(head, state.san, d)
6372         if state.san == state.sim then state.sim = new end
6373         local d = node.new(DIR)
6374         d.dir = '-TLT'
6375         _, new = node.insert_after(head, state.ean, d)
6376         if state.ean == state.eim then state.eim = new end
6377     end
6378     new_state.san, new_state.ean = nil, nil
6379     return head, new_state
6380 end
6381
6382 -- TODO - \hbox with an explicit dir can lead to wrong results
6383 -- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>. A small attempt
6384 -- was s made to improve the situation, but the problem is the 3-dir
6385 -- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
6386 -- well.
6387
6388 function Babel.bidi(head, ispar, hdir)
6389     local d -- d is used mainly for computations in a loop
6390     local prev_d = ''
6391     local new_d = false
6392
6393     local nodes = {}
6394     local outer_first = nil
6395     local inmath = false
6396
6397     local glue_d = nil
6398     local glue_i = nil
6399
6400     local has_en = false
6401     local first_et = nil
6402
6403     local ATDIR = luatexbase.registernumber'bbl@attr@dir'
6404
6405     local save_outer
6406     local temp = node.get_attribute(head, ATDIR)
6407     if temp then
6408         temp = temp % 3
6409         save_outer = (temp == 0 and 'l') or
6410                     (temp == 1 and 'r') or
6411                     (temp == 2 and 'al')
6412     elseif ispar then -- Or error? Shouldn't happen
6413         save_outer = ('TRT' == tex.pardir) and 'r' or 'l'

```



```

6414 else -- Or error? Shouldn't happen
6415     save_outer = ('TRT' == hdir) and 'r' or 'l'
6416 end
6417 -- when the callback is called, we are just _after_ the box,
6418 -- and the textdir is that of the surrounding text
6419 -- if not ispar and hdir ~= tex.textdir then
6420 --     save_outer = ('TRT' == hdir) and 'r' or 'l'
6421 -- end
6422 local outer = save_outer
6423 local last = outer
6424 -- 'al' is only taken into account in the first, current loop
6425 if save_outer == 'al' then save_outer = 'r' end
6426
6427 local fontmap = Babel.fontmap
6428
6429 for item in node.traverse(head) do
6430
6431     -- In what follows, #node is the last (previous) node, because the
6432     -- current one is not added until we start processing the neutrals.
6433
6434     -- three cases: glyph, dir, otherwise
6435     if item.id == GLYPH
6436         or (item.id == 7 and item.subtype == 2) then
6437
6438         local d_font = nil
6439         local item_r
6440         if item.id == 7 and item.subtype == 2 then
6441             item_r = item.replace -- automatic discs have just 1 glyph
6442         else
6443             item_r = item
6444         end
6445         local chardata = characters[item_r.char]
6446         d = chardata and chardata.d or nil
6447         if not d or d == 'nsm' then
6448             for nn, et in ipairs(ranges) do
6449                 if item_r.char < et[1] then
6450                     break
6451                 elseif item_r.char <= et[2] then
6452                     if not d then d = et[3]
6453                     elseif d == 'nsm' then d_font = et[3]
6454                     end
6455                     break
6456                 end
6457             end
6458         end
6459         d = d or 'l'
6460
6461         -- A short 'pause' in bidi for mapfont
6462         d_font = d_font or d
6463         d_font = (d_font == 'l' and 0) or
6464             (d_font == 'nsm' and 0) or
6465             (d_font == 'r' and 1) or
6466             (d_font == 'al' and 2) or
6467             (d_font == 'an' and 2) or nil
6468         if d_font and fontmap and fontmap[d_font][item_r.font] then
6469             item_r.font = fontmap[d_font][item_r.font]
6470         end
6471
6472         if new_d then

```

```

6473     table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
6474     if inmath then
6475         attr_d = 0
6476     else
6477         attr_d = node.get_attribute(item, ATDIR)
6478         attr_d = attr_d % 3
6479     end
6480     if attr_d == 1 then
6481         outer_first = 'r'
6482         last = 'r'
6483     elseif attr_d == 2 then
6484         outer_first = 'r'
6485         last = 'al'
6486     else
6487         outer_first = 'l'
6488         last = 'l'
6489     end
6490     outer = last
6491     has_en = false
6492     first_et = nil
6493     new_d = false
6494 end
6495
6496 if glue_d then
6497     if (d == 'l' and 'l' or 'r') ~= glue_d then
6498         table.insert(nodes, {glue_i, 'on', nil})
6499     end
6500     glue_d = nil
6501     glue_i = nil
6502 end
6503
6504 elseif item.id == DIR then
6505     d = nil
6506     new_d = true
6507
6508 elseif item.id == node.id'glue' and item.subtype == 13 then
6509     glue_d = d
6510     glue_i = item
6511     d = nil
6512
6513 elseif item.id == node.id'math' then
6514     inmath = (item.subtype == 0)
6515
6516 else
6517     d = nil
6518 end
6519
6520 -- AL <= EN/ET/ES      -- W2 + W3 + W6
6521 if last == 'al' and d == 'en' then
6522     d = 'an'            -- W3
6523 elseif last == 'al' and (d == 'et' or d == 'es') then
6524     d = 'on'            -- W6
6525 end
6526
6527 -- EN + CS/ES + EN      -- W4
6528 if d == 'en' and #nodes >= 2 then
6529     if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
6530         and nodes[#nodes-1][2] == 'en' then
6531         nodes[#nodes][2] = 'en'

```

```

6532     end
6533 end
6534
6535 -- AN + CS + AN          -- W4 too, because uax9 mixes both cases
6536 if d == 'an' and #nodes >= 2 then
6537   if (nodes[#nodes][2] == 'cs')
6538     and nodes[#nodes-1][2] == 'an' then
6539     nodes[#nodes][2] = 'an'
6540   end
6541 end
6542
6543 -- ET/EN                -- W5 + W7->l / W6->on
6544 if d == 'et' then
6545   first_et = first_et or (#nodes + 1)
6546 elseif d == 'en' then
6547   has_en = true
6548   first_et = first_et or (#nodes + 1)
6549 elseif first_et then      -- d may be nil here !
6550   if has_en then
6551     if last == 'l' then
6552       temp = 'l'      -- W7
6553     else
6554       temp = 'en'     -- W5
6555     end
6556   else
6557     temp = 'on'       -- W6
6558   end
6559   for e = first_et, #nodes do
6560     if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
6561   end
6562   first_et = nil
6563   has_en = false
6564 end
6565
6566 -- Force mathdir in math if ON (currently works as expected only
6567 -- with 'l')
6568 if inmath and d == 'on' then
6569   d = ('TRT' == tex.mathdir) and 'r' or 'l'
6570 end
6571
6572 if d then
6573   if d == 'al' then
6574     d = 'r'
6575     last = 'al'
6576   elseif d == 'l' or d == 'r' then
6577     last = d
6578   end
6579   prev_d = d
6580   table.insert(nodes, {item, d, outer_first})
6581 end
6582
6583   outer_first = nil
6584
6585 end
6586
6587 -- TODO -- repeated here in case EN/ET is the last node. Find a
6588 -- better way of doing things:
6589 if first_et then      -- dir may be nil here !
6590   if has_en then

```

```

6591     if last == 'l' then
6592         temp = 'l'      -- W7
6593     else
6594         temp = 'en'     -- W5
6595     end
6596 else
6597     temp = 'on'         -- W6
6598 end
6599 for e = first_et, #nodes do
6600     if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
6601 end
6602 end
6603
6604 -- dummy node, to close things
6605 table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
6606
6607 ----- NEUTRAL -----
6608
6609 outer = save_outer
6610 last = outer
6611
6612 local first_on = nil
6613
6614 for q = 1, #nodes do
6615     local item
6616
6617     local outer_first = nodes[q][3]
6618     outer = outer_first or outer
6619     last = outer_first or last
6620
6621     local d = nodes[q][2]
6622     if d == 'an' or d == 'en' then d = 'r' end
6623     if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
6624
6625     if d == 'on' then
6626         first_on = first_on or q
6627     elseif first_on then
6628         if last == d then
6629             temp = d
6630         else
6631             temp = outer
6632         end
6633         for r = first_on, q - 1 do
6634             nodes[r][2] = temp
6635             item = nodes[r][1]      -- MIRRORING
6636             if Babel.mirroring_enabled and item.id == GLYPH
6637                 and temp == 'r' and characters[item.char] then
6638                 local font_mode = font.fonts[item.font].properties.mode
6639                 if font_mode ~= 'harf' and font_mode ~= 'plug' then
6640                     item.char = characters[item.char].m or item.char
6641                 end
6642             end
6643         end
6644         first_on = nil
6645     end
6646
6647     if d == 'r' or d == 'l' then last = d end
6648 end
6649

```

```

6650 ----- IMPLICIT, REORDER -----
6651
6652 outer = save_outer
6653 last = outer
6654
6655 local state = {}
6656 state.has_r = false
6657
6658 for q = 1, #nodes do
6659     local item = nodes[q][1]
6660
6661     outer = nodes[q][3] or outer
6662
6663     local d = nodes[q][2]
6664
6665     if d == 'nsm' then d = last end          -- W1
6666     if d == 'en' then d = 'an' end
6667     local isdir = (d == 'r' or d == 'l')
6668
6669     if outer == 'l' and d == 'an' then
6670         state.san = state.san or item
6671         state.ean = item
6672     elseif state.san then
6673         head, state = insert_numeric(head, state)
6674     end
6675
6676     if outer == 'l' then
6677         if d == 'an' or d == 'r' then      -- im -> implicit
6678             if d == 'r' then state.has_r = true end
6679             state.sim = state.sim or item
6680             state.eim = item
6681         elseif d == 'l' and state.sim and state.has_r then
6682             head, state = insert_implicit(head, state, outer)
6683         elseif d == 'l' then
6684             state.sim, state.eim, state.has_r = nil, nil, false
6685         end
6686     else
6687         if d == 'an' or d == 'l' then
6688             if nodes[q][3] then -- nil except after an explicit dir
6689                 state.sim = item -- so we move sim 'inside' the group
6690             else
6691                 state.sim = state.sim or item
6692             end
6693             state.eim = item
6694         elseif d == 'r' and state.sim then
6695             head, state = insert_implicit(head, state, outer)
6696         elseif d == 'r' then
6697             state.sim, state.eim = nil, nil
6698         end
6699     end
6700 end
6701
6702 if isdir then
6703     last = d          -- Don't search back - best save now
6704 elseif d == 'on' and state.san then
6705     state.san = state.san or item
6706     state.ean = item
6707 end
6708

```

```

6709 end
6710
6711 return node.prev(head) or head
6712 end
6713 </basic>

```

## 14 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.

## 15 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.

```

6714 <nil>
6715 \ProvidesLanguage{nil}[<<date>> <<version>> Nil language]
6716 \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.

```

6717 \ifx\l@nil\@undefined
6718 \newlanguage\l@nil
6719 \@namedef{bbl@hyphendata@the\l@nil}{\{}}% Remove warning
6720 \let\bbl@elt\relax
6721 \edef\bbl@languages{% Add it to the list of languages
6722 \bbl@languages\bbl@elt{nil}{the\l@nil}{\{}}
6723 \fi

```

This macro is used to store the values of the hyphenation parameters `\lefthyphenmin` and `\righthyphenmin`.

```

6724 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}

```

The next step consists of defining commands to switch to (and from) the ‘nil’ language.

```

\captionnil
\datenil
6725 \let\captionnil\@empty
6726 \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.

```

6727 \ldf@finish{nil}
6728 </nil>

```

## 16 Support for Plain T<sub>E</sub>X (plain.def)

### 16.1 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 `localhyphen.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`.

```
6729 <(*bplain | blplain)
6730 \catcode`\{=1 % left brace is begin-group character
6731 \catcode`\}=2 % right brace is end-group character
6732 \catcode`\#=6 % hash mark is macro parameter character
```

If a file called `hyphen.cfg` can be found, we make sure that *it* will be read instead of the file `hyphen.tex`. 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).

```
6733 \openin 0 hyphen.cfg
6734 \ifeof0
6735 \else
6736   \let\input
```

Then `\input` is defined to forget about its argument and load `hyphen.cfg` instead. Once that’s done the original meaning of `\input` can be restored and the definition of `\a` can be forgotten.

```
6737   \def\input #1 {%
6738     \let\input\input\input\input
6739     \a hyphen.cfg
6740     \let\input\input\input\input
6741   }
6742 \fi
6743 >/bplain | blplain)
```

Now that we have made sure that `hyphen.cfg` will be loaded at the right moment it is time to load `plain.tex`.

```
6744 <bplain>\a plain.tex
6745 <blplain>\a 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.

```
6746 <bplain>\def\fmtname{babel-plain}
6747 <blplain>\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.

### 16.2 Emulating some L<sup>A</sup>T<sub>E</sub>X features

The following code duplicates or emulates parts of L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> that are needed for `babel`.

```
6748 <<(*Emulate LaTeX)>> ≡
```

```

6749 % == Code for plain ==
6750 \def\@empty{}
6751 \def\loadlocalcfg#1{%
6752   \openin0#1.cfg
6753   \ifeof0
6754     \closein0
6755   \else
6756     \closein0
6757     {\immediate\write16{*****}%
6758      \immediate\write16{* Local config file #1.cfg used}%
6759      \immediate\write16{*}%
6760     }
6761     \input #1.cfg\relax
6762   \fi
6763 \endofldf}

```

### 16.3 General tools

A number of  $\text{\LaTeX}$  macro's that are needed later on.

```

6764 \long\def\@firstofone#1{#1}
6765 \long\def\@firstoftwo#1#2{#1}
6766 \long\def\@secondoftwo#1#2{#2}
6767 \def\@nnil{\@nil}
6768 \def\@gobbletwo#1#2{}
6769 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
6770 \def\@star@or@long#1{%
6771   \@ifstar
6772   {\let\l@ngrel@x\relax#1}%
6773   {\let\l@ngrel@x\long#1}}
6774 \let\l@ngrel@x\relax
6775 \def\@car#1#2\@nil{#1}
6776 \def\@cdr#1#2\@nil{#2}
6777 \let\@typeset@protect\relax
6778 \let\protected@edef\edef
6779 \long\def\@gobble#1{}
6780 \edef\@backslashchar{\expandafter\@gobble\string\}
6781 \def\strip@prefix#1>{}
6782 \def\g@addto@macro#1#2{%
6783   \toks@\expandafter{#1#2}%
6784   \xdef#1{\the\toks@}}
6785 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
6786 \def\@nameuse#1{\csname #1\endcsname}
6787 \def\@ifundefined#1{%
6788   \expandafter\ifx\csname#1\endcsname\relax
6789     \expandafter\@firstoftwo
6790   \else
6791     \expandafter\@secondoftwo
6792   \fi}
6793 \def\@expandtwoargs#1#2#3{%
6794   \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
6795 \def\zap@space#1 #2{%
6796   #1%
6797   \ifx#2\@empty\else\expandafter\zap@space\fi
6798   #2}
6799 \let\bbl@trace\@gobble

```

$\text{\LaTeX}_{2\epsilon}$  has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.



```

6800 \ifx\@preamblecmds\undefined
6801   \def\@preamblecmds{}
6802 \fi
6803 \def\@onlypreamble#1{%
6804   \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
6805     \@preamblecmds\do#1}}
6806 \@onlypreamble\@onlypreamble

```

Mimick L<sup>A</sup>T<sub>E</sub>X's \AtBeginDocument; for this to work the user needs to add \begindocument to his file.

```

6807 \def\begindocument{%
6808   \@begindocumenthook
6809   \global\let\@begindocumenthook\@undefined
6810   \def\do##1{\global\let##1\@undefined}%
6811   \@preamblecmds
6812   \global\let\do\noexpand}
6813 \ifx\@begindocumenthook\@undefined
6814   \def\@begindocumenthook{}
6815 \fi
6816 \@onlypreamble\@begindocumenthook
6817 \def\AtBeginDocument{\g@addto@macro\@begindocumenthook}

```

We also have to mimick L<sup>A</sup>T<sub>E</sub>X's \AtEndOfPackage. Our replacement macro is much simpler; it stores its argument in \endoflfd.

```

6818 \def\AtEndOfPackage#1{\g@addto@macro\endoflfd{#1}}
6819 \@onlypreamble\AtEndOfPackage
6820 \def\endoflfd{}
6821 \@onlypreamble\endoflfd
6822 \let\bbl@afterlang\@empty
6823 \chardef\bbl@opt@hyphenmap\z@

```

L<sup>A</sup>T<sub>E</sub>X needs to be able to switch off writing to its auxiliary files; plain doesn't have them by default. There is a trick to hide some conditional commands from the outer \ifx. The same trick is applied below.

```

6824 \catcode`\&=\z@
6825 \ifx&\if@files\@undefined
6826   \expandafter\let\csname if@files\expandafter\endcsname
6827     \csname iffalse\endcsname
6828 \fi
6829 \catcode`\&=4

```

Mimick L<sup>A</sup>T<sub>E</sub>X's commands to define control sequences.

```

6830 \def\newcommand{\@star@or@long\new@command}
6831 \def\new@command#1{%
6832   \@testopt{\@newcommand#1}0}
6833 \def\@newcommand#1[#2]{%
6834   \@ifnextchar [{\@xargdef#1[#2]}%
6835     {\@argdef#1[#2]}}
6836 \long\def\@argdef#1[#2]#3{%
6837   \@yargdef#1\@ne{#2}{#3}}
6838 \long\def\@xargdef#1[#2][#3]#4{%
6839   \expandafter\def\expandafter#1\expandafter{%
6840     \expandafter\@protected@testopt\expandafter #1%
6841     \csname\string#1\expandafter\endcsname{#3}}}%
6842 \expandafter\@yargdef \csname\string#1\endcsname
6843 \tw@{#2}{#4}}
6844 \long\def\@yargdef#1#2#3{%
6845   \@tempcnta#3\relax
6846   \advance \@tempcnta \@ne
6847   \let\@hash@\relax

```

```

6848 \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
6849 \@tempcntb #2%
6850 \@whilenum\@tempcntb <\@tempcnta
6851 \do{%
6852   \edef\reserved@a{\reserved@a\@hash@the\@tempcntb}%
6853   \advance\@tempcntb \@ne}%
6854 \let\@hash@###
6855 \l@ngrel@x\expandafter\def\expandafter#1\reserved@a}
6856 \def\providecommand{\@star@or@long\provide@command}
6857 \def\provide@command#1{%
6858   \begingroup
6859   \escapechar\m@ne\xdef\@gtempa{\string#1}%
6860   \endgroup
6861   \expandafter\ifundefined\@gtempa
6862     {\def\reserved@a{\new@command#1}}%
6863     {\let\reserved@a\relax
6864      \def\reserved@a{\new@command\reserved@a}}%
6865   \reserved@a}%
6866 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
6867 \def\declare@robustcommand#1{%
6868   \edef\reserved@a{\string#1}%
6869   \def\reserved@b{#1}%
6870   \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
6871   \edef#1{%
6872     \ifx\reserved@a\reserved@b
6873       \noexpand\x@protect
6874       \noexpand#1%
6875     \fi
6876     \noexpand\protect
6877     \expandafter\noexpand\csname
6878       \expandafter\@gobble\string#1 \endcsname
6879   }%
6880   \expandafter\new@command\csname
6881     \expandafter\@gobble\string#1 \endcsname
6882 }
6883 \def\x@protect#1{%
6884   \ifx\protect\@typeset@protect\else
6885     \x@protect#1%
6886   \fi
6887 }
6888 \catcode`\&=\z@ % Trick to hide conditionals
6889 \def\x@protect#1&fi#2#3{&fi\protect#1}

```

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`.

```

6890 \def\bbl@tempa{\csname newif\endcsname&fin@}
6891 \catcode`\&=4
6892 \ifx\in@\@undefined
6893   \def\in@#1#2{%
6894     \def\in@##1#1##2##3\in@{%
6895       \ifx\in@##2\in@false\else\in@true\fi}%
6896     \in@#2#1\in@\in@@}
6897 \else
6898   \let\bbl@tempa\@empty
6899 \fi
6900 \bbl@tempa

```

$\LaTeX$  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  $\TeX$  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).

```
6901 \def\ifpackagewith#1#2#3#4{#3}
```

The  $\TeX$  macro `\ifl@aded` checks whether a file was loaded. This functionality is not needed for plain  $\TeX$  but we need the macro to be defined as a no-op.

```
6902 \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  $\TeX 2_{\epsilon}$  versions; just enough to make things work in plain  $\TeX$  environments.

```
6903 \ifx\@tempcnta\@undefined
6904   \csname newcount\endcsname\@tempcnta\relax
6905 \fi
6906 \ifx\@tempcntb\@undefined
6907   \csname newcount\endcsname\@tempcntb\relax
6908 \fi
```

To prevent wasting two counters in  $\TeX$  2.09 (because counters with the same name are allocated later by it) we reset the counter that holds the next free counter (`\count10`).

```
6909 \ifx\bye\@undefined
6910   \advance\count10 by -2\relax
6911 \fi
6912 \ifx\@ifnextchar\@undefined
6913   \def\@ifnextchar#1#2#3{%
6914     \let\reserved@d=#1%
6915     \def\reserved@a{#2}\def\reserved@b{#3}%
6916     \futurelet\@let@token\@ifnch}
6917   \def\@ifnch{%
6918     \ifx\@let@token\@sptoken
6919       \let\reserved@c\@xifnch
6920     \else
6921       \ifx\@let@token\reserved@d
6922         \let\reserved@c\reserved@a
6923       \else
6924         \let\reserved@c\reserved@b
6925       \fi
6926     \fi
6927     \reserved@c}
6928   \def\:{\let\@sptoken= } \: % this makes \@sptoken a space token
6929   \def\:{\@xifnch} \expandafter\def\:{\futurelet\@let@token\@ifnch}
6930 \fi
6931 \def\@testopt#1#2{%
6932   \@ifnextchar[#{#1}{#1[#2]}}
6933 \def\@protected@testopt#1{%
6934   \ifx\protect\@typeset@protect
6935     \expandafter\@testopt
6936   \else
6937     \@x@protect#1%
6938   \fi}
6939 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
6940   #2\relax}\fi}
6941 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
6942   \else\expandafter\@gobble\fi{#1}}
```

## 16.4 Encoding related macros

Code from `ltoutenc.dtx`, adapted for use in the plain  $\TeX$  environment.

```

6943 \def\DeclareTextCommand{%
6944   \@dec@text@cmd\providecommand
6945 }
6946 \def\ProvideTextCommand{%
6947   \@dec@text@cmd\providecommand
6948 }
6949 \def\DeclareTextSymbol#1#2#3{%
6950   \@dec@text@cmd\chardef#1{#2}#3\relax
6951 }
6952 \def\@dec@text@cmd#1#2#3{%
6953   \expandafter\def\expandafter#2%
6954     \expandafter{%
6955       \csname#3-cmd\expandafter\endcsname
6956       \expandafter#2%
6957       \csname#3\string#2\endcsname
6958     }%
6959 %   \let\@ifdefinable\rc@ifdefinable
6960   \expandafter#1\csname#3\string#2\endcsname
6961 }
6962 \def\@current@cmd#1{%
6963   \ifx\protect\@typeset@protect\else
6964     \noexpand#1\expandafter\@gobble
6965   \fi
6966 }
6967 \def\@changed@cmd#1#2{%
6968   \ifx\protect\@typeset@protect
6969     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
6970       \expandafter\ifx\csname ?\string#1\endcsname\relax
6971         \expandafter\def\csname ?\string#1\endcsname{%
6972           \@changed@x@err{#1}%
6973         }%
6974       \fi
6975       \global\expandafter\let
6976         \csname\cf@encoding \string#1\expandafter\endcsname
6977         \csname ?\string#1\endcsname
6978       \fi
6979       \csname\cf@encoding\string#1%
6980       \expandafter\endcsname
6981     \else
6982       \noexpand#1%
6983     \fi
6984 }
6985 \def\@changed@x@err#1{%
6986   \errhelp{Your command will be ignored, type <return> to proceed}%
6987   \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
6988 \def\DeclareTextCommandDefault#1{%
6989   \DeclareTextCommand#1?%
6990 }
6991 \def\ProvideTextCommandDefault#1{%
6992   \ProvideTextCommand#1?%
6993 }
6994 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
6995 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
6996 \def\DeclareTextAccent#1#2#3{%
6997   \DeclareTextCommand#1{#2}[1]{\accent#3 #1}
6998 }
6999 \def\DeclareTextCompositeCommand#1#2#3#4{%
7000   \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
7001   \edef\reserved@b{\string#1}%

```

```

7002 \edef\reserved@c{%
7003   \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
7004 \ifx\reserved@b\reserved@c
7005   \expandafter\expandafter\expandafter\ifx
7006     \expandafter\@car\reserved@a\relax\relax\@nil
7007     \@text@composite
7008   \else
7009     \edef\reserved@b##1{%
7010       \def\expandafter\noexpand
7011         \csname#2\string#1\endcsname####1{%
7012           \noexpand\@text@composite
7013             \expandafter\noexpand\csname#2\string#1\endcsname
7014             ####1\noexpand\@empty\noexpand\@text@composite
7015             {##1}%
7016       }%
7017     }%
7018     \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
7019   \fi
7020   \expandafter\def\csname\expandafter\string\csname
7021     #2\endcsname\string#1-\string#3\endcsname{#4}
7022 \else
7023   \errhelp{Your command will be ignored, type <return> to proceed}%
7024   \errmessage{\string\DeclareTextCompositeCommand\space used on
7025     inappropriate command \protect#1}
7026 \fi
7027 }
7028 \def\@text@composite#1#2#3\@text@composite{%
7029   \expandafter\@text@composite@x
7030   \csname\string#1-\string#2\endcsname
7031 }
7032 \def\@text@composite@x#1#2{%
7033   \ifx#1\relax
7034     #2%
7035   \else
7036     #1%
7037   \fi
7038 }
7039 %
7040 \def\@strip@args#1:#2-#3\@strip@args{#2}
7041 \def\DeclareTextComposite#1#2#3#4{%
7042   \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
7043   \bgroup
7044     \lccode`\@=#4%
7045     \lowercase{%
7046   \egroup
7047     \reserved@a @%
7048   }%
7049 }
7050 %
7051 \def\UseTextSymbol#1#2{#2}
7052 \def\UseTextAccent#1#2#3{}
7053 \def\@use@text@encoding#1{}
7054 \def\DeclareTextSymbolDefault#1#2{%
7055   \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
7056 }
7057 \def\DeclareTextAccentDefault#1#2{%
7058   \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
7059 }
7060 \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.

```
7061 \DeclareTextAccent{"}{OT1}{127}
7062 \DeclareTextAccent{'}{OT1}{19}
7063 \DeclareTextAccent{^}{OT1}{94}
7064 \DeclareTextAccent`}{OT1}{18}
7065 \DeclareTextAccent{~}{OT1}{126}
```

The following control sequences are used in `babel.def` but are not defined for PLAIN  $\text{\TeX}$ .

```
7066 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
7067 \DeclareTextSymbol{\textquotedblright}{OT1}{`"}
7068 \DeclareTextSymbol{\textquoteleft}{OT1}{``}
7069 \DeclareTextSymbol{\textquoteright}{OT1}{''}
7070 \DeclareTextSymbol{\i}{OT1}{16}
7071 \DeclareTextSymbol{\ss}{OT1}{25}
```

For a couple of languages we need the  $\text{\LaTeX}$ -control sequence `\scriptsize` to be available. Because plain  $\text{\TeX}$  doesn't have such a sophisticated font mechanism as  $\text{\LaTeX}$  has, we just `\let` it to `\sevenrm`.

```
7072 \ifx\scriptsize@undefined
7073   \let\scriptsize\sevenrm
7074 \fi
7075 % End of code for plain
7076 <</Emulate LaTeX>>
```

A proxy file:

```
7077 <*plain>
7078 \input babel.def
7079 </plain>
```

## 17 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] Johannes Braams, Victor Eijkhout and Nico Poppelier, *The development of national  $\text{\LaTeX}$  styles*, *TUGboat* 10 (1989) #3, p. 401–406.
- [3] Yannis Haralambous, *Fonts & Encodings*, O'Reilly, 2007.
- [4] Donald E. Knuth, *The  $\text{\TeX}$ book*, Addison-Wesley, 1986.
- [5] Jukka K. Korpela, *Unicode Explained*, O'Reilly, 2006.
- [6] Leslie Lamport,  *$\text{\LaTeX}$ , A document preparation System*, Addison-Wesley, 1986.
- [7] Leslie Lamport, in:  $\text{\TeX}$ hax Digest, Volume 89, #13, 17 February 1989.
- [8] Ken Lunde, *CJKV Information Processing*, O'Reilly, 2nd ed., 2009.
- [9] Hubert Partl, *German  $\text{\TeX}$* , *TUGboat* 9 (1988) #1, p. 70–72.
- [10] Joachim Schrod, *International  $\text{\LaTeX}$  is ready to use*, *TUGboat* 11 (1990) #1, p. 87–90.
- [11] Apostolos Syropoulos, Antonis Tsolomitis and Nick Sofroniu, *Digital typography using  $\text{\LaTeX}$* , Springer, 2002, p. 301–373.
- [12] K.F. Treebus, *Tekstwijzer, een gids voor het grafisch verwerken van tekst*, SDU Uitgeverij ('s-Gravenhage, 1988).