

# Babel

Version 3.60.2399  
2021/06/10

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 . . . . .	8
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 . . . . .	15
1.12	The base option . . . . .	17
1.13	ini files . . . . .	18
1.14	Selecting fonts . . . . .	26
1.15	Modifying a language . . . . .	28
1.16	Creating a language . . . . .	29
1.17	Digits and counters . . . . .	33
1.18	Dates . . . . .	34
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 . . . . .	40
1.23	Selecting scripts . . . . .	41
1.24	Selecting directions . . . . .	42
1.25	Language attributes . . . . .	46
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 . . . . .	51
<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 . . . . .	57
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>62</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>63</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 . . . . .	79
7.9	Basic bidi support . . . . .	80
7.10	Local Language Configuration . . . . .	85
7.11	Language options . . . . .	86
<b>8</b>	<b>The kernel of Babel (<code>babel.def</code>, <code>common</code>)</b>	<b>89</b>
8.1	Tools . . . . .	90
<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 . . . . .	121
9.10	Multiencoding strings . . . . .	122
9.11	Macros common to a number of languages . . . . .	129
9.12	Making glyphs available . . . . .	129
9.12.1	Quotation marks . . . . .	129
9.12.2	Letters . . . . .	131
9.12.3	Shorthands for quotation marks . . . . .	131
9.12.4	Umlauts and tremas . . . . .	132
9.13	Layout . . . . .	134
9.14	Load engine specific macros . . . . .	134
9.15	Creating and modifying languages . . . . .	134
<b>10</b>	<b>Adjusting the Babel behavior</b>	<b>155</b>
<b>11</b>	<b>Loading hyphenation patterns</b>	<b>157</b>
<b>12</b>	<b>Font handling with <code>fontspec</code></b>	<b>162</b>

<b>13</b>	<b>Hooks for XeTeX and LuaTeX</b>	<b>166</b>
13.1	XeTeX	166
13.2	Layout	168
13.3	LuaTeX	170
13.4	Southeast Asian scripts	175
13.5	CJK line breaking	177
13.6	Arabic justification	179
13.7	Common stuff	183
13.8	Automatic fonts and ids switching	183
13.9	Layout	197
13.10	Auto bidi with basic and basic-r	200
<b>14</b>	<b>Data for CJK</b>	<b>211</b>
<b>15</b>	<b>The ‘nil’ language</b>	<b>211</b>
<b>16</b>	<b>Support for Plain T<sub>E</sub>X (plain.def)</b>	<b>212</b>
16.1	Not renaming hyphen.tex	212
16.2	Emulating some L <sup>A</sup> T <sub>E</sub> X features	213
16.3	General tools	213
16.4	Encoding related macros	217
<b>17</b>	<b>Acknowledgements</b>	<b>219</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	8
Unknown language ‘LANG’	9
Argument of \language@active@arg” has an extra }	12
Package fontspec Warning: ‘Language ‘LANG’ not available for font ‘FONT’ with script ‘SCRIPT’ ‘Default’ language used instead’	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 `e-Plain` and `pdf-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 site](#). 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  $\TeX$  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  $\text{\LaTeX}$ , the preamble of the document:

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

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

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

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

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

**NOTE** Some classes load babel with a hardcoded language option. Sometimes, the main language 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 three-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.

### 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>

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

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

```
! 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 e-Plain and pdf-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 those formats. Please, refer to [Using babel with Plain](#) for further details.

## 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`  $\{ \langle language \rangle \}$

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:

```
\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.

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

**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 `bidir` 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 `other language*` does not.

## 1.9 More on selection

**`\babeltags`** `{\langle tag1 \rangle = \langle language1 \rangle, \langle tag2 \rangle = \langle language2 \rangle, ...}`

**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{\langle tag1 \rangle}{\langle text \rangle}` to be `\foreignlanguage{\langle language1 \rangle}{\langle text \rangle}`, and `\begin{\langle tag1 \rangle}` to be `\begin{other language*}{\langle language1 \rangle}`, and so on. Note `\langle tag1 \rangle` 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  $\text{\TeX}$  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{\langle tag \rangle}`, namely, it is not affected by `\MakeUppercase` (while `\foreignlanguage` is).

**`\babelensure`** `[include=\langle commands \rangle, exclude=\langle commands \rangle, fontenc=\langle encoding \rangle]{\langle language \rangle}`

**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,  $\text{\TeX}$  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>4</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 `\kernbcode`, 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`).

**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` `{\shorthands-list}`

**\shorthandoff** `*{\<shorthands-list>}`

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.

**\usesshorthands** `*{\<char>}`

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

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

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

**\defineshorthand** `[\<language>,\<language>,...]{\<shorthand>}{\<code>}`

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{\<lang>}` to the corresponding `\extras{\<lang>}`, 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:

---

<sup>4</sup>With it, encoded strings may not work as expected.

```
\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>5</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:

```
\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>6</sup>

**Languages with no shorthands** Croatian, English (any variety), Indonesian, Hebrew, Interlingua, Irish, Lower Sorbian, Malaysian, North Sami, Romanian, Scottish, Welsh

<sup>5</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.

<sup>6</sup>Thanks to Enrico Gregorio

**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>7</sup>

**\ifbabelshorthand** {<character>}{<true>}{<false>}

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

**\aliasshorthand** {<original>}{<alias>}

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

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

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

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

## 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.



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

<sup>7</sup>This declaration serves to nothing, but it is preserved for backward compatibility.

- noconfigs** Global and language default config files are not loaded, so you can make sure your document is not spoiled by an unexpected .cfg file. However, if the key config is set, this file is loaded.
- showlanguages** Prints to the log the list of languages loaded when the format was created: number (remember dialects can share it), name, hyphenation file and exceptions file.
- nocase** **New 3.9l** Language settings for uppercase and lowercase mapping (as set by \SetCase) are ignored. Use only if there are incompatibilities with other packages.
- silent** **New 3.9l** No warnings and no *infos* are written to the log file.<sup>8</sup>
- strings=** generic | unicode | encoded | *<label>* | *<font encoding>*  
 Selects the encoding of strings in languages supporting this feature. Predefined labels are generic (for traditional T<sub>E</sub>X, L<sup>A</sup>T<sub>E</sub>X and ASCII strings), unicode (for engines like xetex and luatex) and encoded (for special cases requiring mixed encodings). Other allowed values are font encoding codes (T1, T2A, LGR, L7X...), but only in languages supporting them. Be aware with encoded captions are protected, but they work in \MakeUppercase and the like (this feature misuses some internal L<sup>A</sup>T<sub>E</sub>X tools, so use it only as a last resort).
- hyphenmap=** off | first | select | other | other\*  
**New 3.9g** Sets the behavior of case mapping for hyphenation, provided the language defines it.<sup>9</sup> It can take the following values:  
**off** deactivates this feature and no case mapping is applied;  
**first** sets it at the first switching commands in the current or parent scope (typically, when the aux file is first read and at \begin{document}, but also the first \selectlanguage in the preamble), and it's the default if a single language option has been stated;<sup>10</sup>  
**select** sets it only at \selectlanguage;  
**other** also sets it at otherlanguage;  
**other\*** also sets it at otherlanguage\* as well as in heads and foots (if the option headfoot is used) and in auxiliary files (ie, at \select@language), and it's the default if several language options have been stated. The option first can be regarded as an optimized version of other\* for monolingual documents.<sup>11</sup>
- bidi=** default | basic | basic-r | bidi-l | bidi-r  
**New 3.14** Selects the bidi algorithm to be used in luatex and xetex. See sec. 1.24.
- 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

<sup>8</sup>You can use alternatively the package silence.

<sup>9</sup>Turned off in plain.

<sup>10</sup>Duplicated options count as several ones.

<sup>11</sup>Providing foreign is pointless, because the case mapping applied is that at the end of the paragraph, but if either xetex or luatex change this behavior it might be added. On the other hand, other is provided even if I [JBL] think it isn't really useful, but who knows.

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

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

This command is currently the only provided by base. Executes  $\langle code \rangle$  when the file loaded by the corresponding package option is finished (at `\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 Locale 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.

**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]{lᦺ lᦴ lᦶ lᦸ lᦺ lᦴ lᦶ} % 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 short texts the `ini` files should be fine, CJK texts are best set with a dedicated framework (CJK, luatexja, kotex, CTeX, etc.). This is what the class `ltjbook` does with luatex, which can be used in conjunction with the `ldf` for japanese, because the following piece of code loads luatexja:

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

**Latin, Greek, Cyrillic** Combining chars with the default luatex font renderer might be wrong; on the other hand, with the Harfbuzz renderer diacritics are stacked correctly, but many hyphenation 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>	bo	Tibetan <sup>u</sup>
agq	Aghem	brx	Bodo
ak	Akan	bs-Cyrl	Bosnian
am	Amharic <sup>ul</sup>	bs-Latn	Bosnian <sup>ul</sup>
ar	Arabic <sup>ul</sup>	bs	Bosnian <sup>ul</sup>
ar-DZ	Arabic <sup>ul</sup>	ca	Catalan <sup>ul</sup>
ar-MA	Arabic <sup>ul</sup>	ce	Chechen
ar-SY	Arabic <sup>ul</sup>	cgg	Chiga
as	Assamese	chr	Cherokee
asa	Asu	ckb	Central Kurdish
ast	Asturian <sup>ul</sup>	cop	Coptic
az-Cyrl	Azerbaijani	cs	Czech <sup>ul</sup>
az-Latn	Azerbaijani	cu	Church Slavic
az	Azerbaijani <sup>ul</sup>	cu-Cyrs	Church Slavic
bas	Basaa	cu-Glag	Church Slavic
be	Belarusian <sup>ul</sup>	cy	Welsh <sup>ul</sup>
bem	Bemba	da	Danish <sup>ul</sup>
bez	Bena	dav	Taita
bg	Bulgarian <sup>ul</sup>	de-AT	German <sup>ul</sup>
bm	Bambara	de-CH	German <sup>ul</sup>
bn	Bangla <sup>ul</sup>	de	German <sup>ul</sup>

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
en-NZ	English <sup>ul</sup>	ki	Kikuyu
en-US	English <sup>ul</sup>	kk	Kazakh
en	English <sup>ul</sup>	kkj	Kako
eo	Esperanto <sup>ul</sup>	kl	Kalaallisut
es-MX	Spanish <sup>ul</sup>	kln	Kalenjin
es	Spanish <sup>ul</sup>	km	Khmer
et	Estonian <sup>ul</sup>	kn	Kannada <sup>ul</sup>
eu	Basque <sup>ul</sup>	ko	Korean
ewo	Ewondo	kok	Konkani
fa	Persian <sup>ul</sup>	ks	Kashmiri
ff	Fulah	ksb	Shambala
fi	Finnish <sup>ul</sup>	ksf	Bafia
fil	Filipino	ksh	Colognian
fo	Faroese	kw	Cornish
fr	French <sup>ul</sup>	ky	Kyrgyz
fr-BE	French <sup>ul</sup>	lag	Langi
fr-CA	French <sup>ul</sup>	lb	Luxembourgish
fr-CH	French <sup>ul</sup>	lg	Ganda
fr-LU	French <sup>ul</sup>	lkt	Lakota
fur	Friulian <sup>ul</sup>	ln	Lingala
fy	Western Frisian	lo	Lao <sup>ul</sup>
ga	Irish <sup>ul</sup>	lrc	Northern Luri
gd	Scottish Gaelic <sup>ul</sup>	lt	Lithuanian <sup>ul</sup>
gl	Galician <sup>ul</sup>	lu	Luba-Katanga
grc	Ancient Greek <sup>ul</sup>	luo	Luo
gsw	Swiss German	luy	Luyia
gu	Gujarati	lv	Latvian <sup>ul</sup>
guz	Gusii	mas	Masai
gv	Manx	mer	Meru
ha-GH	Hausa	mfe	Morisyen
ha-NE	Hausa <sup>1</sup>	mg	Malagasy
ha	Hausa	mgh	Makhuwa-Meetto
haw	Hawaiian	mgo	Meta'
he	Hebrew <sup>ul</sup>	mk	Macedonian <sup>ul</sup>
hi	Hindi <sup>u</sup>	ml	Malayalam <sup>ul</sup>
hr	Croatian <sup>ul</sup>	mn	Mongolian
hsb	Upper Sorbian <sup>ul</sup>	mr	Marathi <sup>ul</sup>
hu	Hungarian <sup>ul</sup>	ms-BN	Malay <sup>1</sup>
hy	Armenian <sup>u</sup>	ms-SG	Malay <sup>1</sup>
ia	Interlingua <sup>ul</sup>	ms	Malay <sup>ul</sup>
id	Indonesian <sup>ul</sup>	mt	Maltese
ig	Igbo	mua	Mundang

my	Burmese	sn	Shona
mzn	Mazanderani	so	Somali
naq	Nama	sq	Albanian <sup>ul</sup>
nb	Norwegian Bokmål <sup>ul</sup>	sr-Cyrl-BA	Serbian <sup>ul</sup>
nd	North Ndebele	sr-Cyrl-ME	Serbian <sup>ul</sup>
ne	Nepali	sr-Cyrl-XK	Serbian <sup>ul</sup>
nl	Dutch <sup>ul</sup>	sr-Cyrl	Serbian <sup>ul</sup>
nmg	Kwasio	sr-Latn-BA	Serbian <sup>ul</sup>
nn	Norwegian Nynorsk <sup>ul</sup>	sr-Latn-ME	Serbian <sup>ul</sup>
nnh	Ngiemboon	sr-Latn-XK	Serbian <sup>ul</sup>
nus	Nuer	sr-Latn	Serbian <sup>ul</sup>
nyn	Nyankole	sr	Serbian <sup>ul</sup>
om	Oromo	sv	Swedish <sup>ul</sup>
or	Odia	sw	Swahili
os	Ossetic	ta	Tamil <sup>u</sup>
pa-Arab	Punjabi	te	Telugu <sup>ul</sup>
pa-Guru	Punjabi	teo	Teso
pa	Punjabi	th	Thai <sup>ul</sup>
pl	Polish <sup>ul</sup>	ti	Tigrinya
pms	Piedmontese <sup>ul</sup>	tk	Turkmen <sup>ul</sup>
ps	Pashto	to	Tongan
pt-BR	Portuguese <sup>ul</sup>	tr	Turkish <sup>ul</sup>
pt-PT	Portuguese <sup>ul</sup>	twq	Tasawaq
pt	Portuguese <sup>ul</sup>	tzm	Central Atlas Tamazight
qu	Quechua	ug	Uyghur
rm	Romansh <sup>ul</sup>	uk	Ukrainian <sup>ul</sup>
rn	Rundi	ur	Urdu <sup>ul</sup>
ro	Romanian <sup>ul</sup>	uz-Arab	Uzbek
rof	Rombo	uz-Cyrl	Uzbek
ru	Russian <sup>ul</sup>	uz-Latn	Uzbek
rw	Kinyarwanda	uz	Uzbek
rwk	Rwa	vai-Latn	Vai
sa-Beng	Sanskrit	vai-Vaii	Vai
sa-Deva	Sanskrit	vai	Vai
sa-Gujr	Sanskrit	vi	Vietnamese <sup>ul</sup>
sa-Knda	Sanskrit	vun	Vunjo
sa-Mlym	Sanskrit	wae	Walser
sa-Telu	Sanskrit	xog	Soga
sa	Sanskrit	yav	Yangben
sah	Sakha	yi	Yiddish
saq	Samburu	yo	Yoruba
sbp	Sangu	yue	Cantonese
se	Northern Sami <sup>ul</sup>	zgh	Standard Moroccan Tamazight
seh	Sena		
ses	Koyraboro Senni	zh-Hans-HK	Chinese
sg	Sango	zh-Hans-MO	Chinese
shi-Latn	Tachelhit	zh-Hans-SG	Chinese
shi-Tfng	Tachelhit	zh-Hans	Chinese
shi	Tachelhit	zh-Hant-HK	Chinese
si	Sinhala	zh-Hant-MO	Chinese
sk	Slovak <sup>ul</sup>	zh-Hant	Chinese
sl	Slovenian <sup>ul</sup>	zh	Chinese
smn	Inari Sami	zu	Zulu

---

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

---

aghem	cantonese
akan	catalan
albanian	centralatlastamazight
american	centralkurdish
amharic	chechen
ancientgreek	cherokee
arabic	chiga
arabic-algeria	chinese-hans-hk
arabic-DZ	chinese-hans-mo
arabic-morocco	chinese-hans-sg
arabic-MA	chinese-hans
arabic-syria	chinese-hant-hk
arabic-SY	chinese-hant-mo
armenian	chinese-hant
assamese	chinese-simplified-hongkongsarchina
asturian	chinese-simplified-macausarchina
asu	chinese-simplified-singapore
australian	chinese-simplified
austrian	chinese-traditional-hongkongsarchina
azerbaijani-cyrillic	chinese-traditional-macausarchina
azerbaijani-cyrl	chinese-traditional
azerbaijani-latin	chinese
azerbaijani-latn	churchslavic
azerbaijani	churchslavic-cyrs
bafia	churchslavic-oldcyrillic <sup>12</sup>
bambara	churchsslavic-glag
basaa	churchsslavic-glagolitic
basque	cognian
belarusian	cornish
bemba	croatian
bena	czech
bengali	danish
bodo	duala
bosnian-cyrillic	dutch
bosnian-cyrl	dzongkha
bosnian-latin	embu
bosnian-latn	english-au
bosnian	english-australia
brazilian	english-ca
breton	english-canada
british	english-gb
bulgarian	english-newzealand
burmese	english-nz
canadian	english-unitedkingdom

---

<sup>12</sup>The name in the CLDR is Old Church Slavonic Cyrillic, but it has been shortened for practical reasons.



english-unitedstates  
english-us  
english  
esperanto  
estonian  
ewe  
ewondo  
faroese  
filipino  
finnish  
french-be  
french-belgium  
french-ca  
french-canada  
french-ch  
french-lu  
french-luxembourg  
french-switzerland  
french  
friulian  
fulah  
galician  
ganda  
georgian  
german-at  
german-austria  
german-ch  
german-switzerland  
german  
greek  
gujarati  
gusii  
hausa-gh  
hausa-ghana  
hausa-ne  
hausa-niger  
hausa  
hawaiian  
hebrew  
hindi  
hungarian  
icelandic  
igbo  
inarisami  
indonesian  
interlingua  
irish  
italian  
japanese  
jolafonyi  
kabuverdianu  
kabyle  
kako  
kalaallisut

kalenjin  
kamba  
kannada  
kashmiri  
kazakh  
khmer  
kikuyu  
kinyarwanda  
konkani  
korean  
koyraborosenni  
koyrachiini  
kwasio  
kyrgyz  
lakota  
langi  
lao  
latvian  
lingala  
lithuanian  
lowersorbian  
lsorbian  
lubakatanga  
luo  
luxembourgish  
luyia  
macedonian  
machame  
makhuwameetto  
makonde  
malagasy  
malay-bn  
malay-brunei  
malay-sg  
malay-singapore  
malay  
malayalam  
maltese  
manx  
marathi  
masai  
mazanderani  
meru  
meta  
mexican  
mongolian  
morisyen  
mundang  
nama  
nepali  
newzealand  
ngiemboon  
ngomba  
norsk

northernluri  
northernnsami  
northndebele  
norwegianbokmal  
norwegiannynorsk  
nswissgerman  
nuer  
nyankole  
nynorsk  
occitan  
oriya  
oromo  
ossetic  
pashto  
persian  
piedmontese  
polish  
polytonicgreek  
portuguese-br  
portuguese-brazil  
portuguese-portugal  
portuguese-pt  
portuguese  
punjabi-arab  
punjabi-arabic  
punjabi-gurmukhi  
punjabi-guru  
punjabi  
quechua  
romanian  
romansh  
rombo  
rundi  
russian  
rwa  
sakha  
samburu  
samin  
sango  
sangu  
sanskrit-beng  
sanskrit-bengali  
sanskrit-deva  
sanskrit-devanagari  
sanskrit-gujarati  
sanskrit-gujr  
sanskrit-kannada  
sanskrit-knda  
sanskrit-malayalam  
sanskrit-mlym  
sanskrit-telu  
sanskrit-telugu  
sanskrit  
scottishgaelic

sena  
serbian-cyrillic-bosniaherzegovina  
serbian-cyrillic-kosovo  
serbian-cyrillic-montenegro  
serbian-cyrillic  
serbian-cyrl-ba  
serbian-cyrl-me  
serbian-cyrl-xk  
serbian-cyrl  
serbian-latin-bosniaherzegovina  
serbian-latin-kosovo  
serbian-latin-montenegro  
serbian-latin  
serbian-latn-ba  
serbian-latn-me  
serbian-latn-xk  
serbian-latn  
serbian  
shambala  
shona  
sichuanyi  
sinhala  
slovak  
slovene  
slovenian  
soga  
somali  
spanish-mexico  
spanish-mx  
spanish  
standardmoroccantamazight  
swahili  
swedish  
swissgerman  
tachelhit-latin  
tachelhit-latn  
tachelhit-tfng  
tachelhit-tifinagh  
tachelhit  
taita  
tamil  
tasawaq  
telugu  
teso  
thai  
tibetan  
tigrinya  
tongan  
turkish  
turkmen  
ukenglish  
ukrainian  
upporsorbian  
urdu

usenglish	vai-vaii
usorbian	vai
uyghur	vietnam
uzbek-arab	vietnamese
uzbek-arabic	vunjo
uzbek-cyrillic	walser
uzbek-cyrl	welsh
uzbek-latin	westernfrisian
uzbek-latn	yangben
uzbek	yiddish
vai-latin	yoruba
vai-latn	zarma
vai-vai	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=abcdefghijklj`. 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>13</sup>

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

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

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.

---

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

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

**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 imported 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\captionsenglish{%  
  \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\extrasrussian{\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=**  $\langle\textit{language-tag}\rangle$

Loads only the strings. For example:

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

**hyphenrules=**  $\langle\textit{language-list}\rangle$

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 T<sub>E</sub>X 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).

**New 3.58** Another special value is unhyphenated, which activates a line breking mode that allows spaces to be stretched to arbitrary amounts.

**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}
```

Remerber there is an alternative syntax for the latter:

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

**script=**  $\langle\textit{script-name}\rangle$

**New 3.15** Sets the script name to be used by fontspec (eg, Devanagar i). 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=**  $\langle\text{language-name}\rangle$

**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=**  $\langle\text{counter-name}\rangle$

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

**Alph=**  $\langle\text{counter-name}\rangle$

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=**  $\langle\text{base}\rangle$   $\langle\text{shrink}\rangle$   $\langle\text{stretch}\rangle$

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=**  $\langle\text{penalty}\rangle$

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).

**justification=** `kashida` | `elongated` | `unhyphenated`

**New 3.59** There are currently three options, mainly for the Arabic script. It sets the linebreaking and justification method, which can be based on the the ARABIC TATWEEL character or in the ‘justification alternatives’ OpenType table (`jalt`). For an explanation see the [babel site](#).

**linebreaking=** **New 3.59** Just a synonymous for justification.

**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 T<sub>E</sub>X code). This means the local digits have the correct bidirectional behavior (unlike `Numbers=Arabic` in fonts, 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):

- `\localenumerals{<style>}{<number>}`, like `\localenumerals{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, syllabe, 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` [`<calendar=.., variant=..>`]{`<year>`}{`<month>`}{`<day>`}

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. Çileyä Pêşîn 2019*, but with `variant=iza fa` it prints *31'ê Çileyä Pêşînê 2019*.

## 1.19 Accessing language info

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

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

**\iflanguage** `{\language}{\true}{\false}`

If more than one language is used, it might be necessary to know which language is active at a specific time. This can be checked by a call to `\iflanguage`, but note here “language” is used in the  $\TeX$ 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** `{\field}`

**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 פֶּרֶק.

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 `\babenullhyphen`; (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}** {<language>} ... \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 other language\* (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** [*<language>*, *<language>*, ...]{*<patterns>*}

**New 3.9m** *In luatex only*,<sup>14</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

<sup>14</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.

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

Transforms (only luatex) provide a way to process the text on the typesetting level in several language-dependent ways, like non-standard hyphenation, special line breaking rules, script to script conversion, spacing conventions and so on.<sup>15</sup>

It currently embraces `\babelprehyphenation` and `\babelposthyphenation`.

**New 3.57** Several ini files predefine some transforms. They are activated with the key `transforms` in `\babelprovide`, either if the locale is being defined with this macro or the languages has been previously loaded as a class or package option, as the following example illustrates:

```
\usepackage[magyar]{babel}
\babelprovide[transforms = digraphs.hyphen]{magyar}
```

Here are the transforms currently predefined. (More to follow in future releases.)

Arabic	<code>transliteration.dad</code>	Applies the transliteration system devised by Yannis Haralambous for dad (simple and T <sub>E</sub> X-friendly). Not yet complete, but sufficient for most texts.
Croatian	<code>digraphs.ligatures</code>	Ligatures <i>DŽ, Dž, dž, LJ, Lj, lj, NJ, Nj, nj</i> . It assumes they exist. This is not the recommended way to make these transformations (the best way is with OTF features), but it can get you out of a hurry.
Czech, Polish, Portuguese, Slovak, Spanish	<code>hyphen.repeat</code>	Explicit hyphens behave like <code>\babelhyphen{repeat}</code> .
Czech, Polish, Slovak	<code>oneletter.nobreak</code>	Converts a space after a non-syllabic preposition or conjunction into a non-breaking space.
Greek	<code>diaeresis.hyphen</code>	Removes the diaeresis above iota and upsilon if hyphenated just before. It works with the three variants.
Hindi, Sanskrit	<code>transliteration.hk</code>	The Harvard-Kyoto system to romanize Devanagari.
Hindi, Sanskrit	<code>punctuation.space</code>	Inserts a space before the following four characters: <i>!?:;</i> .
Hungarian	<code>digraphs.hyphen</code>	Hyphenates the long digraphs <i>ccs, ddz, ggy, lly, nny, ssz, tty</i> and <i>zsz</i> as <i>cs-cs, dz-dz</i> , etc.

<sup>15</sup>They are similar in concept, but not the same, as those in Unicode.

Arabic, Persian	kashida.plain	Experimental. A very simple and basic transform for ‘plain’ Arabic fonts, which attempts to distribute the tatwil as evenly as possible (starting at the end of the line). See the news for version 3.59.
Serbian	transliteration.gajica	(Note serbian with ini files refers to the Cyrillic script, which is here the target.) The standard system devised by Ljudevit Gaj.

**\babelposthyphenation** {<hyphenrules-name>}{<lua-pattern>}{<replacement>}

**New 3.37-3.39** With *luatex* it is now possible to define non-standard hyphenation rules, like  $f-f \rightarrow ff-f$ , repeated hyphens, ranked ruled (or more precisely, ‘penalized’ hyphenation points), and so on. Only a few rules are currently provided (see below), but they can be defined as shown in the following example, where {1} is the first captured char (between ( ) in the pattern):

```
\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 ([îû]), the replacement could be {1|îû|íú}, which maps î to í, and û to ú, 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, which is particularly useful in transliterations. There are other differences: (1) the first argument is the locale instead of the name of the 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. 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
}
```

**NOTE** With luatex there is another approach to make text transformations, with the function `fonts.handlers.otf.addfeature`, which adds new features to an OTF font (substitution and positioning). These features can be made language-dependent, and babel by default recognizes this setting if the font has been declared with `\babelfont`. The *transforms* mechanism supplements rather than replaces OTF features.

With xetex, where *transforms* are not available, there is still another approach, with font mappings, mainly meant to perform encoding conversions and transliterations. Mappings, however, are linked to fonts, not to languages.

## 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: `fr-Latn-FR` → `fr-Latn` → `fr-FR` → `fr`. Languages with the same resolved name are considered the same. Case is normalized before, so that `fr-latn-fr` → `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>16</sup>

Some languages sharing the same script define macros to switch it (eg, `\textcyrillic`), but be aware they may also set the language to a certain default. Even the babel core defined `\textlatin`, but it was somewhat buggy because in some cases it messed up encodings and fonts (for example, if the main Latin encoding was LY1), and therefore it has been deprecated.<sup>17</sup>

`\ensureascii` `{⟨text⟩}`

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

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

<sup>16</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>17</sup>But still defined for backwards compatibility.

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:

```
\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 (الاريفي) بـ
    Arabia أو Aravia (بالاريفية 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 فصحى العصر \textit{fuṣḥā l-‘aṣr} (MSA) and
    فصحى التراث \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 `\textthe` must be defined to select the main language):

```

\newcommand\refrange[2]{\babelsublr{\textthe{\ref{#1}}-\textthe{\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>18</sup>

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

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

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

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

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

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

**tabular** required in `luatex` for R `tabular`, 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 `\LaTeXe` **New 3.19** .

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

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

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

<sup>18</sup>Next on the roadmap are counters and numeral systems in general. Expect some minor readjustments.

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

```
RTL A ltr text \thechapter{} and still ltr RTL B
```

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

```
RTL A \foreignlanguage{english}{ltr text \thechapter{} and still ltr} RTL B
```

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

Mainly for bidi text, but it 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`**  $\langle cmd \rangle \langle local\text{-}language \rangle \langle before \rangle \langle after \rangle$

**New 3.17** Something like:

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

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

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

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

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

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

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

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

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

## 1.25 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 T<sub>E</sub>X 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, francaais, 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>19</sup>  
**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).

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

<sup>19</sup>The two last name comes from the times when they had to be shortened to 8 characters

**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{\_}{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 \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**  $\langle key-value-list \rangle$

**New 3.36** Sometimes you might need to disable some babel features. Currently this macro understands the following keys (and only for luatex), with values on or off: bidi.text, bidi.mirroring, bidi.mapdigits, layout.lists, layout.tabular, linebreak.sea, linebreak.cjk, justify.arabic. For example, you can set \babeladjust{bidi.text=off} if you are using an alternative algorithm or with large sections not requiring it. 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), L<sup>A</sup>T<sub>E</sub>X 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 T<sub>E</sub>X only takes into account the values when the paragraph is hyphenated, i.e., when it has been finished.<sup>20</sup> So, if you write a chunk of French text with `\foreignlanguage`, the apostrophes might not be taken into account. This is a limitation of T<sub>E</sub>X, not of babel. Alternatively, you may use `\usesorthands` to activate ' and `\definesorthand`, or redefine `\textquoteright` (the latter is called by the non-ASCII right quote).
- `\bibitem` is out of sync with `\selectlanguage` in the .aux file. The reason is `\bibitem` uses `\immediate` (and others, in fact), while `\selectlanguage` doesn't. There is no known workaround.
- Babel does not take into account `\normalsfcodes` and (non-)French spacing is not always properly (un)set by languages. However, problems are unlikely to happen and therefore this part remains untouched in version 3.9 (but it is in the 'to do' list).
- Using a character mathematically active (ie, with math code "8000) as a shorthand can make T<sub>E</sub>X enter in an infinite loop in some rare cases. (Another issue in the 'to do' list, although there is a partial solution.)

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

**csquotes** Logical markup for quotes.

**iflang** Tests correctly the current language.

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

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

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

**biblatex** Programmable bibliographies and citations.

**bicaption** Bilingual captions.

**babelbib** Multilingual bibliographies.

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

Ligatures can be disabled.

**substitutefont** Combines fonts in several encodings.

**mkpattern** Generates hyphenation patterns.

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

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

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

### 1.31 Current and future work

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

Useful additions would be, for example, time, currency, addresses and personal names.<sup>21</sup>

But that is the easy part, because they don't require modifying the L<sup>A</sup>T<sub>E</sub>X internals.

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

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

<sup>20</sup>This explains why L<sup>A</sup>T<sub>E</sub>X assumes the lowercase mapping of T1 and does not provide a tool for multiple mappings. Unfortunately, `\savingsphcodes` is not a solution either, because lccodes for hyphenation are frozen in the format and cannot be changed.

<sup>21</sup>See for example POSIX, ISO 14652 and the Unicode Common Locale Data Repository (CLDR). Those systems, however, have limited application to T<sub>E</sub>X because their aim is just to display information and not fine typesetting.

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

### 1.32 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 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`

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

**New 3.9q** With luatex, however, patterns are loaded on the fly when requested by the language (except the “0th” language, typically english, which is preloaded always).<sup>22</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>23</sup>

### 2.1 Format

In that file the person who maintains a  $\text{\TeX}$  environment has to record for which languages he has hyphenation patterns *and* in which files these are stored<sup>24</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  $\text{\LaTeX}$  that the hyphenation patterns just processed have to be known under an alternative name. Here is an example:

<sup>22</sup>This feature was added to 3.9o, but it was buggy. Both 3.9o and 3.9p are deprecated.

<sup>23</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>24</sup>This is because different operating systems sometimes use very different file-naming conventions.

```
% 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>25</sup> For example:

```
german:T1 hyphenT1.ger
german hyphen.ger
```

With the previous settings, if the encoding when the language is selected is T1 then the patterns in `hyphenT1.ger` are used, but otherwise use those in `hyphen.ger` (note the encoding 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

<sup>25</sup>This is not a new feature, but in former versions it didn't work correctly.

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>26</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.

<sup>26</sup>But not removed, for backward compatibility.

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

The following page provides a starting point for ldf files:

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

<https://latex3.github.io/babel/guides/list-of-locale-templates.html>.

If you need further assistance and technical advice in the development of language styles, I am willing to help you. And of course, you can make any suggestion you like.

## 3.2 Basic macros

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

**\addlanguage** The macro \addlanguage is a non-outer version of the macro \newlanguage, defined in plain.tex version 3.x. Here “language” is used in the TeX sense of set of hyphenation patterns.

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

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

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

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

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

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

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

**\extras<lang>** The macro \extras<lang> contains all the extra definitions needed for a specific language. This macro, like the following, is a hook – you can add things to it, but it must not be used directly.

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



<code>\bbl@declare@ttribute</code>	This is a command to be used in the language definition files for declaring a language attribute. It takes three arguments: the name of the language, the attribute to be defined, and the code to be executed when the attribute is to be used.
<code>\main@language</code>	To postpone the activation of the definitions needed for a language until the beginning of a document, all language definition files should use <code>\main@language</code> instead of <code>\selectlanguage</code> . This will just store the name of the language, and the proper language will be activated at the start of the document.
<code>\ProvidesLanguage</code>	The macro <code>\ProvidesLanguage</code> should be used to identify the language definition files. Its syntax is similar to the syntax of the $\TeX$ command <code>\ProvidesPackage</code> .
<code>\LdfInit</code>	The macro <code>\LdfInit</code> performs a couple of standard checks that must be made at the beginning of a language definition file, such as checking the category code of the <code>@</code> -sign, preventing the <code>.ldf</code> file from being processed twice, etc.
<code>\ldf@quit</code>	The macro <code>\ldf@quit</code> does work needed if a <code>.ldf</code> file was processed earlier. This includes resetting the category code of the <code>@</code> -sign, preparing the language to be activated at <code>\begin{document}</code> time, and ending the input stream.
<code>\ldf@finish</code>	The macro <code>\ldf@finish</code> does work needed at the end of each <code>.ldf</code> file. This includes resetting the category code of the <code>@</code> -sign, loading a local configuration file, and preparing the language to be activated at <code>\begin{document}</code> time.
<code>\loadlocalcfg</code>	After processing a language definition file, $\TeX$ can be instructed to load a local configuration file. This file can, for instance, be used to add strings to <code>\captions{&lt;lang&gt;}</code> to support local document classes. The user will be informed that this configuration file has been loaded. This macro is called by <code>\ldf@finish</code> .
<code>\substitutefontfamily</code>	(Deprecated.) This command takes three arguments, a font encoding and two font family names. It creates a font description file for the first font in the given encoding. This <code>.fd</code> file will instruct $\TeX$ 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>

\bbl@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\monthiname{<name of first month>}
% More strings

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

\StartBabelCommands*{<dialect>}{date}
\SetString\monthiname{<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  $\text{\LaTeX}$  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` The command `\bbl@activate` is used to change the way an active character expands.  
`\bbl@deactivate` `\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` The  $\text{\TeX}$ book states: “Plain  $\text{\TeX}$  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]  
`\bbl@remove@special` 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`.  $\text{\LaTeX}$  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>27</sup>.

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

`\babel@savevariable` A second macro is provided to save the current value of a variable. In this context, anything that is allowed after the `\the` primitive is considered to be a variable. The macro takes one argument, the `\variable`.  
The effect of the preceding macros is to append a piece of code to the current definition of `\originalTeX`. When `\originalTeX` is expanded, this code restores the previous definition of the control sequence or the previous value of the variable.

### 3.6 Support for extending macros

`\addto` The macro `\addto{<control sequence>}{<TeX code>}` can be used to extend the definition of a macro. The macro need not be defined (ie, it can be undefined or `\relax`). This macro can, for instance, be used in adding instructions to a macro like `\extrasenglish`. Be careful when using this macro, because depending on the case the assignment 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

<sup>27</sup>This mechanism was introduced by Bernd Raichle.

below). In other words, the old way of defining/switching strings still works and it's used by default.

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

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

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

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

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

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

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

The  $\langle\textit{category}\rangle$  is either `captions`, `date` or `extras`. You must stick to these three categories, even if no error is raised when using other name.<sup>28</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:

---

<sup>28</sup>In future releases further categories may be added.

```

\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}

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

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


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

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

\EndBabelCommands

```

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

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

The starred version just forces strings to take a value – if not set as package option, then the default for the engine is used. This is not done by default to prevent backward incompatibilities, but if you are creating a new language this version is better. It's up to the maintainers of the current languages to decide if using it is appropriate.<sup>29</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  $\backslash EndBabelCommands$ .

<sup>29</sup>This replaces in 3.9g a short-lived  $\backslash UseStrings$  which has been removed because it did not work.

**\SetString** {*<macro-name>*}{*<string>*}

Adds *<macro-name>* to the current category, and defines globally *<lang-macro-name>* to *<code>* (after applying the transformation corresponding to the current charset or defined with the hook `stringprocess`).

Use this command to define strings, without including any “logic” if possible, which should be a separated macro. See the example above for the date.

**\SetStringLoop** {*<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  $\TeX$ , 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  $\TeX$  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  $\TeX$  primitive (`\lccode`), babel sets them separately.

There are three helper macros to be used inside `\SetHyphenMap`:

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

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

```
\SetHyphenMap{\BabelLowerMM{"100"}{"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`. 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  $\TeX$  package, which sets options and loads language styles.

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

**hyphen.cfg** is the file to be used when generating the formats to load hyphenation patterns.

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 counter s has been devised to have arbitrary keys, so you can add lowercased keys if you want.

## 7 Tools

```
1 <<version=3.60.2399>>
2 <<date=2021/06/10>>
```

**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  $\text{\LaTeX}$  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` 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>30</sup>. These macros will break if another `\if... \fi` statement appears in one of the arguments and it is not enclosed in braces.

`\bbl@afterfi`

```
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}
```

<sup>30</sup>This code is based on code presented in TUGboat vol. 12, no2, June 1991 in “An expansion Power Lemma” by Sonja Maus.



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

```

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

```

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

```

`\bbl@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\bbl@ifblank#1{%
69   \bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
70 \long\def\bbl@ifblank@i#1#2\@nil#3#4#5\@nil{#4}
71 \def\bbl@ifset#1#2#3{%
72   \bbl@ifunset{#1}{#3}{\bbl@exp{\bbl@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\bbl@forkv#1#2{%
74   \def\bbl@kvcmd##1##2##3{#2}%
75   \bbl@kvnext#1,\@nil,}
76 \def\bbl@kvnext#1,{%

```

```

77 \ifx\@nil#1\relax\else
78 \bbl@ifblank{#1}{\bbl@forkv@eq#1=\@empty=\@nil{#1}}%
79 \expandafter\bbl@kvnext
80 \fi}
81 \def\bbl@forkv@eq#1=#2=#3\@nil#4{%
82 \bbl@trim@def\bbl@forkv@a{#1}%
83 \bbl@trim{\expandafter\bbl@kvcmd\expandafter{\bbl@forkv@a}}{#2}{#4}}

```

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

```

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 \def\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<sub>ε</sub>X, 1 is luatex, and 2 is 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  $\LaTeX$  macro. The following code is placed before them to define (and then undefine) if not in  $\LaTeX$ .

```
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  $\LaTeX 2.09$ . In that case the file `plain.def` is needed (which also defines `\AtBeginDocument`, and therefore it is not loaded twice). We need the first part when the format is created, and `\orig@dump` is used as a flag. Otherwise, we need to use the second part, so `\orig@dump` is not defined (`plain.def` undefines it). Check if the current version of `switch.def` has been previously loaded (mainly, `hyphen.cfg`). If not, load it now. We cannot load `babel.def` here because we first need to declare and process the package options.

## 7.2 The Package File ( $\LaTeX$ , `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\bbbl@trace[1]{\message{^^J[ #1 ]}}%
196    \let\bbbl@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' 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\bbl@add\csname#1.ldf-h@@k\endcsname}%

```

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

```

273 \ifx\bbl@languages\undefined\else
274 \begingroup
275 \catcode\^^I=12
276 \@ifpackagewith{babel}{showlanguages}{%
277 \begingroup
278 \def\bbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}%
279 \wlog{<*languages>}%
280 \bbl@languages
281 \wlog{</languages>}%
282 \endgroup}{%
283 \endgroup
284 \def\bbl@elt#1#2#3#4{%
285 \ifnum#2=\z@
286 \gdef\bbl@nulllanguage{#1}%
287 \def\bbl@elt##1##2##3##4{}}%
288 \fi}%
289 \bbl@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 \bbl@trace{Defining option 'base'}
292 \@ifpackagewith{babel}{base}{%
293 \let\bbl@onlyswitch\@empty
294 \let\bbl@provide@locale\relax
295 \input babel.def
296 \let\bbl@onlyswitch\@undefined
297 \ifx\directlua\@undefined
298 \DeclareOption*{\bbl@patterns{\CurrentOption}}%
299 \else
300 \input luababel.def
301 \DeclareOption*{\bbl@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\bb1@sh@string
401 \fi}
402 \ifx\bb1@opt@shorthands\@nnil
403 \def\bb1@ifshorthand#1#2#3{#2}%
404 \else\ifx\bb1@opt@shorthands\@empty
405 \def\bb1@ifshorthand#1#2#3{#3}%
406 \else

```

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

```

407 \def\bb1@ifshorthand#1{%
408 \bb1@xin@{\string#1}{\bb1@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\bb1@opt@shorthands{%
415 \expandafter\bb1@sh@string\bb1@opt@shorthands\@empty}%

```

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

```

416 \bb1@ifshorthand{'}%
417 {\PassOptionsToPackage{activeacute}{babel}}{}
418 \bb1@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\bb1@opt@headfoot\@nnil\else
422 \g@addto@macro\@resetactivechars{%
423 \set@typeset@protect
424 \expandafter\select@language@x\expandafter{\bb1@opt@headfoot}%
425 \let\protect\noexpand}
426 \fi

```

For the option safe we use a different approach – \bb1@opt@safe says which macros are redefined (B for bibs and R for refs). By default, both are set.

```

427 \ifx\bb1@opt@safe\@undefined
428 \def\bb1@opt@safe{BR}
429 \fi
430 \ifx\bb1@opt@main\@nnil\else
431 \edef\bb1@language@opts{%
432 \ifx\bb1@language@opts\@empty\else\bb1@language@opts,\fi
433 \bb1@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 \bb1@trace{Defining IfBabelLayout}
436 \ifx\bb1@opt@layout\@nnil
437 \newcommand\IfBabelLayout[3]{#3}%
438 \else
439 \newcommand\IfBabelLayout[1]{%
440 \@expandtwoargs\in@{.#1.}{.\bb1@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@activetrue
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     \@tempwattrue
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@activetrue
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 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@redefineroobust\ref#1{%
487 \@safe@activetrue\org@ref{#1}\@safe@activesfalse}
488 \bbl@redefineroobust\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 Bi $\TeX$  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 `\bibcite` is needed we define `\bibcite` in such a way that it redefines itself with the proper definition. We call `\bbl@cite@choice` to select the proper definition for `\bibcite`. This new definition is then activated.

```

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

```

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

```

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

```

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

```

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

```

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

```

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@activetrue\org@@bibitem{#1}\@safe@activesfalse}
525 \else
526   \let\org@nocite\nocite
527   \let\org@@citex\@citex
528   \let\org@bibcite\bibcite
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\bbbl@bidimode\else % Only with bidi. See also above
539     \edef\thepage{%
540         \noexpand\babelsublr{\unexpanded\expandafter{\thepage}}}%
541     \fi}%
542 \fi}
543 {\ifbbbl@single\else
544     \bbbl@ifunset{markright } \bbbl@redefine\bbbl@redefineroobust
545     \markright#1{%
546         \bbbl@ifblank{#1}%
547         {\org@markright{}}}%
548         {\toks@{#1}}%
549         \bbbl@exp{%
550             \\org@markright{\\protect\\foreignlanguage{\language}%
551                 {\\protect\\bbbl@restore@actives\the\toks@}}}%

```

`\markboth`    The definition of `\markboth` is equivalent to that of `\markright`, except that we need two token registers. The documentclasses report and book define and set the headings for the page. While doing so they also store a copy of `\markboth` in `\@mkboth`. Therefore we need to check whether `\@mkboth` has already been set. If so we need to do that again with the new definition of `\markboth`. (As of Oct 2019,  $\text{\LaTeX}$  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\bbbl@tempc{\let\@mkboth\markboth}
554     \else
555     \def\bbbl@tempc{}
556     \fi
557     \bbbl@ifunset{markboth } \bbbl@redefine\bbbl@redefineroobust
558     \markboth#1#2{%
559         \protected@edef\bbbl@tempb##1{%
560             \protect\foreignlanguage
561             {\language}{\protect\bbbl@restore@actives##1}}%
562         \bbbl@ifblank{#1}%
563         {\toks@{}}%
564         {\toks@\expandafter{\bbbl@tempb{#1}}}%
565         \bbbl@ifblank{#2}%
566         {\@temptokena{}}%
567         {\@temptokena\expandafter{\bbbl@tempb{#2}}}%
568         \bbbl@exp{\\org@markboth{\the\toks@}{\the\@temptokena}}
569         \bbbl@tempc
570     \fi} % end ifbbbl@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@activetrue
586        #2}%
587       {\let\pageref\bbl@temp@pref
588        \let\ref\bbl@temp@ref
589        \@safe@activetrue
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  
`\vrefpagemum` to prevent problems when an active character ends up in the argument of `\vref`. The same needs to  
`\Ref` happen for `\vrefpagemum`.

```

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

```

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 L<sup>A</sup>T<sub>E</sub>X.

```
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]^^J
630     \string\DeclareFontFamily{#1}{#2}{^^J
631     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{^^J
632     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{^^J
633     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{^^J
634     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{^^J
635     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{^^J
636     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{^^J
637     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{^^J
638     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{^^J
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  $\text{\TeX}$  and  $\text{\LaTeX}$  always come out in the right encoding. There is a list of non-ASCII encodings. Unfortunately, fontenc deletes its package options, so we must guess which encodings has been loaded by traversing  $\text{\@filelist}$  to search for  $\langle enc \rangle \text{enc.def}$ . If a non-ASCII has been loaded, we define versions of  $\text{\TeX}$  and  $\text{\LaTeX}$  for them using  $\text{\ensureascii}$ . The default ASCII encoding is set, too (in reverse order): the “main” encoding (when the document begins), the last loaded, or OT1.

$\text{\ensureascii}$

```
643 \bbl@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  $\text{\begin{document}}$ , which latin fontencoding to use.

$\text{\latinencoding}$

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  $\text{\begin{document}}$  whether it was loaded with the T1 option. The normal way to do this (using  $\text{\@ifpackageloaded}$ ) is disabled for this package. Now we have to revert to parsing the internal macro  $\text{\@filelist}$  which contains all the filenames loaded.



```

680 \AtBeginDocument{%
681   \ifpackageloaded{fontspec}%
682     {\xdef\latinencoding{%
683       \ifx\UTFencname\@undefined
684         EU\ifcase\bb1@engine\or2\or1\fi
685       \else
686         \UTFencname
687       \fi}}%
688   {\gdef\latinencoding{OT1}%
689     \ifx\cf@encoding\bb1@t@one
690       \xdef\latinencoding{\bb1@t@one}%
691     \else
692       \ifx\@fontenc@load@list\@undefined
693         \ifl@aded{def}{t1enc}{\xdef\latinencoding{\bb1@t@one}}}%
694       \else
695         \def\@elt#1{,#1,}%
696         \edef\bb1@tempa{\expandafter\@gobbletwo\@fontenc@load@list}%
697         \let\@elt\relax
698         \bb1@xin@{,T1,}\bb1@tempa
699         \ifin@
700           \xdef\latinencoding{\bb1@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 `LuaTeX-jā` 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 `LTEX`. 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\bbbl@bidimode\or
761 require('babel-bidi-basic.lua')
762 \or
763 require('babel-bidi-basic-r.lua')
764 \fi}
765 % TODO - to locale_props, not as separate attribute
766 \newattribute\bbbl@attr@dir
767 % TODO. I don't like it, hackish:
768 \bbbl@exp{\output{\bodydir\pagedir\the\output}}
769 \AtEndOfPackage{\EnableBabelHook{babel-bidi}}
770 \fi\fi
771 \else
772 \ifnum\bbbl@bidimode>100 \ifnum\bbbl@bidimode<200
773 \bbbl@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\bbbl@beforeforeign\leavevmode
779 \AtEndOfPackage{%
780 \EnableBabelHook{babel-bidi}%
781 \bbbl@xebidipar}
782 \fi\fi
783 \def\bbbl@loadxebidi#1{%
784 \ifx\RTLfootnotetext\@undefined
785 \AtEndOfPackage{%
786 \EnableBabelHook{babel-bidi}%
787 \ifx\fontspec\@undefined
788 \bbbl@loadfontspec % bidi needs fontspec
789 \fi
790 \usepackage#1{bidi}}%
791 \fi}
792 \ifnum\bbbl@bidimode>200
793 \ifcase\expandafter\@gobbletwo\the\bbbl@bidimode\or
794 \bbbl@tentative{bidi=bidi}
795 \bbbl@loadxebidi{}
796 \or
797 \bbbl@loadxebidi{[rldocument]}
798 \or
799 \bbbl@loadxebidi{}
800 \fi
801 \fi
802 \fi
803 \ifnum\bbbl@bidimode=\@ne
804 \let\bbbl@beforeforeign\leavevmode
805 \ifodd\bbbl@engine
806 \newattribute\bbbl@attr@dir
807 \bbbl@exp{\output{\bodydir\pagedir\the\output}}%
808 \fi
809 \AtEndOfPackage{%
810 \EnableBabelHook{babel-bidi}%
811 \ifodd\bbbl@engine\else
812 \bbbl@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 \bbbl@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@lsys@\languagename}{\bbl@provide@lsys{\languagename}}{}%
846   \bbl@ifunset{bbl@wdir@\languagename}{\bbl@provide@dirs{\languagename}}{}%
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 rl
868     \ifcase#3\relax
869       \ifcase\bbl@getluadir{#1}\relax\else
870         #2 TLT\relax
871       \fi
872     \else

```

```

873     \ifcase\bb1@getluadir{#1}\relax
874     #2 TRT\relax
875     \fi
876   \fi}
877 \def\bb1@textdir#1{%
878   \bb1@setluadir{text}\textdir{#1}%
879   \chardef\bb1@thetextdir#1\relax
880   \setattribute\bb1@attr@dir{\numexpr\bb1@thepardir*3+#1}}
881 \def\bb1@pardir#1{%
882   \bb1@setluadir{par}\pardir{#1}%
883   \chardef\bb1@thepardir#1\relax}
884 \def\bb1@bodydir{\bb1@setluadir{body}\bodydir}
885 \def\bb1@pagedir{\bb1@setluadir{page}\pagedir}
886 \def\bb1@dirparastext{\pardir\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 \ifnum\bb1@bidimode>\z@
890   \def\bb1@mathboxdir{%
891     \ifcase\bb1@thetextdir\relax
892       \everyhbox{\bb1@mathboxdir@aux L}%
893     \else
894       \everyhbox{\bb1@mathboxdir@aux R}%
895     \fi}
896   \def\bb1@mathboxdir@aux#1{%
897     \@ifnextchar\egroup{}\textdir T#1T\relax}}
898   \frozen@everymath\expandafter{%
899     \expandafter\bb1@mathboxdir\the\frozen@everymath}
900   \frozen@everydisplay\expandafter{%
901     \expandafter\bb1@mathboxdir\the\frozen@everydisplay}
902   \fi
903 \else % pdftex=0, xetex=2
904   \newcount\bb1@dirlevel
905   \chardef\bb1@thetextdir\z@
906   \chardef\bb1@thepardir\z@
907   \def\bb1@textdir#1{%
908     \ifcase#1\relax
909       \chardef\bb1@thetextdir\z@
910       \bb1@textdir@i\beginL\endL
911     \else
912       \chardef\bb1@thetextdir\@ne
913       \bb1@textdir@i\beginR\endR
914     \fi}
915   \def\bb1@textdir@i#1#2{%
916     \ifhmode
917       \ifnum\currentgrouplevel>\z@
918         \ifnum\currentgrouplevel=\bb1@dirlevel
919           \bb1@error{Multiple bidi settings inside a group}%
920           {I'll insert a new group, but expect wrong results.}%
921           \bgroup\aftergroup#2\aftergroup\egroup
922         \else
923           \ifcase\currentgrouptype\or % 0 bottom
924             \aftergroup#2% 1 simple {}
925           \or
926             \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
927           \or
928             \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
929           \or\or\or % vbox vtop align
930           \or
931             \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign

```

```

932      \or\or\or\or\or\or % output math disc insert vcent mathchoice
933      \or
934      \aftergroup#2% 14 \begingroup
935      \else
936      \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
937      \fi
938      \fi
939      \bbl@dirlevel\currentgrouplevel
940      \fi
941      #1%
942      \fi}
943 \def\bbl@pdir#1{\chardef\bbl@thepardir#1\relax}
944 \let\bbl@bodydir\@gobble
945 \let\bbl@pagedir\@gobble
946 \def\bbl@dirparastext{\chardef\bbl@thepardir\bbl@thetextdir}

```

The following command is executed only if there is a right-to-left script (once). It activates the `\everypar` hack for xetex, to properly handle the par direction. Note text and par dirs are decoupled to some extent (although not completely).

```

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

```

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

```

973 \DeclareRobustCommand\babelsublr[1]{\leavevmode\bbl@textdir\z@#1}}
974 \AtBeginDocument{%
975   \ifx\pdfstringdefDisableCommands\undefined\else
976     \ifx\pdfstringdefDisableCommands\relax\else
977       \pdfstringdefDisableCommands{\let\babelsublr\@firstofone}%
978     \fi
979   \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`.

```

980 \bbl@trace{Local Language Configuration}
981 \ifx\loadlocalcfg\undefined
982   \@ifpackagewith{babel}{noconfigs}%
983   {\let\loadlocalcfg@gobble}%
984   {\def\loadlocalcfg#1{%
985     \InputIfFileExists{#1.cfg}%
986     {\typeout{*****^J%
987               * Local config file #1.cfg used^^J%
988               *}}}%
989   \@empty}}
990 \fi

```

## 7.11 Language options

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

```

991 \bbl@trace{Language options}
992 \let\bbl@afterlang\relax
993 \let\BabelModifiers\relax
994 \let\bbl@loaded\@empty
995 \def\bbl@load@language#1{%
996   \InputIfFileExists{#1.ldf}%
997   {\edef\bbl@loaded{\CurrentOption
998     \ifx\bbl@loaded\@empty\else,\bbl@loaded\fi}%
999     \expandafter\let\expandafter\bbl@afterlang
1000     \csname\CurrentOption.ldf-h@@k\endcsname
1001     \expandafter\let\expandafter\BabelModifiers
1002     \csname bbl@mod@\CurrentOption\endcsname}%
1003   {\bbl@error{%
1004     Unknown option '\CurrentOption'. Either you misspelled it\\%
1005     or the language definition file \CurrentOption.ldf was not found}}%
1006     Valid options are, among others: shorthands=, KeepShorthandsActive,\\%
1007     activeacute, activegrave, noconfigs, safe=, main=, math=\\%
1008     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.

```

1009 \def\bbl@try@load@lang#1#2#3{%
1010   \IfFileExists{\CurrentOption.ldf}%
1011   {\bbl@load@language{\CurrentOption}}%
1012   {#1\bbl@load@language{#2}#3}}
1013 \DeclareOption{hebrew}{%
1014   \input{rlbabel.def}%
1015   \bbl@load@language{hebrew}}
1016 \DeclareOption{hungarian}{\bbl@try@load@lang{}{magyar}{}}
1017 \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}{}}
1018 \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}{}}
1019 \DeclareOption{polutonikogreek}{%
1020   \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}}
1021 \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}{}}
1022 \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}{}}
1023 \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.

```

1024 \ifx\bbl@opt@config\@nnil
1025 \@ifpackagewith{babel}{noconfigs}{}%
1026 {\InputIfFileExists{bblopts.cfg}%
1027 {\typeout{*****^J%
1028          * Local config file bblopts.cfg used^^J%
1029          *}}}%
1030 {}}%
1031 \else
1032 \InputIfFileExists{\bbl@opt@config.cfg}%
1033 {\typeout{*****^J%
1034          * Local config file \bbl@opt@config.cfg used^^J%
1035          *}}}%
1036 {\bbl@error{%
1037   Local config file '\bbl@opt@config.cfg' not found}{%
1038   Perhaps you misspelled it.}}}%
1039 \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.

```

1040 \let\bbl@tempc\relax
1041 \bbl@foreach\bbl@language@opts{%
1042 \ifcase\bbl@iniflag % Default
1043 \bbl@ifunset{ds@#1}%
1044 {\DeclareOption{#1}{\bbl@load@language{#1}}}%
1045 {}%
1046 \or % provide=*
1047 \@gobble % case 2 same as 1
1048 \or % provide+=*
1049 \bbl@ifunset{ds@#1}%
1050 {\IfFileExists{#1.ldf}{}%
1051 {\IfFileExists{babel-#1.tex}{\@namedef{ds@#1}}}%
1052 {}%
1053 \bbl@ifunset{ds@#1}%
1054 {\def\bbl@tempc{#1}%
1055 \DeclareOption{#1}{%
1056 \ifnum\bbl@iniflag>\@ne
1057 \bbl@ldfinit
1058 \babelprovide[import]{#1}%
1059 \bbl@afterldf}%
1060 \else
1061 \bbl@load@language{#1}%
1062 \fi}}%
1063 {}%
1064 \or % provide*=*
1065 \def\bbl@tempc{#1}%
1066 \bbl@ifunset{ds@#1}%
1067 {\DeclareOption{#1}{%
1068 \bbl@ldfinit
1069 \babelprovide[import]{#1}%
1070 \bbl@afterldf}}}%
1071 {}%
1072 \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.

```

1073 \let\bbl@tempb\@nnil
1074 \bbl@foreach\@classoptionslist{%
1075   \bbl@ifunset{ds@#1}%
1076     {\IfFileExists{#1.ldf}%
1077       {\def\bbl@tempb{#1}%
1078         \DeclareOption{#1}{%
1079           \ifnum\bbl@iniflag>\@ne
1080             \bbl@ldfinit
1081             \babelprovide[import]{#1}%
1082             \bbl@afterldf}%
1083           \else
1084             \bbl@load@language{#1}%
1085           \fi}}%
1086     {\IfFileExists{babel-#1.tex}% TODO. Copypaste pattern
1087       {\def\bbl@tempb{#1}%
1088         \DeclareOption{#1}{%
1089           \ifnum\bbl@iniflag>\@ne
1090             \bbl@ldfinit
1091             \babelprovide[import]{#1}%
1092             \bbl@afterldf}%
1093           \else
1094             \bbl@load@language{#1}%
1095           \fi}}%
1096       {}}}%
1097   {}}

```

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

```

1098 \ifnum\bbl@iniflag=\z@ \else
1099   \ifx\bbl@opt@main\@nnil
1100     \ifx\bbl@tempc\relax
1101       \let\bbl@opt@main\bbl@tempb
1102     \else
1103       \let\bbl@opt@main\bbl@tempc
1104     \fi
1105   \fi
1106 \fi
1107 \ifx\bbl@opt@main\@nnil \else
1108   \expandafter
1109   \let\expandafter\bbl@loadmain\csname ds@\bbl@opt@main\endcsname
1110   \expandafter\let\csname ds@\bbl@opt@main\endcsname\@empty
1111 \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):

```

1112 \def\AfterBabelLanguage#1{%
1113   \bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang}{}}
1114 \DeclareOption*{}
1115 \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`.

```

1116 \bbl@trace{Option 'main'}

```

```

1117 \ifx\bbl@opt@main\@nnil
1118 \edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
1119 \let\bbl@tempc\@empty
1120 \bbl@for\bbl@tempb\bbl@tempa{%
1121   \bbl@xin@{\,\bbl@tempb,}{\,\bbl@loaded,}%
1122   \ifin@{\edef\bbl@tempc{\bbl@tempb}\fi}
1123 \def\bbl@tempa#1,#2\@nnil{\def\bbl@tempb{#1}}
1124 \expandafter\bbl@tempa\bbl@loaded,\@nnil
1125 \ifx\bbl@tempb\bbl@tempc\else
1126   \bbl@warning{%
1127     Last declared language option is '\bbl@tempc',\%
1128     but the last processed one was '\bbl@tempb'.\%
1129     The main language can't be set as both a global\%
1130     and a package option. Use 'main=\bbl@tempc' as\%
1131     option. Reported}%
1132   \fi
1133 \else
1134   \ifodd\bbl@iniflag % case 1,3
1135     \bbl@ldfinit
1136     \let\CurrentOption\bbl@opt@main
1137     \bbl@exp{\@babelprovide[import,main]{\bbl@opt@main}}
1138     \bbl@afterldf{}%
1139   \else % case 0,2
1140     \chardef\bbl@iniflag\z@ % Force ldf
1141     \expandafter\let\csname ds@\bbl@opt@main\endcsname\bbl@loadmain
1142     \ExecuteOptions{\bbl@opt@main}
1143     \DeclareOption*{}%
1144     \ProcessOptions*
1145   \fi
1146 \fi
1147 \def\AfterBabelLanguage{%
1148   \bbl@error
1149   {Too late for \string\AfterBabelLanguage}%
1150   {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.

1151 \ifx\bbl@main@language\@undefined
1152   \bbl@info{%
1153     You haven't specified a language. I'll use 'nil'\%
1154     as the main language. Reported}
1155   \bbl@load@language{nil}
1156 \fi
1157 \</package>
1158 \<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  $\LaTeX$ , some of it is for the  $\LaTeX$  case only.

Plain formats based on `etex` (`etex`, `xetex`, `luatex`) don't load `hyphen.cfg` but `etex.src`, which follows a different naming convention, so we need to define the babel names. It presumes `language.def`

exists and it is the same file used when formats were created.

## 8.1 Tools

```

1159 \ifx\ldf@quit\undefined\else
1160 \endinput\fi % Same line!
1161 <<Make sure ProvidesFile is defined>>
1162 \ProvidesFile{babel.def}[\<date>] \<version>] Babel common definitions]

```

The file `babel.def` expects some definitions made in the  $\text{\LaTeX} 2_{\epsilon}$  style file. So, In  $\text{\LaTeX} 2.09$  and Plain we must provide at least some predefined values as well some tools to set them (even if not all options are available). There 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).

```

1163 \ifx\AtBeginDocument\undefined % TODO. change test.
1164 <<Emulate LaTeX>>
1165 \def\language{english}%
1166 \let\bbl@opt@shorthands\@nnil
1167 \def\bbl@ifshorthand#1#2#3{#2}%
1168 \let\bbl@language@opts\@empty
1169 \ifx\babeloptionstrings\undefined
1170 \let\bbl@opt@strings\@nnil
1171 \else
1172 \let\bbl@opt@strings\babeloptionstrings
1173 \fi
1174 \def\BabelStringsDefault{generic}
1175 \def\bbl@tempa{normal}
1176 \ifx\babeloptionmath\bbl@tempa
1177 \def\bbl@mathnormal{\noexpand\textormath}
1178 \fi
1179 \def\AfterBabelLanguage#1#2{}
1180 \ifx\BabelModifiers\undefined\let\BabelModifiers\relax\fi
1181 \let\bbl@afterlang\relax
1182 \def\bbl@opt@safe{BR}
1183 \ifx\@uclclist\undefined\let\@uclclist\@empty\fi
1184 \ifx\bbl@trace\undefined\def\bbl@trace#1{}\fi
1185 \expandafter\newif\csname ifbbl@single\endcsname
1186 \chardef\bbl@bidimode\z@
1187 \fi

```

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

```

1188 \ifx\bbl@trace\undefined
1189 \let\ldfInit\endinput
1190 \def\ProvidesLanguage#1{\endinput}
1191 \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.

```

1192 <<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.

```

1193 \def\bbl@version{\<version>}
1194 \def\bbl@date{\<date>}

```

```

1195 \def\adddialect#1#2{%
1196   \global\chardef#1#2\relax
1197   \bbl@usehooks{adddialect}{#1}{#2}}%
1198   \begingroup
1199     \count@#1\relax
1200     \def\bbl@elt##1##2##3##4{%
1201       \ifnum\count@=##2\relax
1202         \edef\bbl@tempa{\expandafter@gobbletwo\string#1}%
1203         \bbl@info{Hyphen rules for '\expandafter@gobble\bbl@tempa'
1204           set to \expandafter\string\csname l@##1\endcsname\\%
1205           (\string\language\the\count@). Reported}%
1206         \def\bbl@elt####1####2####3####4{%
1207           \fi}%
1208         \bbl@cs{languages}%
1209       \endgroup}

```

\bbl@iflanguage executes code only if the language l@ exists. Otherwise raises an error. The argument of \bbl@fixname has to be a macro name, as it may get “fixed” if casing (lc/uc) is wrong. It’s intended to fix a long-standing bug when \foreignlanguage and the like appear in a \MakeXXXcase. However, a lowercase form is not imposed to improve backward compatibility (perhaps you defined a language named MYLANG, but unfortunately mixed case names cannot be trapped). Note l@ is encapsulated, so that its case does not change.

```

1210 \def\bbl@fixname#1{%
1211   \begingroup
1212   \def\bbl@tempe{l@}%
1213   \edef\bbl@tempd{\noexpand@ifundefined{\noexpand\bbl@tempe#1}}%
1214   \bbl@tempd
1215     {\lowercase\expandafter{\bbl@tempd}%
1216     {\uppercase\expandafter{\bbl@tempd}%
1217     \@empty
1218     {\edef\bbl@tempd{\def\noexpand#1{#1}}%
1219     \uppercase\expandafter{\bbl@tempd}}}%
1220     {\edef\bbl@tempd{\def\noexpand#1{#1}%
1221     \lowercase\expandafter{\bbl@tempd}}}%
1222     \@empty
1223   \edef\bbl@tempd{\endgroup\def\noexpand#1{#1}}%
1224   \bbl@tempd
1225   \bbl@exp{\bbl@usehooks{language}{\language}{#1}}%
1226 \def\bbl@iflanguage#1{%
1227   \ifundefined{l@#1}{\@nolanner{#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.

```

1228 \def\bbl@bcpcase#1#2#3#4\@@#5{%
1229   \ifx\@empty#3%
1230     \uppercase{\def#5{#1#2}}%
1231   \else
1232     \uppercase{\def#5{#1}}%
1233     \lowercase{\edef#5{#5#2#3#4}}%
1234   \fi}
1235 \def\bbl@bcpllookup#1-#2-#3-#4\@@{%
1236   \let\bbl@bcp\relax
1237   \lowercase{\def\bbl@tempa{#1}}%
1238   \ifx\@empty#2%
1239     \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{%
1240     \else\ifx\@empty#3%

```

```

1241 \bbl@bcpcase#2\@empty\@empty\@@\bbl@tempb
1242 \IfFileExists{babel-\bbl@tempa-\bbl@tempb.ini}%
1243 {\edef\bbl@bcp{\bbl@tempa-\bbl@tempb}}}%
1244 {}%
1245 \ifx\bbl@bcp\relax
1246 \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}}}%
1247 \fi
1248 \else
1249 \bbl@bcpcase#2\@empty\@empty\@@\bbl@tempb
1250 \bbl@bcpcase#3\@empty\@empty\@@\bbl@tempc
1251 \IfFileExists{babel-\bbl@tempa-\bbl@tempb-\bbl@tempc.ini}%
1252 {\edef\bbl@bcp{\bbl@tempa-\bbl@tempb-\bbl@tempc}}}%
1253 {}%
1254 \ifx\bbl@bcp\relax
1255 \IfFileExists{babel-\bbl@tempa-\bbl@tempc.ini}%
1256 {\edef\bbl@bcp{\bbl@tempa-\bbl@tempc}}}%
1257 {}%
1258 \fi
1259 \ifx\bbl@bcp\relax
1260 \IfFileExists{babel-\bbl@tempa-\bbl@tempc.ini}%
1261 {\edef\bbl@bcp{\bbl@tempa-\bbl@tempc}}}%
1262 {}%
1263 \fi
1264 \ifx\bbl@bcp\relax
1265 \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}}}%
1266 \fi
1267 \fi\fi}
1268 \let\bbl@initoload\relax
1269 \def\bbl@provide@locale{%
1270 \ifx\babelprovide\undefined
1271 \bbl@error{For a language to be defined on the fly 'base'\\%
1272 is not enough, and the whole package must be\\%
1273 loaded. Either delete the 'base' option or\\%
1274 request the languages explicitly}%
1275 {See the manual for further details.}%
1276 \fi
1277 % TODO. Option to search if loaded, with \LocaleForEach
1278 \let\bbl@auxname\language\name % Still necessary. TODO
1279 \bbl@ifunset{bbl@bcp@map@\language\name}}}% Move uplevel??
1280 {\edef\language\name{\@nameuse{bbl@bcp@map@\language\name}}}%
1281 \ifbbl@bcp@allowed
1282 \expandafter\ifx\csname date\language\name\endcsname\relax
1283 \expandafter
1284 \bbl@bcp@lookup\language\name-\@empty-\@empty-\@empty\@@
1285 \ifx\bbl@bcp\relax\else % Returned by \bbl@bcp@lookup
1286 \edef\language\name{\bbl@bcp@prefix\bbl@bcp}%
1287 \edef\localename{\bbl@bcp@prefix\bbl@bcp}%
1288 \expandafter\ifx\csname date\language\name\endcsname\relax
1289 \let\bbl@initoload\bbl@bcp
1290 \bbl@exp{\bbl@babelprovide[\bbl@autoload@bcptoptions]{\language\name}}}%
1291 \let\bbl@initoload\relax
1292 \fi
1293 \bbl@csarg\xdef{bcp@map@\bbl@bcp}{\localename}%
1294 \fi
1295 \fi
1296 \fi
1297 \expandafter\ifx\csname date\language\name\endcsname\relax
1298 \IfFileExists{babel-\language\name.tex}%
1299 {\bbl@exp{\bbl@babelprovide[\bbl@autoload@options]{\language\name}}}%

```

```

1300      {}%
1301    \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.

```

1302 \def\iflanguage#1{%
1303   \bbl@iflanguage{#1}{%
1304     \ifnum\csname l@#1\endcsname=\language
1305       \expandafter\@firstoftwo
1306     \else
1307       \expandafter\@secondoftwo
1308     \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.

```

1309 \let\bbl@select@type\z@
1310 \edef\selectlanguage{%
1311   \noexpand\protect
1312   \expandafter\noexpand\csname selectlanguage \endcsname}

```

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

```

1313 \ifx\@undefined\protect\let\protect\relax\fi

```

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

```

1314 \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.

```

1315 \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`

```

1316 \def\bbl@push@language{%
1317   \ifx\language\undefined\else
1318     \xdef\bbl@language@stack{\language+\bbl@language@stack}%
1319   \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`.

```
1320 \def\bbl@pop@lang#1+#2\@@{%
1321   \edef\language{#1}%
1322   \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).

```
1323 \let\bbl@ifrestoring\@secondoftwo
1324 \def\bbl@pop@language{%
1325   \expandafter\bbl@pop@lang\bbl@language@stack\@@
1326   \let\bbl@ifrestoring\@firstoftwo
1327   \expandafter\bbl@set@language\expandafter{\language}%
1328   \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).

```
1329 \chardef\localeid\z@
1330 \def\bbl@id@last{0} % No real need for a new counter
1331 \def\bbl@id@assign{%
1332   \bbl@ifunset\bbl@id@\language}%
1333   {\count@\bbl@id@last\relax
1334     \advance\count@\@ne
1335     \bbl@csarg\chardef{id@\language}\count@
1336     \edef\bbl@id@last{\the\count@}%
1337     \ifcase\bbl@engine\or
1338       \directlua{
1339         Babel = Babel or {}
1340         Babel.locale_props = Babel.locale_props or {}
1341         Babel.locale_props[\bbl@id@last] = {}
1342         Babel.locale_props[\bbl@id@last].name = '\language'
1343       }%
1344     \fi}%
1345   }%
1346   \chardef\localeid\bbl@cl{id@}}
```

The unprotected part of `\selectlanguage`.

```
1347 \expandafter\def\csname selectlanguage \endcsname#1{%
1348   \ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel\tw@\fi
1349   \bbl@push@language
1350   \aftergroup\bbl@pop@language
1351   \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.

```

1352 \def\BabelContentsFiles{toc,lof,lot}
1353 \def\bbl@set@language#1{% from selectlanguage, pop@
1354 % The old buggy way. Preserved for compatibility.
1355 \edef\language{%
1356   \ifnum\escapechar=\expandafter`\string#1\@empty
1357   \else\string#1\@empty\fi}%
1358 \ifcat\relax\noexpand#1%
1359   \expandafter\ifx\csname date\language\endcsname\relax
1360     \edef\language{#1}%
1361     \let\localname\language
1362   \else
1363     \bbl@info{Using '\string\language' instead of 'language' is\\%
1364       deprecated. If what you want is to use a\\%
1365       macro containing the actual locale, make\\%
1366       sure it does not not match any language.\\%
1367       Reported}%
1368     % I'll\\%
1369     % try to fix '\string\localname', but I cannot promise\\%
1370     % anything. Reported}%
1371     \ifx\scantokens\@undefined
1372       \def\localname{??}%
1373     \else
1374       \scantokens\expandafter{\expandafter
1375         \def\expandafter\localname\expandafter{\language}}%
1376     \fi
1377   \fi
1378 \else
1379   \def\localname{#1}% This one has the correct catcodes
1380 \fi
1381 \select@language{\language}%
1382 % write to auxs
1383 \expandafter\ifx\csname date\language\endcsname\relax\else
1384   \if@filesw
1385     \ifx\babel@aux\@gobbletwo\else % Set if single in the first, redundant
1386       % \bbl@savelastskip
1387       \protected@write\@auxout{}\{\string\babel@aux{\bbl@auxname}\}%
1388       % \bbl@restorelastskip
1389     \fi
1390     \bbl@usehooks{write}\}%
1391   \fi
1392 \fi}
1393 % The following is used above to deal with skips before the write
1394 % whatsit. Adapted from hyperref, but it might fail, so for the moment
1395 % it's not activated. TODO.
1396 \def\bbl@savelastskip{%
1397   \let\bbl@restorelastskip\relax
1398   \ifvmode
1399     \ifdim\lastskip=\z@
1400       \let\bbl@restorelastskip\nobreak
1401     \else
1402       \bbl@exp{%
1403         \def\\bbl@restorelastskip{%
1404           \skip@=\the\lastskip
1405           \\nobreak \vskip-\skip@ \vskip\skip@}}%
1406       \fi
1407     \fi}
1408 \newif\ifbbl@bcpallowed
1409 \bbl@bcpallowedfalse
1410 \def\select@language#1{% from set@, babel@aux

```



```

1411 % set hymap
1412 \ifnum\bbl@hymapsel=\@ccclv\chardef\bbl@hymapsel4\relax\fi
1413 % set name
1414 \edef\language{#1}%
1415 \bbl@fixname\language
1416 % TODO. name@map must be here?
1417 \bbl@provide@locale
1418 \bbl@iflanguage\language{%
1419   \expandafter\ifx\csname date\language\endcsname\relax
1420     \bbl@error
1421     {Unknown language '\language'. Either you have\\%
1422       misspelled its name, it has not been installed,\\%
1423       or you requested it in a previous run. Fix its name,\\%
1424       install it or just rerun the file, respectively. In\\%
1425       some cases, you may need to remove the aux file}%
1426     {You may proceed, but expect wrong results}%
1427   \else
1428     % set type
1429     \let\bbl@select@type\z@
1430     \expandafter\bbl@switch\expandafter{\language}%
1431     \fi}}
1432 \def\babel@aux#1#2{% TODO. See how to avoid undefined nil's
1433   \select@language{#1}%
1434   \bbl@foreach\BabelContentsFiles{%
1435     \@writefile{##1}{\babel@toc{#1}{#2}}}% % TODO - ok in plain?
1436 \def\babel@toc#1#2{%
1437   \select@language{#1}}

```

First, check if the user asks for a known language. If so, update the value of `\language` and call `\originalTeX` to bring  $\TeX$  in a certain pre-defined state.

The name of the language is stored in the control sequence `\language`.

Then we have to redefine `\originalTeX` to compensate for the things that have been activated. To save memory space for the macro definition of `\originalTeX`, we construct the control sequence name for the `\noextras<lang>` command at definition time by expanding the `\csname` primitive. Now activate the language-specific definitions. This is done by constructing the names of three macros by concatenating three words with the argument of `\selectlanguage`, and calling these macros.

The switching of the values of `\lefthyphenmin` and `\righthyphenmin` is somewhat different. First we save their current values, then we check if `\<lang>hyphenmins` is defined. If it is not, we set default values (2 and 3), otherwise the values in `\<lang>hyphenmins` will be used.

```

1438 \newif\ifbbl@usedategroup
1439 \def\bbl@switch#1{% from select@, foreign@
1440   % make sure there is info for the language if so requested
1441   \bbl@ensureinfo{#1}%
1442   % restore
1443   \originalTeX
1444   \expandafter\def\expandafter\originalTeX\expandafter{%
1445     \csname noextras#1\endcsname
1446     \let\originalTeX\@empty
1447     \babel@beginsave}%
1448   \bbl@usehooks{afterreset}{}%
1449   \languageshorthands{none}%
1450   % set the locale id
1451   \bbl@id@assign
1452   % switch captions, date
1453   % No text is supposed to be added here, so we remove any
1454   % spurious spaces.
1455   \bbl@bsphack
1456   \ifcase\bbl@select@type

```

```

1457     \csname captions#1\endcsname\relax
1458     \csname date#1\endcsname\relax
1459     \else
1460         \bbl@xin@{,captions,}{, \bbl@select@opts,}%
1461         \ifin@
1462             \csname captions#1\endcsname\relax
1463             \fi
1464             \bbl@xin@{,date,}{, \bbl@select@opts,}%
1465             \ifin@ % if \foreign... within \<lang>date
1466                 \csname date#1\endcsname\relax
1467             \fi
1468         \fi
1469     \bbl@esphack
1470     % switch extras
1471     \bbl@usehooks{beforeextras}{}%
1472     \csname extras#1\endcsname\relax
1473     \bbl@usehooks{afterextras}{}%
1474     % > babel-ensure
1475     % > babel-sh-<short>
1476     % > babel-bidi
1477     % > babel-fontspec
1478     % hyphenation - case mapping
1479     \ifcase\bbl@opt@hyphenmap\or
1480         \def\BabelLower##1##2{\lccode##1=##2\relax}%
1481         \ifnum\bbl@hymapsel>4\else
1482             \csname\language @bbl@hyphenmap\endcsname
1483             \fi
1484         \chardef\bbl@opt@hyphenmap\z@
1485     \else
1486         \ifnum\bbl@hymapsel>\bbl@opt@hyphenmap\else
1487             \csname\language @bbl@hyphenmap\endcsname
1488             \fi
1489         \fi
1490         \let\bbl@hymapsel@cclv
1491     % hyphenation - select rules
1492     \ifnum\csname l@\language\endcsname=\l@unhyphenated
1493         \edef\bbl@tempa{u}%
1494     \else
1495         \edef\bbl@tempa{\bbl@ccl\lnbrk}%
1496     \fi
1497     % linebreaking - handle u, e, k (v in the future)
1498     \bbl@xin@{/u}{/\bbl@tempa}%
1499     \ifin@ \else \bbl@xin@{/e}{/\bbl@tempa} \fi % elongated forms
1500     \ifin@ \else \bbl@xin@{/k}{/\bbl@tempa} \fi % only kashida
1501     \ifin@ \else \bbl@xin@{/v}{/\bbl@tempa} \fi % variable font
1502     \ifin@
1503         % unhyphenated/kashida/elongated = allow stretching
1504         \language\l@unhyphenated
1505         \babel@savevariable\emergencystretch
1506         \emergencystretch\maxdimen
1507         \babel@savevariable\hbadness
1508         \hbadness\@M
1509     \else
1510         % other = select patterns
1511         \bbl@patterns{#1}%
1512     \fi
1513     % hyphenation - mins
1514     \babel@savevariable\lefthyphenmin
1515     \babel@savevariable\righthyphenmin

```

```

1516 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
1517   \set@hyphenmins\tw@\thr@@\relax
1518 \else
1519   \expandafter\expandafter\expandafter\set@hyphenmins
1520   \csname #1hyphenmins\endcsname\relax
1521 \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.

```

1522 \long\def\otherlanguage#1{%
1523   \ifnum\bbl@hymapsel=\ccclv\let\bbl@hymapsel\thr@@\fi
1524   \csname selectlanguage \endcsname{#1}%
1525   \ignorespaces}

```

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

```

1526 \long\def\endotherlanguage{%
1527   \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`.

```

1528 \expandafter\def\csname otherlanguage*\endcsname{%
1529   \@ifnextchar[\bbl@otherlanguage@s{\bbl@otherlanguage@s[]}}
1530 \def\bbl@otherlanguage@s[#1]#2{%
1531   \ifnum\bbl@hymapsel=\ccclv\chardef\bbl@hymapsel4\relax\fi
1532   \def\bbl@select@opts{#1}%
1533   \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”.

```

1534 \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`.

```

1535 \providecommand\bbl@beforeforeign{}
1536 \edef\foreignlanguage{%
1537   \noexpand\protect
1538   \expandafter\noexpand\csname foreignlanguage \endcsname}
1539 \expandafter\def\csname foreignlanguage \endcsname{%
1540   \@ifstar\bbl@foreign@s\bbl@foreign@x}
1541 \providecommand\bbl@foreign@x[3][]{%
1542   \begingroup
1543     \def\bbl@select@opts{#1}%
1544     \let\BabelText\@firstofone
1545     \bbl@beforeforeign
1546     \foreign@language{#2}%
1547     \bbl@usehooks{foreign}{}%
1548     \BabelText{#3}% Now in horizontal mode!
1549   \endgroup}
1550 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \@setpar, ?\@@par
1551   \begingroup
1552     {\par}%
1553     \let\bbl@select@opts\@empty
1554     \let\BabelText\@firstofone
1555     \foreign@language{#1}%
1556     \bbl@usehooks{foreign*}{}%
1557     \bbl@dirparastext
1558     \BabelText{#2}% Still in vertical mode!
1559   {\par}%
1560   \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`.

```

1561 \def\foreign@language#1{%
1562   % set name
1563   \edef\language#1}%
1564   \ifbbl@usedategroup
1565     \bbl@add\bbl@select@opts{,date,}%
1566     \bbl@usedategroupfalse
1567   \fi
1568   \bbl@fixname\language#1
1569   % TODO. name@map here?
1570   \bbl@provide@locale
1571   \bbl@iflanguage\language#1{%
1572     \expandafter\ifx\csname date\language#1\endcsname\relax
1573       \bbl@warning % TODO - why a warning, not an error?
1574         {Unknown language '#1'. Either you have\\%
1575           misspelled its name, it has not been installed,\\%
1576           or you requested it in a previous run. Fix its name,\\%
1577           install it or just rerun the file, respectively. In\\%
1578           some cases, you may need to remove the aux file.\\%
1579           I'll proceed, but expect wrong results.\\%
1580           Reported}%
1581     \fi
1582     % set type
1583     \let\bbl@select@type\@ne
1584     \expandafter\bbl@switch\expandafter{\language#1}}

```

`\bbl@patterns` This macro selects the hyphenation patterns by changing the `\language` register. If special hyphenation patterns are available specifically for the current font encoding, use them instead of the default.

It also sets hyphenation exceptions, but only once, because they are global (here language \lccode's has been set, too). \bbl@hyphenation@ is set to relax until the very first \babelhyphenation, so do nothing with this value. If the exceptions for a language (by its number, not its name, so that :ENC is taken into account) has been set, then use \hyphenation with both global and language exceptions and empty the latter to mark they must not be set again.

```

1585 \let\bbl@hyphlist\@empty
1586 \let\bbl@hyphenation@\relax
1587 \let\bbl@pttnlist\@empty
1588 \let\bbl@patterns@\relax
1589 \let\bbl@hymapsel=\@cclv
1590 \def\bbl@patterns#1{%
1591   \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
1592     \csname l@#1\endcsname
1593     \edef\bbl@tempa{#1}%
1594   \else
1595     \csname l@#1:\f@encoding\endcsname
1596     \edef\bbl@tempa{#1:\f@encoding}%
1597   \fi
1598   \@expandtwoargs\bbl@usehooks{patterns}{{#1}{\bbl@tempa}}%
1599   % > luatex
1600   \@ifundefined{bbl@hyphenation@}{}{% Can be \relax!
1601     \begingroup
1602       \bbl@xin@{,\number\language,}{,\bbl@hyphlist}%
1603       \ifin@ \else
1604         \@expandtwoargs\bbl@usehooks{hyphenation}{{#1}{\bbl@tempa}}%
1605         \hyphenation{%
1606           \bbl@hyphenation@
1607           \@ifundefined{bbl@hyphenation@#1}%
1608             \@empty
1609             {\space\csname bbl@hyphenation@#1\endcsname}}%
1610         \xdef\bbl@hyphlist{\bbl@hyphlist\bbl@hyphlist\number\language,}%
1611       \fi
1612     \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*`.

```

1613 \def\hyphenrules#1{%
1614   \edef\bbl@tempf{#1}%
1615   \bbl@fixname\bbl@tempf
1616   \bbl@iflanguage\bbl@tempf{%
1617     \expandafter\bbl@patterns\expandafter{\bbl@tempf}%
1618     \ifx\languageshortands\undefined\else
1619       \languageshortands{none}%
1620     \fi
1621     \expandafter\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
1622       \set@hyphenmins\tw@\thr@@\relax
1623     \else
1624       \expandafter\expandafter\expandafter\set@hyphenmins
1625       \csname\bbl@tempf hyphenmins\endcsname\relax
1626     \fi}}
1627 \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.

```

1628 \def\providehyphenmins#1#2{%

```

```

1629 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
1630 \namedef{#1hyphenmins}{#2}%
1631 \fi}

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

1632 \def\set@hyphenmins#1#2{%
1633 \lefthyphenmin#1\relax
1634 \righthyphenmin#2\relax}

\ProvidesLanguage The identification code for each file is something that was introduced in  $\TeX$  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.

1635 \ifx\ProvidesFile\@undefined
1636 \def\ProvidesLanguage#1[#2 #3 #4]{%
1637 \wlog{Language: #1 #4 #3 <#2>}%
1638 }
1639 \else
1640 \def\ProvidesLanguage#1{%
1641 \begingroup
1642 \catcode`\ 10 %
1643 \@makeother\/%
1644 \ifnextchar[%]
1645 {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}
1646 \def\@provideslanguage#1[#2]{%
1647 \wlog{Language: #1 #2}%
1648 \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
1649 \endgroup}
1650 \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.

1651 \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.

1652 \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’:

1653 \providecommand\setlocale{%
1654 \bbl@error
1655 {Not yet available}%
1656 {Find an armchair, sit down and wait}}
1657 \let\uselocale\setlocale
1658 \let\locale\setlocale
1659 \let\selectlocale\setlocale
1660 \let\localename\setlocale
1661 \let\textlocale\setlocale
1662 \let\textlanguage\setlocale
1663 \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.

`\@nopatterns`

`\@noopterr` When the package was loaded without options not everything will work as expected. An error message is issued in that case.

When the format knows about `\PackageError` it must be  $\text{\TeX 2}\epsilon$ , 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.

```

1664 \edef\bbl@nulllanguage{\string\language=0}
1665 \ifx\PackageError\undefined % TODO. Move to Plain
1666   \def\bbl@error#1#2{%
1667     \begingroup
1668       \newlinechar=`^^J
1669       \def\{^^J(babel) }%
1670       \errhelp{#2}\errmessage{\{#1}%
1671     \endgroup}
1672   \def\bbl@warning#1{%
1673     \begingroup
1674       \newlinechar=`^^J
1675       \def\{^^J(babel) }%
1676       \message{\{#1}%
1677     \endgroup}
1678   \let\bbl@infowarn\bbl@warning
1679   \def\bbl@info#1{%
1680     \begingroup
1681       \newlinechar=`^^J
1682       \def\{^^J}%
1683       \wlog{#1}%
1684     \endgroup}
1685 \fi
1686 \def\bbl@nocaption{\protect\bbl@nocaption@i}
1687 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
1688   \global\@namedef{#2}{\textbf{?#1?}}%
1689   \@nameuse{#2}%
1690   \edef\bbl@tempa{#1}%
1691   \bbl@sreplace\bbl@tempa{name}{}%
1692   \bbl@warning{% TODO.
1693     \@backslashchar#1 not set for '\language'. Please,\%
1694     define it after the language has been loaded\%
1695     (typically in the preamble) with:\%
1696     \string\setlocalecaption{\language}{\bbl@tempa}{..\}%
1697     Reported}}
1698 \def\bbl@tentative{\protect\bbl@tentative@i}
1699 \def\bbl@tentative@i#1{%
1700   \bbl@warning{%
1701     Some functions for '#1' are tentative.\%
1702     They might not work as expected and their behavior\%
1703     could change in the future.\%
1704     Reported}}
1705 \def\@nolanerr#1{%
1706   \bbl@error
1707   {You haven't defined the language '#1' yet.\%
1708     Perhaps you misspelled it or your installation\%
1709     is not complete}%
1710   {Your command will be ignored, type <return> to proceed}}
1711 \def\@nopatterns#1{%
1712   \bbl@warning
1713   {No hyphenation patterns were preloaded for\%
1714     the language '#1' into the format.\%
1715     Please, configure your TeX system to add them and\%

```

```

1716      rebuild the format. Now I will use the patterns\\%
1717      preloaded for \bbl@nulllanguage\space instead}}
1718 \let\bbl@usehooks\@gobbletwo
1719 \ifx\bbl@onlyswitch\@empty\endinput\fi
1720 % Here ended switch.def

      Here ended switch.def.

1721 \ifx\directlua\@undefined\else
1722   \ifx\bbl@luapatterns\@undefined
1723     \input luababel.def
1724   \fi
1725 \fi
1726 <<Basic macros>>
1727 \bbl@trace{Compatibility with language.def}
1728 \ifx\bbl@languages\@undefined
1729   \ifx\directlua\@undefined
1730     \openin1 = language.def % TODO. Remove hardcoded number
1731     \ifeof1
1732       \closein1
1733       \message{I couldn't find the file language.def}
1734     \else
1735       \closein1
1736       \begingroup
1737         \def\addlanguage#1#2#3#4#5{%
1738           \expandafter\ifx\csname lang@#1\endcsname\relax\else
1739             \global\expandafter\let\csname l@#1\endcsname
1740               \csname lang@#1\endcsname
1741           \fi}%
1742         \def\uselanguage#1{%
1743           \input language.def
1744         \endgroup
1745       \fi
1746     \fi
1747     \chardef\l@english\z@
1748 \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.

```

1749 \def\addto#1#2{%
1750   \ifx#1\@undefined
1751     \def#1{#2}%
1752   \else
1753     \ifx#1\relax
1754       \def#1{#2}%
1755     \else
1756       {\toks@\expandafter{#1#2}%
1757        \xdef#1{\the\toks@}}%
1758     \fi
1759   \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?

```

1760 \def\bbl@withactive#1#2{%
1761   \begingroup
1762     \lccode`~=#2\relax
1763     \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  $\TeX$  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`.

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

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

```
1769 \def\bbl@redefine@long#1{%
1770   \edef\bbl@tempa{\bbl@stripslash#1}%
1771   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
1772   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
1773 \@onlypreamble\bbl@redefine@long
```

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

```
1774 \def\bbl@redefineroobust#1{%
1775   \edef\bbl@tempa{\bbl@stripslash#1}%
1776   \bbl@ifunset{\bbl@tempa\space}%
1777   {\expandafter\let\csname org@\bbl@tempa\endcsname#1%
1778     \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
1779   {\bbl@exp{\let\<org@\bbl@tempa>\<\bbl@tempa\space>}}}%
1780   \@namedef{\bbl@tempa\space}}
1781 \@onlypreamble\bbl@redefineroobust
```

### 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.

```
1782 \bbl@trace{Hooks}
1783 \newcommand\AddBabelHook[3][{}]{%
1784   \bbl@ifunset{bbl@hk@#2}{\EnableBabelHook{#2}}}%
1785   \def\bbl@tempa##1,#3=##2,##3@empty{\def\bbl@tempb{##2}}%
1786   \expandafter\bbl@tempa\bbl@evargs,#3=,\@empty
1787   \bbl@ifunset{bbl@ev@#2@#3@#1}%
1788   {\bbl@csarg\bbl@add{ev@#3@#1}{\bbl@elth{#2}}}%
1789   {\bbl@csarg\let{ev@#2@#3@#1}\relax}%
1790   \bbl@csarg\newcommand{ev@#2@#3@#1}[\bbl@tempb]}
1791 \newcommand\EnableBabelHook[1]{\bbl@csarg\let{hk@#1}\@firstofone}
1792 \newcommand\DisableBabelHook[1]{\bbl@csarg\let{hk@#1}\@gobble}
1793 \def\bbl@usehooks#1#2{%
1794   \def\bbl@elth##1{%
1795     \bbl@cs{hk@##1}{\bbl@cs{ev@##1@#1}#2}}%
1796     \bbl@cs{ev@#1}%
1797     \ifx\language\undefined\else % Test required for Plain (?)
1798       \def\bbl@elth##1{%
1799         \bbl@cs{hk@##1}{\bbl@cl{ev@##1@#1}#2}}%
1800         \bbl@cl{ev@#1}%
1801         \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).

```
1802 \def\bbl@evargs{,% <- don't delete this comma
1803   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
1804   adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
1805   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
1806   hyphenation=2,initiateactive=3,afterreset=0,foreign=0,foreign*=0,%
1807   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.

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

```

1847         \fi
1848         #####1}}}%
1849     {}%
1850     \toks@\expandafter{##1}%
1851     \edef##1{%
1852         \bbl@csarg\noexpand{ensure@\language}%
1853         {\the\toks@}}%
1854     \fi
1855     \expandafter\bbl@tempb
1856     \fi}%
1857 \expandafter\bbl@tempb\bbl@captionslist\today\@empty
1858 \def\bbl@tempa##1{% elt for include list
1859     \ifx##1\@empty\else
1860         \bbl@csarg\in@{ensure@\language\expandafter}\expandafter{##1}%
1861         \ifin\else
1862             \bbl@tempb##1\@empty
1863         \fi
1864         \expandafter\bbl@tempa
1865     \fi}%
1866 \bbl@tempa#1\@empty}
1867 \def\bbl@captionslist{%
1868     \prefacename\refname\abstractname\bibname\chaptername\appendixname
1869     \contentsname\listfigurename\listtablename\indexname\figurename
1870     \tablename\partname\enclname\ccname\headtoname\pagename\seename
1871     \alsoname\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 at-sign. We make sure that it is a ‘letter’ during the processing of the file. We also save its name as the last called option, even if not loaded.

Another character that needs to have the correct category code during processing of language definition files is the equals sign, ‘=’, because it is sometimes used in constructions with the \let primitive. Therefore we store its current catcode and restore it later on.

Now we check whether we should perhaps stop the processing of this file. To do this we first need to check whether the second argument that is passed to \LdfInit is a control sequence. We do that by looking at the first token after passing #2 through string. When it is equal to \backslashchar we are dealing with a control sequence which we can compare with \@undefined.

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

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

Finally we check \originalTeX.

```

1872 \bbl@trace{Macros for setting language files up}
1873 \def\bbl@ldfinit{%
1874     \let\bbl@screset\@empty
1875     \let\BabelStrings\bbl@opt@string
1876     \let\BabelOptions\@empty
1877     \let\BabelLanguages\relax
1878     \ifx\originalTeX\@undefined
1879         \let\originalTeX\@empty
1880     \else
1881         \originalTeX
1882     \fi}
1883 \def\LdfInit#1#2{%

```

```

1884 \chardef\atcatcode=\catcode`\@
1885 \catcode`\@=11\relax
1886 \chardef\eqcatcode=\catcode`\=
1887 \catcode`\==12\relax
1888 \expandafter\if\expandafter\@backslashchar
1889         \expandafter\@car\string#2\@nil
1890 \ifx#2\@undefined\else
1891     \ldf@quit{#1}%
1892 \fi
1893 \else
1894     \expandafter\ifx\csname#2\endcsname\relax\else
1895         \ldf@quit{#1}%
1896 \fi
1897 \fi
1898 \bbl@ldfinit}

```

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

```

1899 \def\ldf@quit#1{%
1900     \expandafter\main@language\expandafter{#1}%
1901     \catcode`\@=\atcatcode \let\atcatcode\relax
1902     \catcode`\==\eqcatcode \let\eqcatcode\relax
1903 \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.

```

1904 \def\bbl@afterldf#1{% TODO. Merge into the next macro? Unused elsewhere
1905     \bbl@afterlang
1906     \let\bbl@afterlang\relax
1907     \let\BabelModifiers\relax
1908     \let\bbl@screset\relax}%
1909 \def\ldf@finish#1{%
1910     \ifx\loadlocalcfg\@undefined\else % For LaTeX 209
1911         \loadlocalcfg{#1}%
1912     \fi
1913     \bbl@afterldf{#1}%
1914     \expandafter\main@language\expandafter{#1}%
1915     \catcode`\@=\atcatcode \let\atcatcode\relax
1916     \catcode`\==\eqcatcode \let\eqcatcode\relax}

```

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

```

1917 \@onlypreamble\LdfInit
1918 \@onlypreamble\ldf@quit
1919 \@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.

```

1920 \def\main@language#1{%
1921     \def\bbl@main@language{#1}%
1922     \let\language\bbl@main@language % TODO. Set localename
1923     \bbl@id@assign
1924     \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`.

```

1925 \def\bbl@beforestart{%
1926   \bbl@usehooks{beforestart}{}%
1927   \global\let\bbl@beforestart\relax}
1928 \AtBeginDocument{%
1929   \@nameuse{bbl@beforestart}%
1930   \if@filesw
1931     \providecommand\babel@aux[2]{}%
1932     \immediate\write\@mainaux{%
1933       \string\providecommand\string\babel@aux[2]{}%
1934       \immediate\write\@mainaux{\string\@nameuse{bbl@beforestart}}}%
1935   \fi
1936   \expandafter\selectlanguage\expandafter{\bbl@main@language}%
1937   \ifbbl@single % must go after the line above.
1938     \renewcommand\selectlanguage[1]{}%
1939     \renewcommand\foreignlanguage[2]{#2}%
1940     \global\let\babel@aux\@gobbletwo % Also as flag
1941   \fi
1942   \ifcase\bbl@engine\or\pagedir\bodydir\fi} % TODO - a better place

```

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

```

1943 \def\select@language@x#1{%
1944   \ifcase\bbl@select@type
1945     \bbl@ifsamestring\language{#1}{\select@language{#1}}%
1946   \else
1947     \select@language{#1}%
1948   \fi}

```

## 9.5 Shorthands

`\bbl@add@special` The macro `\bbl@add@special` is used to add a new character (or single character control sequence) to the macro `\dospecials` (and `\@sanitize` if  $\text{\LaTeX}$  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.

```

1949 \bbl@trace{Shorhands}
1950 \def\bbl@add@special#1{% 1:a macro like "\, \?, etc.
1951   \bbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
1952   \bbl@ifunset{@sanitize}{\bbl@add\@sanitize{\@makeother#1}}%
1953   \ifx\nfss@catcodes\@undefined\else % TODO - same for above
1954     \begingroup
1955       \catcode`#1\active
1956       \nfss@catcodes
1957       \ifnum\catcode`#1=\active
1958         \endgroup
1959       \bbl@add\nfss@catcodes{\@makeother#1}%
1960     \else
1961       \endgroup
1962     \fi
1963   \fi}

```

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

```

1964 \def\bbl@remove@special#1{%
1965   \begingroup
1966     \def\x##1##2{\ifnum`#1=`##2\noexpand\@empty
1967       \else\noexpand##1\noexpand##2\fi}%
1968   \def\do{\x\do}%

```

```

1969 \def\@makeother{\x\@makeother}%
1970 \edef\x{\endgroup
1971 \def\noexpand\dospecials{\dospecials}%
1972 \expandafter\ifx\csname @sanitize\endcsname\relax\else
1973 \def\noexpand\@sanitize{\@sanitize}%
1974 \fi}%
1975 \x}

```

`\initiate@active@char` A language definition file can call this macro to make a character active. This macro takes one argument, the character that is to be made active. When the character was already active this macro does nothing. Otherwise, this macro defines the control sequence `\normal@char⟨char⟩` to expand to the character in its ‘normal state’ and it defines the active character to expand to `\normal@char⟨char⟩` by default (`⟨char⟩` being the character to be made active). Later its definition can be changed to expand to `\active@char⟨char⟩` by calling `\bbl@activate{⟨char⟩}`. For example, to make the double quote character active one could have `\initiate@active@char{"}` in a language definition file. This defines " as `\active@prefix "\active@char"` (where the first " is the character with its original catcode, when the shorthand is created, and `\active@char"` is a single token). In protected contexts, it expands to `\protect "` or `\noexpand "` (ie, with the original "); otherwise `\active@char"` is executed. This macro in turn expands to `\normal@char` in “safe” contexts (eg, `\label`), but `\user@active` in normal “unsafe” ones. The latter search a definition in the user, language and system levels, in this order, but if none is found, `\normal@char` is used. However, a deactivated shorthand (with `\bbl@deactivate` is defined as `\active@prefix "\normal@char"`. The following macro is used to define shorthands in the three levels. It takes 4 arguments: the (string’ed) character, `\<level>@group`, `<level>@active` and `<next-level>@active` (except in system).

```

1976 \def\bbl@active@def#1#2#3#4{%
1977 \namedef{#3#1}{%
1978 \expandafter\ifx\csname#2@sh@#1\endcsname\relax
1979 \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
1980 \else
1981 \bbl@afterfi\csname#2@sh@#1\endcsname
1982 \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.

```

1983 \long\namedef{#3@arg#1}##1{%
1984 \expandafter\ifx\csname#2@sh@#1@string##1\endcsname\relax
1985 \bbl@afterelse\csname#4#1\endcsname##1%
1986 \else
1987 \bbl@afterfi\csname#2@sh@#1@string##1\endcsname
1988 \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.

```

1989 \def\initiate@active@char#1{%
1990 \bbl@ifunset{active@char\string#1}%
1991 {\bbl@withactive
1992 {\expandafter\@initiate@active@char\expandafter}#1\string#1#1}%
1993 {}}

```

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`).

```

1994 \def\@initiate@active@char#1#2#3{%
1995 \bbl@csarg\edef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%
1996 \ifx#1\@undefined
1997 \bbl@csarg\edef{oridef@#2}{\let\noexpand#1\noexpand\@undefined}%
1998 \else

```

```

1999 \bbl@csarg\let{oridef@@#2}#1%
2000 \bbl@csarg\edef{oridef@#2}{%
2001 \let\noexpand#1%
2002 \expandafter\noexpand\csname bbl@oridef@@#2\endcsname}%
2003 \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`).

```

2004 \ifx#1#3\relax
2005 \expandafter\let\csname normal@char#2\endcsname#3%
2006 \else
2007 \bbl@info{Making #2 an active character}%
2008 \ifnum\mathcode`#2=\ifodd\bbl@engine"1000000 \else"8000 \fi
2009 \@namedef{normal@char#2}{%
2010 \textormath{#3}{\csname bbl@oridef@@#2\endcsname}}%
2011 \else
2012 \@namedef{normal@char#2}{#3}%
2013 \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).

```

2014 \bbl@restoreactive{#2}%
2015 \AtBeginDocument{%
2016 \catcode`#2\active
2017 \if@filesw
2018 \immediate\write\@mainaux{\catcode`\string#2\active}%
2019 \fi}%
2020 \expandafter\bbl@add@special\csname#2\endcsname
2021 \catcode`#2\active
2022 \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⟩`).

```

2023 \let\bbl@tempa\@firstoftwo
2024 \if\string^#2%
2025 \def\bbl@tempa{\noexpand\textormath}%
2026 \else
2027 \ifx\bbl@mathnormal\@undefined\else
2028 \let\bbl@tempa\bbl@mathnormal
2029 \fi
2030 \fi
2031 \expandafter\edef\csname active@char#2\endcsname{%
2032 \bbl@tempa
2033 {\noexpand\if@safe@actives
2034 \noexpand\expandafter
2035 \expandafter\noexpand\csname normal@char#2\endcsname
2036 \noexpand\else
2037 \noexpand\expandafter
2038 \expandafter\noexpand\csname bbl@doactive#2\endcsname

```

```

2039      \noexpand\fi}%
2040      {\expandafter\noexpand\csname normal@char#2\endcsname}}}%
2041      \bbl@csarg\edef{doactive#2}{%
2042      \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!).

```

2043      \bbl@csarg\edef{active@#2}{%
2044      \noexpand\active@prefix\noexpand#1%
2045      \expandafter\noexpand\csname active@char#2\endcsname}%
2046      \bbl@csarg\edef{normal@#2}{%
2047      \noexpand\active@prefix\noexpand#1%
2048      \expandafter\noexpand\csname normal@char#2\endcsname}%
2049      \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.

```

2050      \bbl@active@def#2\user@group{user@active}{language@active}%
2051      \bbl@active@def#2\language@group{language@active}{system@active}%
2052      \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.

```

2053      \expandafter\edef\csname\user@group @sh#2@@\endcsname
2054      {\expandafter\noexpand\csname normal@char#2\endcsname}%
2055      \expandafter\edef\csname\user@group @sh#2@\string\protect\endcsname
2056      {\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 `\prim@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.

```

2057      \if\string'#2%
2058      \let\prim@s\bbl@prim@s
2059      \let\active@math@prime#1%
2060      \fi
2061      \bbl@usehooks{initiateactive}{\{#1\}{#2\}{#3\}}

```

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

```

2062      <<(*More package options)>> ≡
2063      \DeclareOption{math=active}{}
2064      \DeclareOption{math=normal}{\def\bbl@mathnormal{\noexpand\textormath}}
2065      <</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*.

```

2066      \@ifpackagewith{babel}{KeepShorthandsActive}%
2067      {\let\bbl@restoreactive\@gobble}%
2068      {\def\bbl@restoreactive#1{%
2069      \bbl@exp{%

```



```

2070      \\AfterBabelLanguage\\CurrentOption
2071      {\catcode`#1=\the\catcode`#1\relax}%
2072      \\AtEndOfPackage
2073      {\catcode`#1=\the\catcode`#1\relax}}}%
2074      \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.

```

2075 \def\bbl@sh@select#1#2{%
2076   \expandafter\ifx\csname#1@sh@#2@sel\endcsname\relax
2077     \bbl@afterelse\bbl@scndcs
2078   \else
2079     \bbl@afterfi\csname#1@sh@#2@sel\endcsname
2080   \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.

```

2081 \begingroup
2082 \bbl@ifunset{ifincsname}% TODO. Ugly. Correct?
2083 {\gdef\active@prefix#1{%
2084   \ifx\protect\@typeset@protect
2085     \else
2086       \ifx\protect\@unexpandable@protect
2087         \noexpand#1%
2088       \else
2089         \protect#1%
2090       \fi
2091       \expandafter\@gobble
2092     \fi}}
2093 {\gdef\active@prefix#1{%
2094   \ifincsname
2095     \string#1%
2096     \expandafter\@gobble
2097   \else
2098     \ifx\protect\@typeset@protect
2099     \else
2100       \ifx\protect\@unexpandable@protect
2101         \noexpand#1%
2102       \else
2103         \protect#1%
2104       \fi
2105       \expandafter\expandafter\expandafter\@gobble
2106     \fi
2107   \fi}}
2108 \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⟩`.

```

2109 \newif\if@safe@actives
2110 \@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.

```
2111 \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  
`\bbl@deactivate` 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`.

```
2112 \chardef\bbl@activated\z@
2113 \def\bbl@activate#1{%
2114   \chardef\bbl@activated\ne
2115   \bbl@withactive{\expandafter\let\expandafter}#1%
2116   \csname bbl@active@\string#1\endcsname}
2117 \def\bbl@deactivate#1{%
2118   \chardef\bbl@activated\tw@
2119   \bbl@withactive{\expandafter\let\expandafter}#1%
2120   \csname bbl@normal@\string#1\endcsname}
```

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

```
\bbl@scndcs
2121 \def\bbl@firstcs#1#2{\csname#1\endcsname}
2122 \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.

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

```

2147 \bbl@ifunset{#1@sh@\string#2@\string#3@}{}%
2148 {\def\bbl@tempa{#4}%
2149 \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
2150 \else
2151 \bbl@info
2152 {Redefining #1 shorthand \string#2\string#3\%
2153 in language \CurrentOption}%
2154 \fi}%
2155 \@namedef{#1@sh@\string#2@\string#3@}{#4}%
2156 \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.

2157 \def\textormath{%
2158 \ifmmode
2159 \expandafter\@secondoftwo
2160 \else
2161 \expandafter\@firstoftwo
2162 \fi}

\user@group The current concept of ‘shorthands’ supports three levels or groups of shorthands. For each level the
\language@group name of the level or group is stored in a macro. The default is to have a user group; use language
\system@group group ‘english’ and have a system group called ‘system’.

2163 \def\user@group{user}
2164 \def\language@group{english} % TODO. I don't like defaults
2165 \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.

2166 \def\useshorthands{%
2167 \@ifstar\bbl@usesh@s{\bbl@usesh@x{}}
2168 \def\bbl@usesh@s#1{%
2169 \bbl@usesh@x
2170 {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bbl@activate{#1}}}%
2171 {#1}}
2172 \def\bbl@usesh@x#1#2{%
2173 \bbl@ifshorthand{#2}%
2174 {\def\user@group{user}%
2175 \initiate@active@char{#2}%
2176 #1%
2177 \bbl@activate{#2}}%
2178 {\bbl@error
2179 {I can't declare a shorthand turned off (\string#2)}
2180 {Sorry, but you can't use shorthands which have been\\%
2181 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.

2182 \def\user@language@group{user@\language@group}
2183 \def\bbl@set@user@generic#1#2{%
2184 \bbl@ifunset{user@generic@active#1}%
2185 {\bbl@active@def#1\user@language@group{user@active}{user@generic@active}%
2186 \bbl@active@def#1\user@group{user@generic@active}{language@active}%
2187 \expandafter\edef\csname#2@sh@#1@\endcsname{%

```

```

2188     \expandafter\noexpand\csname normal@char#1\endcsname}%
2189     \expandafter\edef\csname#2sh@#1@\string\protect@\endcsname{%
2190     \expandafter\noexpand\csname user@active#1\endcsname}}%
2191     \@empty}
2192 \newcommand\defineshorthand[3][user]{%
2193     \edef\bbl@tempa{\zap@space#1 \@empty}%
2194     \bbl@for\bbl@tempb\bbl@tempa{%
2195         \if*\expandafter\@car\bbl@tempb\@nil
2196             \edef\bbl@tempb{user@\expandafter\@gobble\bbl@tempb}%
2197             \@expandtwoargs
2198             \bbl@set@user@generic{\expandafter\string\@car#2\@nil}\bbl@tempb
2199         \fi
2200     \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].

```

2201 \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`.

```

2202 \def\aliasshorthand#1#2{%
2203     \bbl@ifshorthand{#2}%
2204     {\expandafter\ifx\csname active@char\string#2\endcsname\relax
2205         \ifx\document\@notprerr
2206             \@notshorthand{#2}%
2207         \else
2208             \initiate@active@char{#2}%
2209             \expandafter\let\csname active@char\string#2\expandafter\endcsname
2210             \csname active@char\string#1\endcsname
2211             \expandafter\let\csname normal@char\string#2\expandafter\endcsname
2212             \csname normal@char\string#1\endcsname
2213             \bbl@activate{#2}%
2214         \fi
2215     \fi}%
2216     {\bbl@error
2217         {Cannot declare a shorthand turned off (\string#2)}
2218         {Sorry, but you cannot use shorthands which have been\\%
2219         turned off in the package options}}}

```

`\@notshorthand`

```

2220 \def\@notshorthand#1{%
2221     \bbl@error{%
2222         The character '\string #1' should be made a shorthand character;\\%
2223         add the command \string\usesshorthands\string{#1\string} to
2224         the preamble.\\%
2225         I will ignore your instruction}%
2226     {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.

```

2227 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
2228 \DeclareRobustCommand*\shorthandoff{%
2229     \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
2230 \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.

```

2231 \def\bbl@switch@sh#1#2{%
2232   \ifx#2\@nnil\else
2233     \bbl@ifunset{bbl@active@\string#2}%
2234     {\bbl@error
2235       {I can't switch '\string#2' on or off--not a shorthand}%
2236       {This character is not a shorthand. Maybe you made\\%
2237         a typing mistake? I will ignore your instruction.}}%
2238     {\ifcase#1    off, on, off*
2239       \catcode`#212\relax
2240     \or
2241       \catcode`#2\active
2242       \bbl@ifunset{bbl@shdef@\string#2}%
2243       {}%
2244       {\bbl@withactive{\expandafter\let\expandafter}#2%
2245         \csname bbl@shdef@\string#2\endcsname
2246         \bbl@csarg\let{shdef@\string#2}\relax}%
2247       \ifcase\bbl@activated\or
2248         \bbl@activate{#2}%
2249       \else
2250         \bbl@deactivate{#2}%
2251       \fi
2252     \or
2253       \bbl@ifunset{bbl@shdef@\string#2}%
2254       {\bbl@withactive{\bbl@csarg\let{shdef@\string#2}}#2}%
2255       {}%
2256       \csname bbl@oricat@\string#2\endcsname
2257       \csname bbl@oridef@\string#2\endcsname
2258       \fi}%
2259   \bbl@afterfi\bbl@switch@sh#1%
2260 \fi}

```

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

```

2261 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
2262 \def\bbl@putsh#1{%
2263   \bbl@ifunset{bbl@active@\string#1}%
2264   {\bbl@putsh@i#1\@empty\@nnil}%
2265   {\csname bbl@active@\string#1\endcsname}}
2266 \def\bbl@putsh@i#1#2\@nnil{%
2267   \csname\language@group @sh@\string#1@%
2268     \ifx\@empty#2\else\string#2@\fi\endcsname}
2269 \ifx\bbl@opt@shorthands\@nnil\else
2270   \let\bbl@s@initiate@active@char\initiate@active@char
2271   \def\initiate@active@char#1{%
2272     \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
2273   \let\bbl@s@switch@sh\bbl@switch@sh
2274   \def\bbl@switch@sh#1#2{%
2275     \ifx#2\@nnil\else
2276       \bbl@afterfi
2277       \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
2278     \fi}
2279   \let\bbl@s@activate\bbl@activate
2280   \def\bbl@activate#1{%
2281     \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}

```

```

2282 \let\bbl@s@deactivate\bbl@deactivate
2283 \def\bbl@deactivate#1{%
2284   \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
2285 \fi

```

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

```

2286 \newcommand\ifbabelshorthand[3]{\bbl@ifunset{bbl@active@string#1}{#3}{#2}}

```

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

```

2287 \def\bbl@prim@s{%
2288   \prime\futurelet\@let@token\bbl@pr@m@s}
2289 \def\bbl@if@primes#1#2{%
2290   \ifx#1\@let@token
2291     \expandafter\@firstoftwo
2292   \else\ifx#2\@let@token
2293     \bbl@afterelse\expandafter\@firstoftwo
2294   \else
2295     \bbl@afterfi\expandafter\@secondoftwo
2296   \fi\fi}
2297 \begingroup
2298   \catcode`\^=7 \catcode`\*=\active \lccode`\*=\^
2299   \catcode`\'=12 \catcode`\"=\active \lccode`\"=\'
2300   \lowercase{%
2301     \gdef\bbl@pr@m@s{%
2302       \bbl@if@primes"%
2303         \pr@@s
2304         {\bbl@if@primes*\pr@@@t\egroup}}}
2305 \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).

```

2306 \initiate@active@char{~}
2307 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
2308 \bbl@activate{~}

```

**\OT1dqpos** 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.

```

2309 \expandafter\def\csname OT1dqpos\endcsname{127}
2310 \expandafter\def\csname T1dqpos\endcsname{4}

```

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

```

2311 \ifx\f@encoding\undefined
2312   \def\f@encoding{OT1}
2313 \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.

```
2314 \bbl@trace{Language attributes}
2315 \newcommand\languageattribute[2]{%
2316   \def\bbl@tempc{#1}%
2317   \bbl@fixname\bbl@tempc
2318   \bbl@iflanguage\bbl@tempc{%
2319     \bbl@vforeach{#2}{%
```

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

```
2320     \ifx\bbl@known@attribs\@undefined
2321       \in@false
2322     \else
2323       \bbl@xin@{,\bbl@tempc-##1,}{,\bbl@known@attribs,}%
2324     \fi
2325     \ifin@
2326       \bbl@warning{%
2327         You have more than once selected the attribute '##1'\%
2328         for language #1. Reported}%
2329     \else
```

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

```
2330       \bbl@exp{%
2331         \\bbl@add@list\\bbl@known@attribs{\bbl@tempc-##1}}%
2332       \edef\bbl@tempa{\bbl@tempc-##1}%
2333       \expandafter\bbl@ifknown@trib\expandafter{\bbl@tempa}\bbl@attributes%
2334       {\csname\bbl@tempc @attr##1\endcsname}%
2335       {\@attrerr{\bbl@tempc}{##1}}%
2336     \fi}}}
2337 \@onlypreamble\languageattribute
```

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

```
2338 \newcommand*{\@attrerr}[2]{%
2339   \bbl@error
2340   {The attribute #2 is unknown for language #1.}%
2341   {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}`.

```
2342 \def\bbl@declare@ttribute#1#2#3{%
2343   \bbl@xin@{,#2,}{,\BabelModifiers,}%
2344   \ifin@
2345     \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
2346   \fi
2347   \bbl@add@list\bbl@attributes{#1-#2}%
2348   \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.

```

2349 \def\bbl@ifattributeset#1#2#3#4{%
2350   \ifx\bbl@known@attribs\@undefined
2351     \in@false
2352   \else
2353     \bbl@xin@{,#1-#2,}{,\bbl@known@attribs,}%
2354   \fi
2355   \ifin@
2356     \bbl@afterelse#3%
2357   \else
2358     \bbl@afterfi#4%
2359   \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.

```

2360 \def\bbl@ifknown@ttrib#1#2{%
2361   \let\bbl@tempa\@secondoftwo
2362   \bbl@loopx\bbl@tempb{#2}{%
2363     \expandafter\in@\expandafter{\expandafter,\bbl@tempb,}{,#1,}%
2364   \ifin@
2365     \let\bbl@tempa\@firstoftwo
2366   \else
2367     \fi}%
2368   \bbl@tempa}

```

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

```

2369 \def\bbl@clear@ttribs{%
2370   \ifx\bbl@attributes\@undefined\else
2371     \bbl@loopx\bbl@tempa{\bbl@attributes}{%
2372       \expandafter\bbl@clear@ttrib\bbl@tempa.
2373     }%
2374     \let\bbl@attributes\@undefined
2375   \fi}
2376 \def\bbl@clear@ttrib#1-#2.{%
2377   \expandafter\let\csname#1@attr@#2\endcsname\@undefined}
2378 \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`

```

2379 \bbl@trace{Macros for saving definitions}
2380 \def\babel@beginsave{\babel@savecnt\z@}

```

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

```

2381 \newcount\babel@savecnt
2382 \babel@beginsave

```



`\babel@save` The macro `\babel@save⟨csname⟩` saves the current meaning of the control sequence `⟨csname⟩` to `\originalTeX`<sup>31</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.

```
2383 \def\babel@save#1{%
2384   \expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax
2385   \toks@\expandafter{\originalTeX\let#1=}%
2386   \bbl@exp{%
2387     \def\originalTeX{\the\toks@\<babel@number\babel@savecnt>\relax}}%
2388   \advance\babel@savecnt\@ne}
2389 \def\babel@savevariable#1{%
2390   \toks@\expandafter{\originalTeX #1}%
2391   \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@frenchspacing` 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.

```
2392 \def\bbl@frenchspacing{%
2393   \ifnum\the\sfcode`\.=\@m
2394     \let\bbl@nonfrenchspacing\relax
2395   \else
2396     \frenchspacing
2397     \let\bbl@nonfrenchspacing\nonfrenchspacing
2398   \fi}
2399 \let\bbl@nonfrenchspacing\nonfrenchspacing
2400 \let\bbl@elt\relax
2401 \edef\bbl@fs@chars{%
2402   \bbl@elt{\string.}\@m{3000}\bbl@elt{\string?}\@m{3000}%
2403   \bbl@elt{\string!}\@m{3000}\bbl@elt{\string:}\@m{2000}%
2404   \bbl@elt{\string;}\@m{1500}\bbl@elt{\string,}\@m{1250}}
```

## 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.

```
2405 \bbl@trace{Short tags}
2406 \def\babeltags#1{%
2407   \edef\bbl@tempa{\zap@space#1 \@empty}%
2408   \def\bbl@tempb##1=##2\@{%
2409     \edef\bbl@tempc{%
2410       \noexpand\newcommand
2411       \expandafter\noexpand\csname ##1\endcsname{%
2412         \noexpand\protect
2413         \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
2414       \noexpand\newcommand
2415       \expandafter\noexpand\csname text##1\endcsname{%
2416         \noexpand\foreignlanguage{##2}}
2417       \bbl@tempc}%
2418   \bbl@for\bbl@tempa\bbl@tempa{%
2419     \expandafter\bbl@tempb\bbl@tempa\@{}}
```

<sup>31</sup>`\originalTeX` has to be expandable, i.e. you shouldn't let it to `\relax`.

## 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.

```

2420 \bbl@trace{Hyphens}
2421 \@onlypreamble\babelhyphenation
2422 \AtEndOfPackage{%
2423   \newcommand\babelhyphenation[2][\@empty]{%
2424     \ifx\bbl@hyphenation@relax
2425       \let\bbl@hyphenation@\@empty
2426     \fi
2427     \ifx\bbl@hyphlist\@empty\else
2428       \bbl@warning{%
2429         You must not intermingle \string\selectlanguage\space and\\%
2430         \string\babelhyphenation\space or some exceptions will not\\%
2431         be taken into account. Reported}%
2432     \fi
2433     \ifx\@empty#1%
2434       \protected@edef\bbl@hyphenation@{\bbl@hyphenation@\space#2}%
2435     \else
2436       \bbl@vforeach{#1}{%
2437         \def\bbl@tempa{##1}%
2438         \bbl@fixname\bbl@tempa
2439         \bbl@iflanguage\bbl@tempa{%
2440           \bbl@csarg\protected@edef{hyphenation@\bbl@tempa}{%
2441             \bbl@ifunset{bbl@hyphenation@\bbl@tempa}%
2442             {}%
2443             {\csname bbl@hyphenation@\bbl@tempa\endcsname\space}%
2444             #2}}}%
2445     \fi}}

```

`\bbl@allowhyphens` This macro makes hyphenation possible. Basically its definition is nothing more than `\nobreak` `\hskip 0pt` plus `Opt`<sup>32</sup>.

```

2446 \def\bbl@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}
2447 \def\bbl@t@one{T1}
2448 \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`.

```

2449 \newcommand\babellnullhyphen{\char\hyphenchar\font}
2450 \def\babelhyphen{\active@prefix\babelhyphen\bbl@hyphen}
2451 \def\bbl@hyphen{%
2452   \@ifstar{\bbl@hyphen@i @}{\bbl@hyphen@i \@empty}}
2453 \def\bbl@hyphen@i#1#2{%
2454   \bbl@ifunset{bbl@hy@#1#2\@empty}%
2455   {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%
2456   {\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.

<sup>32</sup> $\TeX$  begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.

```

2457 \def\bbl@usehyphen#1{%
2458   \leavevmode
2459   \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
2460   \nobreak\hskip\z@skip}
2461 \def\bbl@usehyphen#1{%
2462   \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

The following macro inserts the hyphen char.

```

2463 \def\bbl@hyphenchar{%
2464   \ifnum\hyphenchar\font=\m@ne
2465     \babe\nullhyphen
2466   \else
2467     \char\hyphenchar\font
2468   \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.

```

2469 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
2470 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
2471 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
2472 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
2473 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}
2474 \def\bbl@hy@nobreak{\mbox{\bbl@hyphenchar}}
2475 \def\bbl@hy@repeat{%
2476   \bbl@usehyphen{%
2477     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}
2478 \def\bbl@hy@repeat{%
2479   \bbl@usehyphen{%
2480     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}
2481 \def\bbl@hy@empty{\hskip\z@skip}
2482 \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.

```

2483 \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.

```

2484 \bbl@trace{Multiencoding strings}
2485 \def\bbl@tglobal#1{\global\let#1#1}
2486 \def\bbl@recatcode#1{% TODO. Used only once?
2487   \@tempcnta="7F
2488   \def\bbl@tempa{%
2489     \ifnum\@tempcnta>"FF\else
2490       \catcode\@tempcnta=#1\relax
2491       \advance\@tempcnta\@ne
2492       \expandafter\bbl@tempa
2493     \fi}%
2494   \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).

```
2495 \@ifpackagewith{babel}{nocase}%
2496   {\let\bbl@patchuclc\relax}%
2497   {\def\bbl@patchuclc{%
2498     \global\let\bbl@patchuclc\relax
2499     \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}%
2500     \gdef\bbl@uclc##1{%
2501       \let\bbl@encoded\bbl@encoded@uclc
2502       \bbl@ifunset{\language @bbl@uclc}% and resumes it
2503       {##1}%
2504       {\let\bbl@tempa##1\relax % Used by LANG@bbl@uclc
2505         \csname\language @bbl@uclc\endcsname}%
2506       {\bbl@tolower\@empty}{\bbl@toupper\@empty}}}%
2507     \gdef\bbl@tolower{\csname\language @bbl@lc\endcsname}%
2508     \gdef\bbl@toupper{\csname\language @bbl@uc\endcsname}}%
2509 <<(*More package options)>> ≡
2510 \DeclareOption{nocase}{}
2511 <</More package options>>
```

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

```
2512 <<(*More package options)>> ≡
2513 \let\bbl@opt@strings\@nnil % accept strings=value
2514 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
2515 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
2516 \def\BabelStringsDefault{generic}
2517 <</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.

```
2518 \@onlypreamble\StartBabelCommands
2519 \def\StartBabelCommands{%
2520   \begingroup
2521   \bbl@recatcode{11}%
2522   <<Macros local to BabelCommands>>
2523   \def\bbl@provstring##1##2{%
2524     \providecommand##1{##2}%
2525     \bbl@tglobal##1}%
2526   \global\let\bbl@scafter\@empty
2527   \let\StartBabelCommands\bbl@startcmds
2528   \ifx\BabelLanguages\relax
2529     \let\BabelLanguages\CurrentOption
2530   \fi
2531   \begingroup
2532   \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
2533   \StartBabelCommands}
2534 \def\bbl@startcmds{%
2535   \ifx\bbl@screset\@nnil\else
2536     \bbl@usehooks{stopcommands}{}%
2537   \fi
```

```

2538 \endgroup
2539 \begingroup
2540 \@ifstar
2541 {\ifx\bbbl@opt@strings\@nnil
2542 \let\bbbl@opt@strings\BabelStringsDefault
2543 \fi
2544 \bbbl@startcmds@i}%
2545 \bbbl@startcmds@i}
2546 \def\bbbl@startcmds@i#1#2{%
2547 \edef\bbbl@L{\zap@space#1 \@empty}%
2548 \edef\bbbl@G{\zap@space#2 \@empty}%
2549 \bbbl@startcmds@ii}
2550 \let\bbbl@startcommands\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.

```

2551 \newcommand\bbbl@startcmds@ii[1][\@empty]{%
2552 \let\SetString@gobbletwo
2553 \let\bbbl@stringdef@gobbletwo
2554 \let\AfterBabelCommands@gobble
2555 \ifx\@empty#1%
2556 \def\bbbl@sc@label{generic}%
2557 \def\bbbl@encstring##1##2{%
2558 \ProvideTextCommandDefault##1{##2}%
2559 \bbbl@tglobal##1%
2560 \expandafter\bbbl@tglobal\csname\string?\string##1\endcsname}%
2561 \let\bbbl@sctest\in@true
2562 \else
2563 \let\bbbl@sc@charset\space % <- zapped below
2564 \let\bbbl@sc@fontenc\space % <- " "
2565 \def\bbbl@tempa##1=##2\@nil{%
2566 \bbbl@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
2567 \bbbl@foreach{label=#1}{\bbbl@tempa##1\@nil}%
2568 \def\bbbl@tempa##1 ##2{% space -> comma
2569 ##1%
2570 \ifx\@empty##2\else\ifx,##1,\else,\fi\bbbl@afterfi\bbbl@tempa##2\fi}%
2571 \edef\bbbl@sc@fontenc{\expandafter\bbbl@tempa\bbbl@sc@fontenc\@empty}%
2572 \edef\bbbl@sc@label{\expandafter\zap@space\bbbl@sc@label\@empty}%
2573 \edef\bbbl@sc@charset{\expandafter\zap@space\bbbl@sc@charset\@empty}%
2574 \def\bbbl@encstring##1##2{%
2575 \bbbl@foreach\bbbl@sc@fontenc{%
2576 \bbbl@ifunset{T@###1}%
2577 }%
2578 {\ProvideTextCommand##1{####1}{##2}%
2579 \bbbl@tglobal##1%
2580 \expandafter
2581 \bbbl@tglobal\csname####1\string##1\endcsname}}}%
2582 \def\bbbl@sctest{%
2583 \bbbl@xin@{\bbbl@opt@strings,}{\bbbl@sc@label,\bbbl@sc@fontenc,}}%
2584 \fi
2585 \ifx\bbbl@opt@strings\@nnil % ie, no strings key -> defaults
2586 \else\ifx\bbbl@opt@strings\relax % ie, strings=encoded

```

```

2587 \let\AfterBabelCommands\bbl@aftercmds
2588 \let\SetString\bbl@setstring
2589 \let\bbl@stringdef\bbl@encstring
2590 \else % ie, strings=value
2591 \bbl@sctest
2592 \ifin@
2593 \let\AfterBabelCommands\bbl@aftercmds
2594 \let\SetString\bbl@setstring
2595 \let\bbl@stringdef\bbl@provstring
2596 \fi\fi\fi
2597 \bbl@scswitch
2598 \ifx\bbl@G\@empty
2599 \def\SetString##1##2{%
2600 \bbl@error{Missing group for string \string##1}%
2601 {You must assign strings to some category, typically\\%
2602 captions or extras, but you set none}}%
2603 \fi
2604 \ifx\@empty#1%
2605 \bbl@usehooks{defaultcommands}{}%
2606 \else
2607 \@expandtwoargs
2608 \bbl@usehooks{encodedcommands}{\bbl@sc@charset}{\bbl@sc@fontenc}}%
2609 \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).

```

2610 \def\bbl@forlang#1#2{%
2611 \bbl@for#1\bbl@L{%
2612 \bbl@xin@{, #1,}{, \BabelLanguages,}%
2613 \ifin@#2\relax\fi}}
2614 \def\bbl@scswitch{%
2615 \bbl@forlang\bbl@tempa{%
2616 \ifx\bbl@G\@empty\else
2617 \ifx\SetString\@gobbles\else
2618 \edef\bbl@GL{\bbl@G\bbl@tempa}%
2619 \bbl@xin@{, \bbl@GL,}{, \bbl@screset,}%
2620 \ifin@\else
2621 \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
2622 \xdef\bbl@screset{\bbl@screset, \bbl@GL}%
2623 \fi
2624 \fi
2625 \fi}}
2626 \AtEndOfPackage{%
2627 \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{}{#2}}}%
2628 \let\bbl@scswitch\relax}
2629 \@onlypreamble\EndBabelCommands
2630 \def\EndBabelCommands{%
2631 \bbl@usehooks{stopcommands}{}%
2632 \endgroup
2633 \endgroup
2634 \bbl@scafter}
2635 \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.

```

2636 \def\bbl@setstring#1#2{% eg, \prefacename{<string>}
2637   \bbl@forlang\bbl@tempa{%
2638     \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
2639     \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
2640       {\bbl@exp{%
2641         \global\bbbl@add<\bbl@G\bbl@tempa>{\bbbl@scset\#1<\bbl@LC>}}}%
2642       }%
2643     \def\BabelString{#2}%
2644     \bbl@usehooks{stringprocess}{}%
2645     \expandafter\bbl@stringdef
2646     \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`.

```

2647 \ifx\bbl@opt@strings\relax
2648   \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
2649   \bbl@patchuclc
2650   \let\bbl@encoded\relax
2651   \def\bbl@encoded@uclc#1{%
2652     \@inmathwarn#1%
2653     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
2654       \expandafter\ifx\csname ?\string#1\endcsname\relax
2655         \TextSymbolUnavailable#1%
2656       \else
2657         \csname ?\string#1\endcsname
2658       \fi
2659     \else
2660       \csname\cf@encoding\string#1\endcsname
2661     \fi}
2662 \else
2663   \def\bbl@scset#1#2{\def#1{#2}}
2664 \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.

```

2665 <<*Macros local to BabelCommands>> ≡
2666 \def\SetStringLoop##1##2{%
2667   \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
2668   \count@\z@
2669   \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
2670     \advance\count@\@ne
2671     \toks@\expandafter{\bbl@tempa}%
2672     \bbl@exp{%
2673       \\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
2674       \count@=\the\count@\relax}}}%
2675 <</Macros local to BabelCommands>>
```

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

```

2676 \def\bbl@aftercmds#1{%
2677   \toks@\expandafter{\bbl@scafter#1}%
2678   \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.

```
2679 <<*Macros local to BabelCommands>> ≡
2680 \newcommand\SetCase[3][]{%
2681   \bbl@patchuclc
2682   \bbl@forlang\bbl@tempa{%
2683     \expandafter\bbl@encstring
2684     \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
2685     \expandafter\bbl@encstring
2686     \csname\bbl@tempa @bbl@uc\endcsname{##2}%
2687     \expandafter\bbl@encstring
2688     \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
2689 <</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.

```
2690 <<*Macros local to BabelCommands>> ≡
2691 \newcommand\SetHyphenMap[1]{%
2692   \bbl@forlang\bbl@tempa{%
2693     \expandafter\bbl@stringdef
2694     \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
2695 <</Macros local to BabelCommands>>
```

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

```
2696 \newcommand\BabelLower[2]{% one to one.
2697   \ifnum\lccode#1=#2\else
2698     \babel@savevariable{\lccode#1}%
2699     \lccode#1=#2\relax
2700   \fi}
2701 \newcommand\BabelLowerMM[4]{% many-to-many
2702   \@tempcnta=#1\relax
2703   \@tempcntb=#4\relax
2704   \def\bbl@tempa{%
2705     \ifnum\@tempcnta>#2\else
2706       \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
2707       \advance\@tempcnta#3\relax
2708       \advance\@tempcntb#3\relax
2709       \expandafter\bbl@tempa
2710     \fi}%
2711   \bbl@tempa}
2712 \newcommand\BabelLowerM0[4]{% many-to-one
2713   \@tempcnta=#1\relax
2714   \def\bbl@tempa{%
2715     \ifnum\@tempcnta>#2\else
2716       \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
2717       \advance\@tempcnta#3
2718       \expandafter\bbl@tempa
2719     \fi}%
2720   \bbl@tempa}
```

The following package options control the behavior of hyphenation mapping.

```
2721 <<*More package options>> ≡
2722 \DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
2723 \DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\@ne}
2724 \DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
2725 \DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@}
2726 \DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}
```



2727 <</More package options>>

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

```
2728 \AtEndOfPackage{%
2729   \ifx\bbbl@opt@hyphenmap\undefined
2730     \bbbl@xin@{,}{\bbbl@language@opts}%
2731     \chardef\bbbl@opt@hyphenmap\ifin@4\else\@ne\fi
2732   \fi}
```

This sections 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.

```
2733 \newcommand\setlocalecaption{% TODO. Catch typos. What about ensure?
2734   \@ifstar\bbbl@setcaption@s\bbbl@setcaption@x}
2735 \def\bbbl@setcaption@x#1#2#3{% language caption-name string
2736   \bbbl@trim@def\bbbl@tempa{#2}%
2737   \bbbl@xin@{.template}{\bbbl@tempa}%
2738   \ifin@
2739     \bbbl@ini@captions@template{#3}{#1}%
2740   \else
2741     \edef\bbbl@tempd{%
2742       \expandafter\expandafter\expandafter
2743       \strip@prefix\expandafter\meaning\csname captions#1\endcsname}%
2744     \bbbl@xin@
2745       {\expandafter\string\csname #2name\endcsname}%
2746       {\bbbl@tempd}%
2747     \ifin@ % Renew caption
2748       \bbbl@xin@{\string\bbbl@scset}{\bbbl@tempd}%
2749     \ifin@
2750       \bbbl@exp{%
2751         \\bbbl@ifsamestring{\bbbl@tempa}{\language}%
2752         {\bbbl@scset\<#2name>\<#1#2name>}%
2753         {}}%
2754       \else % Old way converts to new way
2755         \bbbl@ifunset{#1#2name}%
2756         {\bbbl@exp{%
2757           \\bbbl@add\<captions#1>{\def\<#2name>{\<#1#2name>}}%
2758           \\bbbl@ifsamestring{\bbbl@tempa}{\language}%
2759           {\def\<#2name>{\<#1#2name>}}%
2760           {}}%
2761         {}}%
2762       \fi
2763     \else
2764       \bbbl@xin@{\string\bbbl@scset}{\bbbl@tempd}% New
2765       \ifin@ % New way
2766         \bbbl@exp{%
2767           \\bbbl@add\<captions#1>{\bbbl@scset\<#2name>\<#1#2name>}%
2768           \\bbbl@ifsamestring{\bbbl@tempa}{\language}%
2769           {\bbbl@scset\<#2name>\<#1#2name>}%
2770           {}}%
2771       \else % Old way, but defined in the new way
2772         \bbbl@exp{%
2773           \\bbbl@add\<captions#1>{\def\<#2name>{\<#1#2name>}}%
2774           \\bbbl@ifsamestring{\bbbl@tempa}{\language}%
2775           {\def\<#2name>{\<#1#2name>}}%
2776           {}}%
2777         \fi%
2778       \fi
2779     \@namedef{#1#2name}{#3}%
```

```

2780 \toks@\expandafter{\bbl@captionslist}%
2781 \bbl@exp{\in@{<#2name>}{\the\toks@}}%
2782 \ifin@else
2783 \bbl@exp{\bbl@add\bbl@captionslist{<#2name>}}%
2784 \bbl@tglobal\bbl@captionslist
2785 \fi
2786 \fi}
2787 % \def\bbl@setcaption@s#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.

```

2788 \bbl@trace{Macros related to glyphs}
2789 \def\set@low@box#1{\setbox\tw@ \hbox{,}\setbox\z@ \hbox{#1}%
2790 \dimen\z@ \ht\z@ \advance\dimen\z@ -\ht\tw@%
2791 \setbox\z@ \hbox{\lower\dimen\z@ \box\z@}\ht\z@ \ht\tw@ \dp\z@ \dp\tw@}

```

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

```

2792 \def\save@sf@q#1{\leavevmode
2793 \begingroup
2794 \edef\SF{\spacefactor\the\spacefactor}#1\SF
2795 \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.

```

2796 \ProvideTextCommand{\quotedblbase}{OT1}{%
2797 \save@sf@q{\set@low@box{\textquotedblright\}}%
2798 \box\z@\kern-.04em\bbl@allowhyphens}}

```

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

```

2799 \ProvideTextCommandDefault{\quotedblbase}{%
2800 \UseTextSymbol{OT1}{\quotedblbase}}

```

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

```

2801 \ProvideTextCommand{\quotesinglbase}{OT1}{%
2802 \save@sf@q{\set@low@box{\textquoteright\}}%
2803 \box\z@\kern-.04em\bbl@allowhyphens}}

```

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

```

2804 \ProvideTextCommandDefault{\quotesinglbase}{%
2805 \UseTextSymbol{OT1}{\quotesinglbase}}

```

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

```

2806 \ProvideTextCommand{\guillemetleft}{OT1}{%
2807 \ifmmode
2808 \ll
2809 \else

```

```

2810 \save@sf@q{\nobreak
2811 \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bbl@allowhyphens}%
2812 \fi}
2813 \ProvideTextCommand{\guillemetright}{OT1}{%
2814 \ifmmode
2815 \gg
2816 \else
2817 \save@sf@q{\nobreak
2818 \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bbl@allowhyphens}%
2819 \fi}
2820 \ProvideTextCommand{\guillemotleft}{OT1}{%
2821 \ifmmode
2822 \ll
2823 \else
2824 \save@sf@q{\nobreak
2825 \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bbl@allowhyphens}%
2826 \fi}
2827 \ProvideTextCommand{\guillemotright}{OT1}{%
2828 \ifmmode
2829 \gg
2830 \else
2831 \save@sf@q{\nobreak
2832 \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bbl@allowhyphens}%
2833 \fi}

```

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

```

2834 \ProvideTextCommandDefault{\guillemetleft}{%
2835 \UseTextSymbol{OT1}{\guillemetleft}}
2836 \ProvideTextCommandDefault{\guillemetright}{%
2837 \UseTextSymbol{OT1}{\guillemetright}}
2838 \ProvideTextCommandDefault{\guillemotleft}{%
2839 \UseTextSymbol{OT1}{\guillemotleft}}
2840 \ProvideTextCommandDefault{\guillemotright}{%
2841 \UseTextSymbol{OT1}{\guillemotright}}

```

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

```

2842 \ProvideTextCommand{\guilsinglleft}{OT1}{%
2843 \ifmmode
2844 <%
2845 \else
2846 \save@sf@q{\nobreak
2847 \raise.2ex\hbox{$\scriptscriptstyle<$}\bbl@allowhyphens}%
2848 \fi}
2849 \ProvideTextCommand{\guilsinglright}{OT1}{%
2850 \ifmmode
2851 >%
2852 \else
2853 \save@sf@q{\nobreak
2854 \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
2855 \fi}

```

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

```

2856 \ProvideTextCommandDefault{\guilsinglleft}{%
2857 \UseTextSymbol{OT1}{\guilsinglleft}}
2858 \ProvideTextCommandDefault{\guilsinglright}{%
2859 \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  
`\IJ` fonts. Therefore we fake it for the OT1 encoding.

```
2860 \DeclareTextCommand{\ij}{OT1}{%
2861   i\kern-0.02em\bbl@allowhyphens j}
2862 \DeclareTextCommand{\IJ}{OT1}{%
2863   I\kern-0.02em\bbl@allowhyphens J}
2864 \DeclareTextCommand{\ij}{T1}{\char188}
2865 \DeclareTextCommand{\IJ}{T1}{\char156}
```

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

```
2866 \ProvideTextCommandDefault{\ij}{%
2867   \UseTextSymbol{OT1}{\ij}}
2868 \ProvideTextCommandDefault{\IJ}{%
2869   \UseTextSymbol{OT1}{\IJ}}
```

`\dj` The croatian language needs the letters `\dj` and `\DJ`; they are available in the T1 encoding, but not in  
`\DJ` 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).

```
2870 \def\crrtic@{\hrule height0.1ex width0.3em}
2871 \def\crttic@{\hrule height0.1ex width0.33em}
2872 \def\ddj@{%
2873   \setbox0\hbox{d}\dimen@=\ht0
2874   \advance\dimen@1ex
2875   \dimen@.45\dimen@
2876   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
2877   \advance\dimen@ii.5ex
2878   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
2879 \def\DDJ@{%
2880   \setbox0\hbox{D}\dimen@=.55\ht0
2881   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
2882   \advance\dimen@ii.15ex % correction for the dash position
2883   \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
2884   \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
2885   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
2886 %
2887 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
2888 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}
```

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

```
2889 \ProvideTextCommandDefault{\dj}{%
2890   \UseTextSymbol{OT1}{\dj}}
2891 \ProvideTextCommandDefault{\DJ}{%
2892   \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.

```
2893 \DeclareTextCommand{\SS}{OT1}{SS}
2894 \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 2895 \ProvideTextCommandDefault{\glq}{%
2896 \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}

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

2897 \ProvideTextCommand{\grq}{T1}{%
2898 \textormath{\kern\z@\textquoteleft}{\mbox{\textquoteleft}}}
2899 \ProvideTextCommand{\grq}{TU}{%
2900 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
2901 \ProvideTextCommand{\grq}{OT1}{%
2902 \save@sf@q{\kern-.0125em
2903 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
2904 \kern.07em\relax}}
2905 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}
```

`\glqq` The ‘german’ double quotes.

```
\grqq 2906 \ProvideTextCommandDefault{\glqq}{%
2907 \textormath{\quotedblbase}{\mbox{\quotedblbase}}}

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

2908 \ProvideTextCommand{\grqq}{T1}{%
2909 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
2910 \ProvideTextCommand{\grqq}{TU}{%
2911 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
2912 \ProvideTextCommand{\grqq}{OT1}{%
2913 \save@sf@q{\kern-.07em
2914 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
2915 \kern.07em\relax}}
2916 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}
```

`\flq` The ‘french’ single guillemets.

```
\frq 2917 \ProvideTextCommandDefault{\flq}{%
2918 \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
2919 \ProvideTextCommandDefault{\frq}{%
2920 \textormath{\guilsinglright}{\mbox{\guilsinglright}}}
```

`\flqq` The ‘french’ double guillemets.

```
\frqq 2921 \ProvideTextCommandDefault{\flqq}{%
2922 \textormath{\guillemetleft}{\mbox{\guillemetleft}}}
2923 \ProvideTextCommandDefault{\frqq}{%
2924 \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  
`\umlautlow` default will be `\umlauthigh` (the normal positioning).

```
2925 \def\umlauthigh{%
2926 \def\bbl@umlauta##1{\leavevmode\bgroup%
2927 \expandafter\accent\csname\fontencoding dpos\endcsname
2928 ##1\bbl@allowhyphens\egroup}%
2929 \let\bbl@umlaute\bbl@umlauta}
2930 \def\umlautlow{%
2931 \def\bbl@umlauta{\protect\lower@umlaut}}
```

```

2932 \def\umlaute\lower{%
2933   \def\bbl@umlaute{\protect\lower@umlaut}}
2934 \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.

```

2935 \expandafter\ifx\csname U@D\endcsname\relax
2936   \csname newdimen\endcsname\U@D
2937 \fi

```

The following code fools T<sub>E</sub>X's `make_accent` procedure about the current x-height of the font to force another placement of the umlaut character. First we have to save the current x-height of the font, because we'll change this font dimension and this is always done globally. Then we compute the new x-height in such a way that the umlaut character is lowered to the base character. The value of `.45ex` depends on the METAFONT parameters with which the fonts were built. (Just try out, which value will look best.) If the new x-height is too low, it is not changed. Finally we call the `\accent` primitive, reset the old x-height and insert the base character in the argument.

```

2938 \def\lower@umlaut#1{%
2939   \leavevmode\bggroup
2940     \U@D 1ex%
2941     {\setbox\z@\hbox{%
2942       \expandafter\char\csname\fontencoding dqpos\endcsname}%
2943       \dimen@ -.45ex\advance\dimen@\ht\z@
2944       \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
2945     \expandafter\accent\csname\fontencoding dqpos\endcsname
2946     \fontdimen5\font\U@D #1%
2947   \egroup}

```

For all vowels we declare `\` to be a composite command which uses `\bbl@umlaut` 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@umlaut` and/or `\bbl@umlaute` for a language in the corresponding `ldf` (using the `babel` switching mechanism, of course).

```

2948 \AtBeginDocument{%
2949   \DeclareTextCompositeCommand{\}{OT1}{a}{\bbl@umlaut{a}}%
2950   \DeclareTextCompositeCommand{\}{OT1}{e}{\bbl@umlaute{e}}%
2951   \DeclareTextCompositeCommand{\}{OT1}{i}{\bbl@umlaute{i}}%
2952   \DeclareTextCompositeCommand{\}{OT1}{\i}{\bbl@umlaute{i}}%
2953   \DeclareTextCompositeCommand{\}{OT1}{o}{\bbl@umlaut{o}}%
2954   \DeclareTextCompositeCommand{\}{OT1}{u}{\bbl@umlaut{u}}%
2955   \DeclareTextCompositeCommand{\}{OT1}{A}{\bbl@umlaut{A}}%
2956   \DeclareTextCompositeCommand{\}{OT1}{E}{\bbl@umlaute{E}}%
2957   \DeclareTextCompositeCommand{\}{OT1}{I}{\bbl@umlaute{I}}%
2958   \DeclareTextCompositeCommand{\}{OT1}{O}{\bbl@umlaut{O}}%
2959   \DeclareTextCompositeCommand{\}{OT1}{U}{\bbl@umlaut{U}}%

```

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

```

2960 \ifx\l@english\undefined
2961   \chardef\l@english\z@
2962 \fi
2963 % The following is used to cancel rules in ini files (see Amharic).
2964 \ifx\l@unhyphenated\undefined
2965   \newlanguage\l@unhyphenated
2966 \fi

```

## 9.13 Layout

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

```
2967 \bbl@trace{Bidi layout}
2968 \providecommand\IfBabelLayout[3]{#3}%
2969 \newcommand\BabelPatchSection[1]{%
2970   \@ifundefined{#1}{}{%
2971     \bbl@exp{\let\<bbl@ss@#1>\<#1>}%
2972     \@namedef{#1}{%
2973       \ifstar{\bbl@presec@s{#1}}%
2974       {\@dblarg{\bbl@presec@x{#1}}}}}%
2975 \def\bbl@presec@x#1[#2]#3{%
2976   \bbl@exp{%
2977     \\\select@language@x{\bbl@main@language}%
2978     \\\bbl@cs{sspre@#1}%
2979     \\\bbl@cs{ss@#1}%
2980     [\\foreignlanguage{\language}{\unexpanded{#2}}]%
2981     {\\foreignlanguage{\language}{\unexpanded{#3}}}%
2982     \\\select@language@x{\language}}}%
2983 \def\bbl@presec@s#1#2{%
2984   \bbl@exp{%
2985     \\\select@language@x{\bbl@main@language}%
2986     \\\bbl@cs{sspre@#1}%
2987     \\\bbl@cs{ss@#1}*%
2988     {\\foreignlanguage{\language}{\unexpanded{#2}}}%
2989     \\\select@language@x{\language}}}%
2990 \IfBabelLayout{sectioning}%
2991   {\BabelPatchSection{part}%
2992    \BabelPatchSection{chapter}%
2993    \BabelPatchSection{section}%
2994    \BabelPatchSection{subsection}%
2995    \BabelPatchSection{subsubsection}%
2996    \BabelPatchSection{paragraph}%
2997    \BabelPatchSection{subparagraph}%
2998    \def\babel@toc#1{%
2999      \select@language@x{\bbl@main@language}}}%
3000 \IfBabelLayout{captions}%
3001   {\BabelPatchSection{caption}}}
```

## 9.14 Load engine specific macros

```
3002 \bbl@trace{Input engine specific macros}
3003 \ifcase\bbl@engine
3004   \input txtbabel.def
3005 \or
3006   \input luababel.def
3007 \or
3008   \input xebabel.def
3009 \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.

```
3010 \bbl@trace{Creating languages and reading ini files}
3011 \newcommand\babelprovide[2][{}]{%
3012   \let\bbl@savelangname\language
3013   \edef\bbl@savlocaleid{\the\localeid}%
```

```

3014 % Set name and locale id
3015 \edef\languagename{#2}%
3016 % \global\@namedef{bbl@lcname@#2}{#2}%
3017 \bbl@id@assign
3018 \let\bbl@KVP@captions\@nil
3019 \let\bbl@KVP@date\@nil
3020 \let\bbl@KVP@import\@nil
3021 \let\bbl@KVP@main\@nil
3022 \let\bbl@KVP@script\@nil
3023 \let\bbl@KVP@language\@nil
3024 \let\bbl@KVP@hyphenrules\@nil
3025 \let\bbl@KVP@linebreaking\@nil
3026 \let\bbl@KVP@justification\@nil
3027 \let\bbl@KVP@mapfont\@nil
3028 \let\bbl@KVP@maparabic\@nil
3029 \let\bbl@KVP@mapdigits\@nil
3030 \let\bbl@KVP@intraspace\@nil
3031 \let\bbl@KVP@intrapenalty\@nil
3032 \let\bbl@KVP@onchar\@nil
3033 \let\bbl@KVP@transforms\@nil
3034 \global\let\bbl@release@transforms\@empty
3035 \let\bbl@KVP@alph\@nil
3036 \let\bbl@KVP@Alph\@nil
3037 \let\bbl@KVP@labels\@nil
3038 \bbl@csarg\let{KVP@labels*}\@nil
3039 \global\let\bbl@inidata\@empty
3040 \bbl@forkv{#1}{% TODO - error handling
3041   \in@{/}{##1}%
3042   \ifin@
3043     \bbl@renewinikey##1\@{##2}%
3044   \else
3045     \bbl@csarg\def{KVP@##1}{##2}%
3046   \fi}%
3047 % == init ==
3048 \ifx\bbl@screset\@undefined
3049   \bbl@ldfinit
3050 \fi
3051 % ==
3052 \let\bbl@lbkflag\relax % \@empty = do setup linebreak
3053 \bbl@ifunset{date#2}%
3054   {\let\bbl@lbkflag\@empty}% new
3055   {\ifx\bbl@KVP@hyphenrules\@nil\else
3056     \let\bbl@lbkflag\@empty
3057     \fi
3058     \ifx\bbl@KVP@import\@nil\else
3059       \let\bbl@lbkflag\@empty
3060     \fi}%
3061 % == import, captions ==
3062 \ifx\bbl@KVP@import\@nil\else
3063   \bbl@exp{\bbl@ifblank{\bbl@KVP@import}}%
3064   {\ifx\bbl@initoload\relax
3065     \begingroup
3066       \def\BabelBeforeIni##1##2{\gdef\bbl@KVP@import{##1}\endinput}%
3067       \bbl@input@texini{#2}%
3068     \endgroup
3069   \else
3070     \xdef\bbl@KVP@import{\bbl@initoload}%
3071   \fi}%
3072 {}%

```



```

3073 \fi
3074 \ifx\bb1@KVP@captions\@nil
3075   \let\bb1@KVP@captions\bb1@KVP@import
3076 \fi
3077 % ==
3078 \ifx\bb1@KVP@transforms\@nil\else
3079   \bb1@replace\bb1@KVP@transforms{ }{,}%
3080 \fi
3081 % Load ini
3082 \bb1@ifunset{date#2}%
3083   {\bb1@provide@new{#2}}%
3084   {\bb1@ifblank{#1}%
3085     }% With \bb1@load@basic below
3086   {\bb1@provide@renew{#2}}%
3087 % Post tasks
3088 % -----
3089 % == ensure captions ==
3090 \ifx\bb1@KVP@captions\@nil\else
3091   \bb1@ifunset{bb1@extracaps@#2}%
3092     {\bb1@exp{\\\babelensure[exclude=\\today]{#2}}}%
3093     {\toks@\\expandafter\\expandafter\\expandafter
3094       {\csname bb1@extracaps@#2\endcsname}%
3095       \bb1@exp{\\\babelensure[exclude=\\today,include=\\the\\toks@]{#2}}}%
3096   \bb1@ifunset{bb1@ensure@\\language}%
3097     {\bb1@exp{%
3098       \\DeclareRobustCommand<bb1@ensure@\\language>[1]{%
3099         \\foreignlanguage{\\language}%
3100         {###1}}}%
3101     }%
3102   \bb1@exp{%
3103     \\bb1@tglobal\\<bb1@ensure@\\language>%
3104     \\bb1@tglobal\\<bb1@ensure@\\language\\space>%
3105   \fi
3106 % ==
3107 % At this point all parameters are defined if 'import'. Now we
3108 % execute some code depending on them. But what about if nothing was
3109 % imported? We just set the basic parameters, but still loading the
3110 % whole ini file.
3111 \bb1@load@basic{#2}%
3112 % == script, language ==
3113 % Override the values from ini or defines them
3114 \ifx\bb1@KVP@script\@nil\else
3115   \bb1@csarg\\edef{sname@#2}{\bb1@KVP@script}%
3116 \fi
3117 \ifx\bb1@KVP@language\@nil\else
3118   \bb1@csarg\\edef{lname@#2}{\bb1@KVP@language}%
3119 \fi
3120 % == onchar ==
3121 \ifx\bb1@KVP@onchar\@nil\else
3122   \bb1@luahyphenate
3123   \directlua{
3124     if Babel.locale_mapped == nil then
3125       Babel.locale_mapped = true
3126       Babel.linebreaking.add_before(Babel.locale_map)
3127       Babel.loc_to_scr = {}
3128       Babel.chr_to_loc = Babel.chr_to_loc or {}
3129     end}%
3130   \bb1@xin@{ ids }{ \bb1@KVP@onchar\\space}%
3131 \ifin@

```

```

3132 \ifx\bb1@starthyphens\@undefined % Needed if no explicit selection
3133 \AddBabelHook{babel-onchar}{beforestart}{\bb1@starthyphens}%
3134 \fi
3135 \bb1@exp{\bb1@add\bb1@starthyphens
3136 {\bb1@patterns@lua{\language\language}}}%
3137 % TODO - error/warning if no script
3138 \directlua{
3139   if Babel.script_blocks['\bb1@cl{sbc}'] then
3140     Babel.loc_to_scr[\the\localeid] =
3141       Babel.script_blocks['\bb1@cl{sbc}']
3142     Babel.locale_props[\the\localeid].lc = \the\localeid\space
3143     Babel.locale_props[\the\localeid].lg = \the\@nameuse{1\language}\space
3144   end
3145 }%
3146 \fi
3147 \bb1@xin{\ fonts }{\bb1@KVP@onchar\space}%
3148 \ifin@
3149 \bb1@ifunset{bb1@lsys\language}{\bb1@provide@lsys\language}}}%
3150 \bb1@ifunset{bb1@wdir\language}{\bb1@provide@dirs\language}}}%
3151 \directlua{
3152   if Babel.script_blocks['\bb1@cl{sbc}'] then
3153     Babel.loc_to_scr[\the\localeid] =
3154       Babel.script_blocks['\bb1@cl{sbc}']
3155   end}%
3156 \ifx\bb1@mapselect\@undefined % TODO. almost the same as mapfont
3157 \AtBeginDocument{%
3158   \expandafter\bb1@add\csname selectfont \endcsname{\bb1@mapselect}}%
3159   {\selectfont}}%
3160 \def\bb1@mapselect{%
3161   \let\bb1@mapselect\relax
3162   \edef\bb1@prefontid{\fontid\font}}%
3163 \def\bb1@mapdir##1{%
3164   {\def\language{##1}%
3165     \let\bb1@ifrestoring\@firstoftwo % To avoid font warning
3166     \bb1@switchfont
3167     \directlua{
3168       Babel.locale_props[\the\csname bb1@id@##1\endcsname]
3169         [\bb1@prefontid] = \fontid\font\space}}}%
3170   \fi
3171   \bb1@exp{\bb1@add\bb1@mapselect{\bb1@mapdir\language}}}%
3172 \fi
3173 % TODO - catch non-valid values
3174 \fi
3175 % == mapfont ==
3176 % For bidi texts, to switch the font based on direction
3177 \ifx\bb1@KVP@mapfont\@nil\else
3178 \bb1@ifsamestring{\bb1@KVP@mapfont}{direction}}}%
3179   {\bb1@error{Option '\bb1@KVP@mapfont' unknown for\%
3180     mapfont. Use 'direction'.%
3181     {See the manual for details.}}}%
3182 \bb1@ifunset{bb1@lsys\language}{\bb1@provide@lsys\language}}}%
3183 \bb1@ifunset{bb1@wdir\language}{\bb1@provide@dirs\language}}}%
3184 \ifx\bb1@mapselect\@undefined % TODO. See onchar
3185 \AtBeginDocument{%
3186   \expandafter\bb1@add\csname selectfont \endcsname{\bb1@mapselect}}%
3187   {\selectfont}}%
3188 \def\bb1@mapselect{%
3189   \let\bb1@mapselect\relax
3190   \edef\bb1@prefontid{\fontid\font}}%

```

```

3191 \def\bb@mapdir##1{%
3192   {\def\languagename{##1}%
3193    \let\bb@ifrestoring\@firstoftwo % avoid font warning
3194    \bb@switchfont
3195    \directlua{Babel.fontmap
3196      [\the\csname \bb@wdir@##1\endcsname]%
3197      [\bb@prefontid]=\fontid\font}}}%
3198 \fi
3199 \bb@exp{\bb@add\bb@mapselect{\bb@mapdir{\languagename}}}%
3200 \fi
3201 % == Line breaking: intraspace, intrapenalty ==
3202 % For CJK, East Asian, Southeast Asian, if interspace in ini
3203 \ifx\bb@KVP@intraspace\@nil\else % We can override the ini or set
3204   \bb@csarg\edef{intsp@#2}{\bb@KVP@intraspace}%
3205 \fi
3206 \bb@provide@intraspace
3207 %
3208 \ifx\bb@KVP@justification\@nil\else
3209   \let\bb@KVP@linebreaking\bb@KVP@justification
3210 \fi
3211 \ifx\bb@KVP@linebreaking\@nil\else
3212   \bb@xin@{,\bb@KVP@linebreaking,}{,elongated,kashida,cjk,unhyphenated,}%
3213   \ifin@
3214     \bb@csarg\xdef
3215       {\lnbrk@\languagename}{\expandafter\@car\bb@KVP@linebreaking\@nil}%
3216   \fi
3217 \fi
3218 \bb@xin@{/e}{/\bb@cl{\lnbrk}}%
3219 \ifin@else\bb@xin@{/k}{/\bb@cl{\lnbrk}}\fi
3220 \ifin@\bb@arabicjust\fi
3221 % == Line breaking: hyphenate.other.locale/.script==
3222 \ifx\bb@lbkflag\@empty
3223   \bb@ifunset{\bb@hyotl@\languagename}{}%
3224   {\bb@csarg\bb@replace{hyotl@\languagename}{ }{,}%
3225    \bb@startcommands*\languagename}{}%
3226   \bb@csarg\bb@foreach{hyotl@\languagename}{%
3227     \ifcase\bb@engine
3228       \ifnum##1<257
3229         \SetHyphenMap{\BabelLower{##1}{##1}}%
3230       \fi
3231     \else
3232       \SetHyphenMap{\BabelLower{##1}{##1}}%
3233     \fi}%
3234   \bb@endcommands}%
3235 \bb@ifunset{\bb@hyots@\languagename}{}%
3236 {\bb@csarg\bb@replace{hyots@\languagename}{ }{,}%
3237  \bb@csarg\bb@foreach{hyots@\languagename}{%
3238    \ifcase\bb@engine
3239      \ifnum##1<257
3240        \global\lccode##1=##1\relax
3241      \fi
3242    \else
3243      \global\lccode##1=##1\relax
3244    \fi}}%
3245 \fi
3246 % == Counters: maparabic ==
3247 % Native digits, if provided in ini (TeX level, xe and lua)
3248 \ifcase\bb@engine\else
3249   \bb@ifunset{\bb@dgnat@\languagename}{}%

```

```

3250     {\expandafter\ifx\csname bbl@dgnat@\language\endcsname\@empty\else
3251       \expandafter\expandafter\expandafter
3252       \bbl@setdigits\csname bbl@dgnat@\language\endcsname
3253       \ifx\bbl@KVP@maparabic\@nil\else
3254         \ifx\bbl@latinarabic\@undefined
3255           \expandafter\let\expandafter\@arabic
3256             \csname bbl@counter@\language\endcsname
3257         \else % ie, if layout=counters, which redefines \@arabic
3258           \expandafter\let\expandafter\bbl@latinarabic
3259             \csname bbl@counter@\language\endcsname
3260         \fi
3261       \fi
3262     \fi}%
3263 \fi
3264 % == Counters: mapdigits ==
3265 % Native digits (lua level).
3266 \ifodd\bbl@engine
3267   \ifx\bbl@KVP@mapdigits\@nil\else
3268     \bbl@ifunset{bbl@dgnat@\language}{}%
3269     {\RequirePackage{luatexbase}%
3270     \bbl@activate@preotf
3271     \directlua{
3272       Babel = Babel or {} %%% -> presets in luababel
3273       Babel.digits_mapped = true
3274       Babel.digits = Babel.digits or {}
3275       Babel.digits[\the\localeid] =
3276         table.pack(string.utfvalue('\bbl@cl{dgnat}'))
3277       if not Babel.numbers then
3278         function Babel.numbers(head)
3279           local LOCALE = luatexbase.registernumber'bbl@attr@locale'
3280           local GLYPH = node.id'glyph'
3281           local inmath = false
3282           for item in node.traverse(head) do
3283             if not inmath and item.id == GLYPH then
3284               local temp = node.get_attribute(item, LOCALE)
3285               if Babel.digits[temp] then
3286                 local chr = item.char
3287                 if chr > 47 and chr < 58 then
3288                   item.char = Babel.digits[temp][chr-47]
3289                 end
3290             end
3291             elseif item.id == node.id'math' then
3292               inmath = (item.subtype == 0)
3293             end
3294           end
3295           return head
3296         end
3297       end
3298     } }%
3299   \fi
3300 \fi
3301 % == Counters: alph, Alph ==
3302 % What if extras<lang> contains a \babel@save\@alph? It won't be
3303 % restored correctly when exiting the language, so we ignore
3304 % this change with the \bbl@alph@saved trick.
3305 \ifx\bbl@KVP@alph\@nil\else
3306   \toks@\expandafter\expandafter\expandafter{%
3307     \csname extras\language\endcsname}%
3308   \bbl@exp{%

```

```

3309 \def<extras\language>{%
3310 \let\\bbl@alph@savd\\@alph
3311 \the\toks@
3312 \let\\@alph\\bbl@alph@savd
3313 \\babel@save\\@alph
3314 \let\\@alph<bbl@cntr@\\bbl@KVP@alph @\language>}}%
3315 \fi
3316 \ifx\bbl@KVP@Alph@nil\else
3317 \toks@\\expandafter\\expandafter\\expandafter{%
3318 \csname extras\language\endcsname}%
3319 \bbl@exp{%
3320 \def<extras\language>{%
3321 \let\\bbl@Alph@savd\\@Alph
3322 \the\toks@
3323 \let\\@Alph\\bbl@Alph@savd
3324 \\babel@save\\@Alph
3325 \let\\@Alph<bbl@cntr@\\bbl@KVP@Alph @\language>}}%
3326 \fi
3327 % == require.babel in ini ==
3328 % To load or reload the babel-*.tex, if require.babel in ini
3329 \ifx\bbl@beforestart\relax\else % But not in doc aux or body
3330 \bbl@ifunset{bbl@rqtex@\language}{}%
3331 {\expandafter\ifx\csname bbl@rqtex@\language\endcsname\@empty\else
3332 \let\BabelBeforeIni@gobbletwo
3333 \chardef\atcatcode=\catcode`\@
3334 \catcode`\@=11\relax
3335 \bbl@input@texini{\bbl@cs{rqtex@\language}}%
3336 \catcode`\@=\atcatcode
3337 \let\atcatcode\relax
3338 \fi}%
3339 \fi
3340 % == Release saved transforms ==
3341 \bbl@release@transforms\relax % \relax closes the last item.
3342 % == main ==
3343 \ifx\bbl@KVP@main@nil % Restore only if not 'main'
3344 \let\language\bbl@savelangname
3345 \chardef\localeid\bbl@savelocaleid\relax
3346 \fi}

```

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

```

3347 \def\bbl@provide@new#1{%
3348 \@namedef{date#1}{}% marks lang exists - required by \StartBabelCommands
3349 \@namedef{extras#1}{}%
3350 \@namedef{noextras#1}{}%
3351 \bbl@startcommands*{#1}{captions}%
3352 \ifx\bbl@KVP@captions@nil % and also if import, implicit
3353 \def\bbl@tempb##1{% elt for \bbl@captionslist
3354 \ifx##1\@empty\else
3355 \bbl@exp{%
3356 \\SetString\\##1{%
3357 \\bbl@nocaption{\bbl@stripslash##1}{#1\bbl@stripslash##1}}%
3358 \expandafter\bbl@tempb
3359 \fi}%
3360 \expandafter\bbl@tempb\bbl@captionslist\@empty
3361 \else
3362 \ifx\bbl@initoload\relax
3363 \bbl@read@ini{\bbl@KVP@captions}2% % Here letters cat = 11
3364 \else
3365 \bbl@read@ini{\bbl@initoload}2% % Same

```

```

3366     \fi
3367     \fi
3368     \StartBabelCommands*{#1}{date}%
3369     \ifx\bbbl@KVP@import\@nil
3370         \bbbl@exp{%
3371             \\\SetString\\today{\\bbbl@nocaption{today}{#1today}}}%
3372     \else
3373         \bbbl@savetoday
3374         \bbbl@savedate
3375     \fi
3376     \bbbl@endcommands
3377     \bbbl@load@basic{#1}%
3378     % == hyphenmins == (only if new)
3379     \bbbl@exp{%
3380         \gdef\<#1hyphenmins>{%
3381             {\bbbl@ifunset{bbbl@lfthm@#1}{2}{\bbbl@cs{lfthm@#1}}}%
3382             {\bbbl@ifunset{bbbl@rgthm@#1}{3}{\bbbl@cs{rgthm@#1}}}%
3383     % == hyphenrules ==
3384     \bbbl@provide@hyphens{#1}%
3385     % == frenchspacing == (only if new)
3386     \bbbl@ifunset{bbbl@frspc@#1}{}%
3387     {\edef\bbbl@tempa{\bbbl@cl{frspc}}}%
3388     \edef\bbbl@tempa{\expandafter\@car\bbbl@tempa\@nil}%
3389     \if u\bbbl@tempa      % do nothing
3390     \else\if n\bbbl@tempa  % non french
3391         \expandafter\bbbl@add\csname extras#1\endcsname{%
3392             \let\bbbl@elt\bbbl@fs@elt@i
3393             \bbbl@fs@chars}%
3394     \else\if y\bbbl@tempa  % french
3395         \expandafter\bbbl@add\csname extras#1\endcsname{%
3396             \let\bbbl@elt\bbbl@fs@elt@ii
3397             \bbbl@fs@chars}%
3398     \fi\fi\fi}%
3399     %
3400     \ifx\bbbl@KVP@main\@nil\else
3401         \expandafter\main@language\expandafter{#1}%
3402     \fi}
3403 % A couple of macros used above, to avoid hashes #####...
3404 \def\bbbl@fs@elt@i#1#2#3{%
3405     \ifnum\sfcode`#1=#2\relax
3406         \babel@savevariable{\sfcode`#1}%
3407         \sfcode`#1=#3\relax
3408     \fi}%
3409 \def\bbbl@fs@elt@ii#1#2#3{%
3410     \ifnum\sfcode`#1=#3\relax
3411         \babel@savevariable{\sfcode`#1}%
3412         \sfcode`#1=#2\relax
3413     \fi}%
3414 %
3415 \def\bbbl@provide@renew#1{%
3416     \ifx\bbbl@KVP@captions\@nil\else
3417         \StartBabelCommands*{#1}{captions}%
3418         \bbbl@read@ini{\bbbl@KVP@captions}2%    % Here all letters cat = 11
3419         \EndBabelCommands
3420     \fi
3421     \ifx\bbbl@KVP@import\@nil\else
3422         \StartBabelCommands*{#1}{date}%
3423         \bbbl@savetoday
3424         \bbbl@savedate

```

```

3425 \EndBabelCommands
3426 \fi
3427 % == hyphenrules ==
3428 \ifx\bbbl@lbkflag\@empty
3429 \bbbl@provide@hyphens{#1}%
3430 \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. (TODO. But preserving previous values would be useful.)

```

3431 \def\bbbl@load@basic#1{%
3432 \bbbl@ifunset{bbbl@inidata@\language\language}\relax
3433 {\getlocaleproperty\bbbl@tempa{\language\language}\identification/load.level}%
3434 \ifcase\bbbl@tempa
3435 \bbbl@csarg\let{lname@\language\language}\relax
3436 \fi}%
3437 \bbbl@ifunset{bbbl@lname@#1}%
3438 {\def\BabelBeforeIni##1##2{%
3439 \begingroup
3440 \let\bbbl@ini@captions@aux\@gobbletwo
3441 \def\bbbl@inidate ####1.####2.####3.####4\relax ####5####6}%
3442 \bbbl@read@ini{##1}1%
3443 \ifx\bbbl@initoload\relax\endinput\fi
3444 \endgroup}%
3445 \begingroup % boxed, to avoid extra spaces:
3446 \ifx\bbbl@initoload\relax
3447 \bbbl@input@texini{##1}%
3448 \else
3449 \setbox\z@\hbox{\BabelBeforeIni{\bbbl@initoload}}}%
3450 \fi
3451 \endgroup}%
3452 {}%

```

The hyphenrules option is handled with an auxiliary macro.

```

3453 \def\bbbl@provide@hyphens#1{%
3454 \let\bbbl@tempa\relax
3455 \ifx\bbbl@KVP@hyphenrules\@nil\else
3456 \bbbl@replace\bbbl@KVP@hyphenrules{ }{,}%
3457 \bbbl@foreach\bbbl@KVP@hyphenrules{%
3458 \ifx\bbbl@tempa\relax % if not yet found
3459 \bbbl@ifsamestring{##1}{+}%
3460 {\bbbl@exp{\addlanguage\<l@##1>}}}%
3461 }%
3462 \bbbl@ifunset{l@##1}%
3463 }%
3464 {\bbbl@exp{\let\bbbl@tempa\<l@##1>}}}%
3465 \fi}%
3466 \fi
3467 \ifx\bbbl@tempa\relax % if no opt or no language in opt found
3468 \ifx\bbbl@KVP@import\@nil
3469 \ifx\bbbl@initoload\relax\else
3470 \bbbl@exp{%
3471 \bbbl@ifblank{\bbbl@cs{hyphr@#1}}}%
3472 }%
3473 {\let\bbbl@tempa\<l@bbbl@cl{hyphr}>}}}%
3474 \fi
3475 \else % if importing
3476 \bbbl@exp{%
3477 \bbbl@ifblank{\bbbl@cs{hyphr@#1}}}%

```

```

3478      {}%
3479      {\let\\bbl@tempa<l@bbl@c1{hyphr}>}}%
3480    \fi
3481  \fi
3482  \bbl@ifunset{bbl@tempa}%      ie, relax or undefined
3483    {\bbl@ifunset{l@#1}%      no hyphenrules found - fallback
3484      {\bbl@exp{\\adddialect<l@#1>\language}}%
3485      {}}%      so, l@<lang> is ok - nothing to do
3486    {\bbl@exp{\\adddialect<l@#1>bbl@tempa}}}% found in opt list or ini

```

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

```

3487 \def\bbl@input@texini#1{%
3488   \bbl@bsphack
3489   \bbl@exp{%
3490     \catcode\\=14 \catcode\\=0
3491     \catcode\\{=1 \catcode\\}=2
3492     \lowercase{\\InputIfFileExists{babel-#1.tex}{}}%
3493     \catcode\\=\\the\catcode\\relax
3494     \catcode\\=\\the\catcode\\relax
3495     \catcode\\{=\\the\catcode\\relax
3496     \catcode\\}=\\the\catcode\\relax}%
3497   \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.

```

3498 \def\bbl@inline#1\bbl@inline{%
3499   \@ifnextchar[\bbl@inisect{\@ifnextchar\bbl@iniskip\bbl@inistore}#1\@@% ]
3500 \def\bbl@inisect[#1]#2\@@{\def\bbl@section{#1}}%
3501 \def\bbl@iniskip#1\@@{%      if starts with ;
3502 \def\bbl@inistore#1=#2\@@{%      full (default)
3503   \bbl@trim@def\bbl@tempa{#1}%
3504   \bbl@trim\toks@{#2}%
3505   \bbl@ifunset{bbl@KVP@\bbl@section/\bbl@tempa}%
3506     {\bbl@exp{%
3507       \\g@addto@macro\\bbl@inidata{%
3508         \\bbl@elt{\bbl@section}{\bbl@tempa}{\the\toks@}}}%
3509     {}}%
3510 \def\bbl@inistore@min#1=#2\@@{%      minimal (maybe set in \bbl@read@ini)
3511   \bbl@trim@def\bbl@tempa{#1}%
3512   \bbl@trim\toks@{#2}%
3513   \bbl@xin@{.identification.}{.\bbl@section.}%
3514   \ifin@
3515     \bbl@exp{\\g@addto@macro\\bbl@inidata{%
3516       \\bbl@elt{identification}{\bbl@tempa}{\the\toks@}}}%
3517   \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.

```

3518 \ifx\bbl@readstream\undefined
3519   \csname newread\endcsname\bbl@readstream
3520 \fi
3521 \def\bbl@read@ini#1#2{%
3522   \openin\bbl@readstream=babel-#1.ini
3523   \ifeof\bbl@readstream

```



```

3524 \bbl@error
3525 {There is no ini file for the requested language\\%
3526 (#1). Perhaps you misspelled it or your installation\\%
3527 is not complete.}%
3528 {Fix the name or reinstall babel.}%
3529 \else
3530 % Store ini data in \bbl@inidata
3531 \catcode\ [=12 \catcode\ ]=12 \catcode\ ==12 \catcode\ &=12
3532 \catcode\ ;=12 \catcode\ |=12 \catcode\ %=14 \catcode\ -=12
3533 \bbl@info{Importing
3534 \ifcase#2font and identification \or basic \fi
3535 data for \language\\%
3536 from babel-#1.ini. Reported}%
3537 \ifnum#2=\z@
3538 \global\let\bbl@inidata\@empty
3539 \let\bbl@inistore\bbl@inistore@min % Remember it's local
3540 \fi
3541 \def\bbl@section{identification}%
3542 \bbl@exp{\ \bbl@inistore tag.ini=#1\ \ \ \ @}%
3543 \bbl@inistore load.level=#2\ @\ @
3544 \loop
3545 \if T\ifeof\bbl@readstream F\fi T\relax % Trick, because inside \loop
3546 \endlinechar\m@ne
3547 \read\bbl@readstream to \bbl@line
3548 \endlinechar\^^M
3549 \ifx\bbl@line\@empty\else
3550 \expandafter\bbl@iniline\bbl@line\bbl@iniline
3551 \fi
3552 \repeat
3553 % Process stored data
3554 \bbl@csarg\xdef{lini@\language}{#1}%
3555 \let\bbl@savestrings\@empty
3556 \let\bbl@savetoday\@empty
3557 \let\bbl@savestate\@empty
3558 \def\bbl@elt##1##2##3{%
3559 \def\bbl@section{##1}%
3560 \in@{=date.}{=##1}% Find a better place
3561 \ifin@
3562 \bbl@ini@calendar{##1}%
3563 \fi
3564 \global\bbl@csarg\let{bbl@KVP###1/##2}\relax
3565 \bbl@ifunset{bbl@inikv###1}{}%
3566 {\csname bbl@inikv###1\endcsname{##2}{##3}}}%
3567 \bbl@inidata
3568 % 'Export' data
3569 \bbl@ini@exports{#2}%
3570 \global\bbl@csarg\let{inidata@\language}\bbl@inidata
3571 \global\let\bbl@inidata\@empty
3572 \bbl@exp{\ \bbl@add@list\ \bbl@ini@loaded{\language}}%
3573 \bbl@toglobal\bbl@ini@loaded
3574 \fi}

```

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

```

3575 \def\bbl@ini@calendar#1{%
3576 \lowercase{\def\bbl@tempa{=##1=}}%
3577 \bbl@replace\bbl@tempa{=date.gregorian}{}%
3578 \bbl@replace\bbl@tempa{=date.}{}%
3579 \in@{.licr}{=##1=}%
3580 \ifin@

```

```

3581 \ifcase\bb1@engine
3582   \bb1@replace\bb1@tempa{.licr={}}}%
3583 \else
3584   \let\bb1@tempa\relax
3585 \fi
3586 \fi
3587 \ifx\bb1@tempa\relax\else
3588   \bb1@replace\bb1@tempa{=}{}}}%
3589   \bb1@exp{%
3590     \def\<bb1@inikv@#1>####1####2{%
3591       \\\bb1@inidate####1...\relax{####2}{\bb1@tempa}}}%
3592 \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 \bb1@inistore above).

```

3593 \def\bb1@renewinikey#1/#2\@#3{%
3594   \edef\bb1@tempa{\zap@space #1 \@empty}%   section
3595   \edef\bb1@tempb{\zap@space #2 \@empty}%   key
3596   \bb1@trim\toks@{#3}%                       value
3597   \bb1@exp{%
3598     \global\let\<bb1@KVP@\bb1@tempa/\bb1@tempb>\\\@empty % just a flag
3599     \\\g@addto@macro\bb1@inidata{%
3600       \\\bb1@elt{\bb1@tempa}{\bb1@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.

```

3601 \def\bb1@exportkey#1#2#3{%
3602   \bb1@ifunset{\bb1@kv@#2}%
3603   {\bb1@csarg\gdef{#1@\language}\{#3}}%
3604   {\expandafter\ifx\csname \bb1@kv@#2\endcsname\@empty
3605     \bb1@csarg\gdef{#1@\language}\{#3}}%
3606   \else
3607     \bb1@exp{\global\let\<bb1@#1@\language>\<bb1@kv@#2>}%
3608   \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 \bb1@ini@exports is called always (via \bb1@inise), while \bb1@after@ini must be called explicitly after \bb1@read@ini if necessary.

```

3609 \def\bb1@iniwarning#1{%
3610   \bb1@ifunset{\bb1@kv@identification.warning#1}{}%
3611   {\bb1@warning{%
3612     From babel-\bb1@cs{lini@\language}.ini:\%
3613     \bb1@cs{@kv@identification.warning#1}\%
3614     Reported }}}
3615 %
3616 \let\bb1@release@transforms\@empty
3617 %
3618 \def\bb1@ini@exports#1{%
3619   % Identification always exported
3620   \bb1@iniwarning{%
3621     \ifcase\bb1@engine
3622       \bb1@iniwarning{.pdf\latex}%
3623     \or
3624       \bb1@iniwarning{.lual\latex}%
3625     \or
3626       \bb1@iniwarning{.xel\latex}%
3627     \fi%

```

```

3628 \bbl@exportkey{elname}{identification.name.english}{}%
3629 \bbl@exp{\bbl@exportkey{lname}{identification.name.opentype}%
3630   {\csname bbl@elname@language\endcsname}}%
3631 \bbl@exportkey{tbc}{identification.tag.bcp47}{}%
3632 \bbl@exportkey{lbc}{identification.language.tag.bcp47}{}%
3633 \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
3634 \bbl@exportkey{esname}{identification.script.name}{}%
3635 \bbl@exp{\bbl@exportkey{sname}{identification.script.name.opentype}%
3636   {\csname bbl@esname@language\endcsname}}%
3637 \bbl@exportkey{sbc}{identification.script.tag.bcp47}{}%
3638 \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}%
3639 % Also maps bcp47 -> language
3640 \ifbbl@bcptoname
3641   \bbl@csarg\xdef{bcp@map@bbl@cl{tbc}}{\language}%
3642 \fi
3643 % Conditional
3644 \ifnum#1>\z@ % 0 = only info, 1, 2 = basic, (re)new
3645   \bbl@exportkey{lnbrk}{typography.linebreaking}{h}%
3646   \bbl@exportkey{hyphr}{typography.hyphenrules}{}%
3647   \bbl@exportkey{lftm}{typography.lefthyphenmin}{2}%
3648   \bbl@exportkey{rgtm}{typography.righthyphenmin}{3}%
3649   \bbl@exportkey{prehc}{typography.prehyphenchar}{}%
3650   \bbl@exportkey{hyotl}{typography.hyphenate.other.locale}{}%
3651   \bbl@exportkey{hyots}{typography.hyphenate.other.script}{}%
3652   \bbl@exportkey{intsp}{typography.intraspaces}{}%
3653   \bbl@exportkey{chrng}{characters.ranges}{}%
3654   \bbl@exportkey{dgnat}{numbers.digits.native}{}%
3655   \bbl@exportkey{frspc}{typography.frenchspacing}{u}%
3656   \ifnum#1=\tw@ % only (re)new
3657     \bbl@exportkey{rqtex}{identification.require.babel}{}%
3658     \bbl@tglobal\bbl@savetoday
3659     \bbl@tglobal\bbl@savestate
3660     \bbl@savestrings
3661   \fi
3662 \fi}

```

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

```

3663 \def\bbl@inikv#1#2{%      key=value
3664   \toks@{#2}%             This hides #'s from ini values
3665   \bbl@csarg\xdef{kv@bbl@section.#1}{the\toks@}}

```

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

```

3666 \let\bbl@inikv@identification\bbl@inikv
3667 \let\bbl@inikv@typography\bbl@inikv
3668 \let\bbl@inikv@characters\bbl@inikv
3669 \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’.

```

3670 \def\bbl@inikv@counters#1#2{%
3671   \bbl@ifsamestring{#1}{digits}%
3672   {\bbl@error{The counter name 'digits' is reserved for mapping\%
3673     decimal digits}%
3674     {Use another name.}}%
3675   }%
3676   \def\bbl@tempc{#1}%
3677   \bbl@trim@def{\bbl@tempb*}{#2}%
3678   \in@{.1$}{#1$}%

```

```

3679 \ifin@
3680 \bbl@replace\bbl@tempc{.1}{}%
3681 \bbl@csarg\protected@xdef{cntr\bbl@tempc @\language}\bbl@tempc@{\%
3682 \noexpand\bbl@alphnumeral{\bbl@tempc}}}%
3683 \fi
3684 \in@{.F.}{#1}%
3685 \ifin\else\in@{.S.}{#1}\fi
3686 \ifin@
3687 \bbl@csarg\protected@xdef{cntr#1@\language}\bbl@tempb*}%
3688 \else
3689 \toks@{}% Required by \bbl@buildifcase, which returns \bbl@tempa
3690 \expandafter\bbl@buildifcase\bbl@tempb* \ \ % Space after \
3691 \bbl@csarg{\global\expandafter\let}{cntr#1@\language}\bbl@tempa
3692 \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.

```

3693 \ifcase\bbl@engine
3694 \bbl@csarg\def{inikv@captions.licr}#1#2{%
3695 \bbl@ini@captions@aux{#1}{#2}}
3696 \else
3697 \def\bbl@inikv@captions#1#2{%
3698 \bbl@ini@captions@aux{#1}{#2}}
3699 \fi

```

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

```

3700 \def\bbl@ini@captions@template#1#2{% string language tempa=capt-name
3701 \bbl@replace\bbl@tempa{.template}{}%
3702 \def\bbl@toreplace{#1}{}%
3703 \bbl@replace\bbl@toreplace{[ ]}{\nobreakspace}}%
3704 \bbl@replace\bbl@toreplace{[ ]}{\csname}%
3705 \bbl@replace\bbl@toreplace{[ ]}{\csname the}%
3706 \bbl@replace\bbl@toreplace{[ ]}{name\endcsname}}}%
3707 \bbl@replace\bbl@toreplace{[ ]}{\endcsname}}}%
3708 \bbl@xin@{,\bbl@tempa,},{,chapter,appendix,part,}%
3709 \ifin@
3710 \@nameuse{bbl@patch\bbl@tempa}%
3711 \global\bbl@csarg\let{\bbl@tempa fmt#2}\bbl@toreplace
3712 \fi
3713 \bbl@xin@{,\bbl@tempa,},{,figure,table,}%
3714 \ifin@
3715 \toks@\expandafter{\bbl@toreplace}%
3716 \bbl@exp{\gdef\<fnun@\bbl@tempa>{\the\toks@}}%
3717 \fi}
3718 \def\bbl@ini@captions@aux#1#2{%
3719 \bbl@trim@def\bbl@tempa{#1}%
3720 \bbl@xin@{.template}{\bbl@tempa}%
3721 \ifin@
3722 \bbl@ini@captions@template{#2}\language
3723 \else
3724 \bbl@ifblank{#2}%
3725 {\bbl@exp{%
3726 \toks@{\bbl@nocaption{\bbl@tempa}{\language\bbl@tempa name}}}%
3727 {\bbl@trim\toks@{#2}}}%
3728 \bbl@exp{%
3729 \bbl@add{\bbl@savestrings{%
3730 \SetString\<\bbl@tempa name>{\the\toks@}}}%
3731 \toks@\expandafter{\bbl@captionslist}}%

```

```

3732 \bbl@exp{\in@{\<\bbl@tempa name>}{\the\toks@}}%
3733 \ifin@%else
3734 \bbl@exp{%
3735 \\\bbl@add\<bbl@extracaps@\language name>{\<\bbl@tempa name>}%
3736 \\\bbl@toglobal\<bbl@extracaps@\language name>}%
3737 \fi
3738 \fi}

```

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

```

3739 \def\bbl@list@the{%
3740 part,chapter,section,subsection,subsubsection,paragraph,%
3741 subparagraph,enumi,enumii,enumiii,enumiv,equation,figure,%
3742 table,page,footnote,mpfootnote,mpfn}
3743 \def\bbl@map@cnt#1{% #1:roman,etc, // #2:enumi,etc
3744 \bbl@ifunset{bbl@map@#1@\language name}%
3745 {\@nameuse{#1}}%
3746 {\@nameuse{bbl@map@#1@\language name}}}%
3747 \def\bbl@inikv@labels#1#2{%
3748 \in@{.map}{#1}%
3749 \ifin@
3750 \ifx\bbl@KVP@labels\@nil%else
3751 \bbl@xin@{ map }{ \bbl@KVP@labels\space}%
3752 \ifin@
3753 \def\bbl@tempc{#1}%
3754 \bbl@replace\bbl@tempc{.map}{}%
3755 \in@{,#2},{,arabic,roman,Roman,alpha,Alpha,fnsymbol,}%
3756 \bbl@exp{%
3757 \gdef\<bbl@map@\bbl@tempc @\language name>%
3758 {\ifin@\<#2>%else\\localecounter{#2}\fi}}%
3759 \bbl@foreach\bbl@list@the{%
3760 \bbl@ifunset{the##1}{}%
3761 {\bbl@exp{\let\\bbl@tempd\<the##1>}%
3762 \bbl@exp{%
3763 \\\bbl@sreplace\<the##1>%
3764 {\<\bbl@tempc>{##1}}{\\\bbl@map@cnt{\bbl@tempc}{##1}}}%
3765 \\\bbl@sreplace\<the##1>%
3766 {\<\@empty @\bbl@tempc>\<c##1>{\\\bbl@map@cnt{\bbl@tempc}{##1}}}%
3767 \expandafter\ifx\csname the##1\endcsname\bbl@tempd%else
3768 \toks@\expandafter\expandafter\expandafter{%
3769 \csname the##1\endcsname}%
3770 \expandafter\def\csname the##1\endcsname{\the\toks@}}%
3771 \fi}}%
3772 \fi
3773 \fi
3774 %
3775 \else
3776 %
3777 % The following code is still under study. You can test it and make
3778 % suggestions. Eg, enumerate.2 = ([enumi]).([enumii]). It's
3779 % language dependent.
3780 \in@{enumerate.}{#1}%
3781 \ifin@
3782 \def\bbl@tempa{#1}%
3783 \bbl@replace\bbl@tempa{enumerate.}{}%
3784 \def\bbl@toreplace{#2}%
3785 \bbl@replace\bbl@toreplace{[ ]}{\nobreakspace}}%
3786 \bbl@replace\bbl@toreplace{[ ]}{\csname the}%
3787 \bbl@replace\bbl@toreplace{ ]}{\endcsname}}%
3788 \toks@\expandafter{\bbl@toreplace}%

```

```

3789 \bbl@exp{%
3790 \\\bbl@add\<extras\language>{%
3791 \\\babel@save\<labelenum\romannumeral\bbl@tempa>%
3792 \def\<labelenum\romannumeral\bbl@tempa>{\the\toks@}}%
3793 \\\bbl@toglobal\<extras\language>}%
3794 \fi
3795 \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.

```

3796 \def\bbl@chapttype{chapter}
3797 \ifx\@makechapterhead\undefined
3798 \let\bbl@patchchapter\relax
3799 \else\ifx\thechapter\undefined
3800 \let\bbl@patchchapter\relax
3801 \else\ifx\ps@headings\undefined
3802 \let\bbl@patchchapter\relax
3803 \else
3804 \def\bbl@patchchapter{%
3805 \global\let\bbl@patchchapter\relax
3806 \bbl@add\appendix{\def\bbl@chapttype{appendix}}% Not harmful, I hope
3807 \bbl@toglobal\appendix
3808 \bbl@sreplace\ps@headings
3809 {\@chapapp\ thechapter}%
3810 {\bbl@chapterformat}%
3811 \bbl@toglobal\ps@headings
3812 \bbl@sreplace\chaptermark
3813 {\@chapapp\ thechapter}%
3814 {\bbl@chapterformat}%
3815 \bbl@toglobal\chaptermark
3816 \bbl@sreplace\@makechapterhead
3817 {\@chapapp\space\thechapter}%
3818 {\bbl@chapterformat}%
3819 \bbl@toglobal\@makechapterhead
3820 \gdef\bbl@chapterformat{%
3821 \bbl@ifunset{bbl@bbl@chapttype fmt@language}%
3822 {\@chapapp\space\thechapter}
3823 {\@nameuse{bbl@bbl@chapttype fmt@language}}}}
3824 \let\bbl@patchappendix\bbl@patchchapter
3825 \fi\fi\fi
3826 \ifx\@part\undefined
3827 \let\bbl@patchpart\relax
3828 \else
3829 \def\bbl@patchpart{%
3830 \global\let\bbl@patchpart\relax
3831 \bbl@sreplace\@part
3832 {\partname\nobreakspace\thepart}%
3833 {\bbl@partformat}%
3834 \bbl@toglobal\@part
3835 \gdef\bbl@partformat{%
3836 \bbl@ifunset{bbl@partfmt@language}%
3837 {\partname\nobreakspace\thepart}
3838 {\@nameuse{bbl@partfmt@language}}}}
3839 \fi

```

**Date.** TODO. Document

3840 % Arguments are `_not_` protected.

```

3841 \let\bbl@calendar\@empty
3842 \DeclareRobustCommand\localedate[1][\bbl@localedate{#1}]
3843 \def\bbl@localedate#1#2#3#4{%
3844   \begingroup
3845     \ifx\@empty#1\@empty\else
3846       \let\bbl@ld@calendar\@empty
3847       \let\bbl@ld@variant\@empty
3848       \edef\bbl@tempa{\zap@space#1 \@empty}%
3849       \def\bbl@tempb##1=##2\@{\@namedef\bbl@ld@##1}{##2}}%
3850       \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
3851       \edef\bbl@calendar{%
3852         \bbl@ld@calendar
3853         \ifx\bbl@ld@variant\@empty\else
3854           .\bbl@ld@variant
3855         \fi}%
3856       \bbl@replace\bbl@calendar{gregorian}{}%
3857     \fi
3858     \bbl@cased
3859     {\@nameuse\bbl@date@\language @\bbl@calendar}{#2}{#3}{#4}}%
3860   \endgroup}
3861 % eg: 1=months, 2=wide, 3=1, 4=dummy, 5=value, 6=calendar
3862 \def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
3863   \bbl@trim@def\bbl@tempa{#1.#2}%
3864   \bbl@ifsamestring{\bbl@tempa}{months.wide}%      to savedate
3865   {\bbl@trim@def\bbl@tempa{#3}%
3866     \bbl@trim\toks@{#5}%
3867     \@temptokena\expandafter{\bbl@savestate}%
3868     \bbl@exp{% Reverse order - in ini last wins
3869       \def\\bbl@savestate{%
3870         \\SetString\<month\romannumeral\bbl@tempa#6name>{\the\toks@}%
3871         \the\@temptokena}}%
3872     {\bbl@ifsamestring{\bbl@tempa}{date.long}%      defined now
3873       {\lowercase{\def\bbl@tempb{#6}}%
3874         \bbl@trim@def\bbl@toreplace{#5}%
3875         \bbl@TG@@date
3876         \bbl@ifunset\bbl@date@\language @}%
3877         {\global\bbl@csarg\let{date@\language @}\bbl@toreplace
3878           % TODO. Move to a better place.
3879         \bbl@exp{%
3880           \gdef\<\language date>{\protect\<\language date >}%
3881           \gdef\<\language date >####1####2####3{%
3882             \\bbl@usedategrouptrue
3883             \<\bbl@ensure@\language >{%
3884               \\localedate{####1}{####2}{####3}}}%
3885             \\bbl@add\\bbl@savestate}%
3886             \\SetString\\today{%
3887               \<\language date>%
3888               {\the\year}{\the\month}{\the\day}}}%
3889           }%
3890         \ifx\bbl@tempb\@empty\else
3891           \global\bbl@csarg\let{date@\language @}\bbl@tempb\bbl@toreplace
3892         \fi}%
3893     {}}}}

```

**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.

```

3894 \let\bbl@calendar\@empty
3895 \newcommand\BabelDateSpace{\nobreakspace}

```

```

3896 \newcommand\BabelDateDot{.\@} % TODO. \let instead of repeating
3897 \newcommand\BabelDated[1]{\number#1}
3898 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
3899 \newcommand\BabelDateM[1]{\number#1}
3900 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}
3901 \newcommand\BabelDateMMMM[1]{%
3902   \csname month\romannumeral#1\bb1@calendar name\endcsname}%
3903 \newcommand\BabelDatey[1]{\number#1}%
3904 \newcommand\BabelDateyy[1]{%
3905   \ifnum#1<10 0\number#1 %
3906   \else\ifnum#1<100 \number#1 %
3907   \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
3908   \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
3909   \else
3910     \bb1@error
3911     {Currently two-digit years are restricted to the\
3912       range 0-9999.}%
3913     {There is little you can do. Sorry.}%
3914   \fi\fi\fi\fi}
3915 \newcommand\BabelDateyyyy[1]{\number#1} % FIXME - add leading 0
3916 \def\bb1@replace@finish@iii#1{%
3917   \bb1@exp{\def\#1####1####2####3{\the\toks@}}
3918 \def\bb1@TG@date{%
3919   \bb1@replace\bb1@toreplace{[ ]}{\BabelDateSpace{}}%
3920   \bb1@replace\bb1@toreplace{[.]}{\BabelDateDot{}}%
3921   \bb1@replace\bb1@toreplace{[d]}{\BabelDated{####3}}%
3922   \bb1@replace\bb1@toreplace{[dd]}{\BabelDatedd{####3}}%
3923   \bb1@replace\bb1@toreplace{[M]}{\BabelDateM{####2}}%
3924   \bb1@replace\bb1@toreplace{[MM]}{\BabelDateMM{####2}}%
3925   \bb1@replace\bb1@toreplace{[MMMM]}{\BabelDateMMMM{####2}}%
3926   \bb1@replace\bb1@toreplace{[y]}{\BabelDatey{####1}}%
3927   \bb1@replace\bb1@toreplace{[yy]}{\BabelDateyy{####1}}%
3928   \bb1@replace\bb1@toreplace{[yyyy]}{\BabelDateyyyy{####1}}%
3929   \bb1@replace\bb1@toreplace{[y|]}{\bb1@datecctr[####1|]}%
3930   \bb1@replace\bb1@toreplace{[m|]}{\bb1@datecctr[####2|]}%
3931   \bb1@replace\bb1@toreplace{[d|]}{\bb1@datecctr[####3|]}%
3932 % Note after \bb1@replace \toks@ contains the resulting string.
3933 % TODO - Using this implicit behavior doesn't seem a good idea.
3934   \bb1@replace@finish@iii\bb1@toreplace}
3935 \def\bb1@datecctr{\expandafter\bb1@xdatecctr\expandafter}
3936 \def\bb1@xdatecctr[#1|#2]{\localnumeral{#2}{#1}}

```

### Transforms.

```

3937 \let\bb1@release@transforms\@empty
3938 \@namedef{bb1@inikv@transforms.prehyphenation}{%
3939   \bb1@transforms\babelprehyphenation}
3940 \@namedef{bb1@inikv@transforms.posthyphenation}{%
3941   \bb1@transforms\babelposthyphenation}
3942 \def\bb1@transforms@aux#1#2#3,#4\relax{#1{#2}{#3}{#4}}
3943 \begingroup
3944   \catcode`\%=12
3945   \catcode`\&=14
3946   \gdef\bb1@transforms#1#2#3{&%
3947     \ifx\bb1@KVP@transforms\@nil\else
3948       \directlua{
3949         str = [=[#2]=]
3950         str = str:gsub('%.%d+%.%d+$', '')
3951         tex.print([[ \def\string\babeltempa{]] .. str .. [{}]])
3952       }&%

```



```

3953 \bbl@xin@{,\babeltempa,}{,\bbl@KVP@transforms,}%&
3954 \ifin@
3955 \in@{.0$}{#2$}&%
3956 \ifin@
3957 \g@addto@macro\bbl@release@transforms{%&
3958 \relax\bbl@transforms@aux#1{\language}\language}{#3}}&%
3959 \else
3960 \g@addto@macro\bbl@release@transforms{, {#3}}&%
3961 \fi
3962 \fi
3963 \fi}
3964 \endgroup

```

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

```

3965 \def\bbl@provide@lsys#1{%
3966 \bbl@ifunset{bbl@lname@#1}%
3967 {\bbl@load@info{#1}}%
3968 }%
3969 \bbl@csarg\let{lsys@#1}\empty
3970 \bbl@ifunset{bbl@sname@#1}{\bbl@csarg\gdef{sname@#1}{Default}}}%
3971 \bbl@ifunset{bbl@sotf@#1}{\bbl@csarg\gdef{sotf@#1}{DFLT}}}%
3972 \bbl@csarg\bbl@add@list{lsys@#1}{Script=\bbl@cs{sname@#1}}%
3973 \bbl@ifunset{bbl@lname@#1}{%
3974 {\bbl@csarg\bbl@add@list{lsys@#1}{Language=\bbl@cs{lname@#1}}}%
3975 \ifcase\bbl@engine\or\or
3976 \bbl@ifunset{bbl@prehc@#1}{}%
3977 {\bbl@exp{\bbl@ifblank{\bbl@cs{prehc@#1}}}%
3978 }%
3979 {\ifx\bbl@xenohyph\@undefined
3980 \let\bbl@xenohyph\bbl@xenohyph@d
3981 \ifx\AtBeginDocument\@notprerr
3982 \expandafter\@secondoftwo % to execute right now
3983 \fi
3984 \AtBeginDocument{%
3985 \expandafter\bbl@add
3986 \csname selectfont \endcsname{\bbl@xenohyph}%
3987 \expandafter\selectlanguage\expandafter{\language}%
3988 \expandafter\bbl@toggle\csname selectfont \endcsname}%
3989 \fi}%
3990 \fi
3991 \bbl@csarg\bbl@toggle{lsys@#1}}
3992 \def\bbl@xenohyph@d{%
3993 \bbl@ifset{bbl@prehc@language}%
3994 {\ifnum\hyphenchar\font=\defaultshyphenchar
3995 \iffontchar\font\bbl@cl{prehc}\relax
3996 \hyphenchar\font\bbl@cl{prehc}\relax
3997 \else\iffontchar\font"200B
3998 \hyphenchar\font"200B
3999 \else
4000 \bbl@warning
4001 {Neither 0 nor ZERO WIDTH SPACE are available\\%
4002 in the current font, and therefore the hyphen\\%
4003 will be printed. Try changing the fontspec's\\%
4004 'HyphenChar' to another value, but be aware\\%
4005 this setting is not safe (see the manual)}}%
4006 \hyphenchar\font\defaultshyphenchar
4007 \fi\fi
4008 \fi}%

```

```

4009     {\hyphenchar\font\defaultthyphenchar}}
4010 % \fi}

```

The following ini reader ignores everything but the identification section. It is called when a font is defined (ie, when the language is first selected) to know which script/language must be enabled. This means we must make sure a few characters are not active. The ini is not read directly, but with a proxy tex file named as the language (which means any code in it must be skipped, too).

```

4011 \def\bbl@load@info#1{%
4012   \def\BabelBeforeIni##1##2{%
4013     \begingroup
4014       \bbl@read@ini{##1}0%
4015       \endinput           % babel- .tex may contain only preamble's
4016       \endgroup}%        boxed, to avoid extra spaces:
4017   {\bbl@input@texini{##1}}}
```

A tool to define the macros for native digits from the list provided in the ini file. Somewhat convoluted because there are 10 digits, but only 9 arguments in T<sub>E</sub>X. Non-digits characters are kept. The first macro is the generic “localized” command.

```

4018 \def\bbl@setdigits#1#2#3#4#5%
4019   \bbl@exp{%
4020     \def\<\language name digits>####1{%      ie, \langdigits
4021       \<\bbl@digits@\language name>####1\\\@nil}%
4022       \let\<\bbl@cntr@digits@\language name>\<\language name digits>%
4023       \def\<\language name counter>####1{%      ie, \langcounter
4024         \\\expandafter\<\bbl@counter@\language name>%
4025         \\\csname c@####1\endcsname}%
4026         \def\<\bbl@counter@\language name>####1{% ie, \bbl@counter@lang
4027           \\\expandafter\<\bbl@digits@\language name>%
4028           \\\number####1\\\@nil}}}%
4029 \def\bbl@tempa##1##2##3##4##5%
4030   \bbl@exp{%    Wow, quite a lot of hashes! :- (
4031     \def\<\bbl@digits@\language name>#####1{%
4032       \\\ifx#####1\\\@nil                % ie, \bbl@digits@lang
4033       \\\else
4034         \\\ifx0#####1#1%
4035         \\\else\\\ifx1#####1#2%
4036         \\\else\\\ifx2#####1#3%
4037         \\\else\\\ifx3#####1#4%
4038         \\\else\\\ifx4#####1#5%
4039         \\\else\\\ifx5#####1##1%
4040         \\\else\\\ifx6#####1##2%
4041         \\\else\\\ifx7#####1##3%
4042         \\\else\\\ifx8#####1##4%
4043         \\\else\\\ifx9#####1##5%
4044         \\\else#####1%
4045         \\\fi\\\fi\\\fi\\\fi\\\fi\\\fi\\\fi\\\fi\\\fi\\\fi\\\fi
4046         \\\expandafter\<\bbl@digits@\language name>%
4047         \\\fi}}}%
4048   \bbl@tempa}

```

Alphabetic counters must be converted from a space separated list to an \ifcase structure.

```

4049 \def\bbl@buildifcase#1 {% Returns \bbl@tempa, requires \toks@={%
4050   \ifx\\#1%           % \\ before, in case #1 is multiletter
4051     \bbl@exp{%
4052       \def\\ \bbl@tempa####1{%
4053         \<ifcase>####1\space\the\toks@\<else>\\@ctrerr\<fi>}}%
4054       \else
4055         \toks@\expandafter{\the\toks@\or #1}%
4056         \expandafter\bbl@buildifcase
4057       \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).

```

4058 \newcommand\localenumeral[2]{\bbl@cs{cntr@#1@\language}\{#2}}
4059 \def\bbl@localecntr#1#2{\localenumeral{#2}{#1}}
4060 \newcommand\localecounter[2]{%
4061   \expandafter\bbl@localecntr
4062   \expandafter{\number\csname c@#2\endcsname}\{#1}}
4063 \def\bbl@alphnumeral#1#2{%
4064   \expandafter\bbl@alphnumeral@i\number#2 76543210\@@{#1}}
4065 \def\bbl@alphnumeral@i#1#2#3#4#5#6#7#8\@@#9{%
4066   \ifcase\car#8\@nil\or   % Currenty <10000, but prepared for bigger
4067     \bbl@alphnumeral@ii{#9}000000#1\or
4068     \bbl@alphnumeral@ii{#9}00000#1#2\or
4069     \bbl@alphnumeral@ii{#9}0000#1#2#3\or
4070     \bbl@alphnumeral@ii{#9}000#1#2#3#4\else
4071     \bbl@alphnum@invalid{>9999}%
4072   \fi}
4073 \def\bbl@alphnumeral@ii#1#2#3#4#5#6#7#8{%
4074   \bbl@ifunset{bbl@cntr@#1.F.\number#5#6#7#8@\language}%
4075     {\bbl@cs{cntr@#1.4@\language}\{#5}
4076     \bbl@cs{cntr@#1.3@\language}\{#6}
4077     \bbl@cs{cntr@#1.2@\language}\{#7}
4078     \bbl@cs{cntr@#1.1@\language}\{#8}
4079     \ifnum#6#7#8>\z@ % TODO. An ad hoc rule for Greek. Ugly.
4080       \bbl@ifunset{bbl@cntr@#1.S.321@\language}\{}}
4081     {\bbl@cs{cntr@#1.S.321@\language}\{}}
4082   \fi}%
4083   {\bbl@cs{cntr@#1.F.\number#5#6#7#8@\language}\{}}
4084 \def\bbl@alphnum@invalid#1{%
4085   \bbl@error{Alphabetic numeral too large (#1)}%
4086   {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.

```

4087 \newcommand\localeinfo[1]{%
4088   \bbl@ifunset{bbl@\csname bbl@info@#1\endcsname @\language}%
4089     {\bbl@error{I've found no info for the current locale.\%
4090       The corresponding ini file has not been loaded\%
4091       Perhaps it doesn't exist}%
4092     {See the manual for details.}}%
4093   {\bbl@cs{\csname bbl@info@#1\endcsname @\language}\{}}
4094 \@namedef{bbl@info@name.locale}\{lname}
4095 \@namedef{bbl@info@tag.ini}\{lini}
4096 \@namedef{bbl@info@name.english}\{elname}
4097 \@namedef{bbl@info@name.opentype}\{lname}
4098 \@namedef{bbl@info@tag.bcp47}\{tbcp}
4099 \@namedef{bbl@info@language.tag.bcp47}\{lbcp}
4100 \@namedef{bbl@info@tag.opentype}\{lotf}
4101 \@namedef{bbl@info@script.name}\{esname}
4102 \@namedef{bbl@info@script.name.opentype}\{sname}
4103 \@namedef{bbl@info@script.tag.bcp47}\{sbcp}
4104 \@namedef{bbl@info@script.tag.opentype}\{sotf}
4105 \let\bbl@ensureinfo\@gobble
4106 \newcommand\BabelEnsureInfo{%
4107   \ifx\InputIfFileExists\undefined\else
4108     \def\bbl@ensureinfo#1{%

```

```

4109 \bbl@ifunset{bbl@lname@##1}{\bbl@load@info{##1}}{}}%
4110 \fi
4111 \bbl@foreach\bbl@loaded{%
4112 \def\language{##1}%
4113 \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`.

```

4114 \newcommand\getlocaleproperty{%
4115 \@ifstar\bbl@getproperty@s\bbl@getproperty@x}
4116 \def\bbl@getproperty@s#1#2#3{%
4117 \let#1\relax
4118 \def\bbl@elt##1##2##3{%
4119 \bbl@ifsamestring{##1/##2}{##3}%
4120 {\providecommand#1{##3}%
4121 \def\bbl@elt####1####2####3{}}}%
4122 {}}%
4123 \bbl@cs{inidata@#2}}%
4124 \def\bbl@getproperty@x#1#2#3{%
4125 \bbl@getproperty@s{#1}{#2}{#3}%
4126 \ifx#1\relax
4127 \bbl@error
4128 {Unknown key for locale '#2':\%
4129 #3}%
4130 \string#1 will be set to \relax}%
4131 {Perhaps you misspelled it.}%
4132 \fi}
4133 \let\bbl@ini@loaded\@empty
4134 \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.

```

4135 \newcommand\babeladjust[1]{% TODO. Error handling.
4136 \bbl@forkv{#1}{%
4137 \bbl@ifunset{bbl@ADJ@##1@##2}%
4138 {\bbl@cs{ADJ@##1}{##2}}%
4139 {\bbl@cs{ADJ@##1@##2}}}%
4140 %
4141 \def\bbl@adjust@lua#1#2{%
4142 \ifvmode
4143 \ifnum\currentgrouplevel=\z@
4144 \directlua{ Babel.#2 }%
4145 \expandafter\expandafter\expandafter\@gobble
4146 \fi
4147 \fi
4148 {\bbl@error % The error is gobbled if everything went ok.
4149 {Currently, #1 related features can be adjusted only\%
4150 in the main vertical list.}%
4151 {Maybe things change in the future, but this is what it is.}}}
4152 \@namedef{bbl@ADJ@bidi.mirroring@on}{%
4153 \bbl@adjust@lua{bidi}{mirroring_enabled=true}}
4154 \@namedef{bbl@ADJ@bidi.mirroring@off}{%
4155 \bbl@adjust@lua{bidi}{mirroring_enabled=false}}
4156 \@namedef{bbl@ADJ@bidi.text@on}{%
4157 \bbl@adjust@lua{bidi}{bidi_enabled=true}}
4158 \@namedef{bbl@ADJ@bidi.text@off}{%

```

```

4159 \bbl@adjust@lua{bidi}{bidi_enabled=false}}
4160 \@namedef{bbl@ADJ@bidi.mapdigits@on}{%
4161 \bbl@adjust@lua{bidi}{digits_mapped=true}}
4162 \@namedef{bbl@ADJ@bidi.mapdigits@off}{%
4163 \bbl@adjust@lua{bidi}{digits_mapped=false}}
4164 %
4165 \@namedef{bbl@ADJ@linebreak.sea@on}{%
4166 \bbl@adjust@lua{linebreak}{sea_enabled=true}}
4167 \@namedef{bbl@ADJ@linebreak.sea@off}{%
4168 \bbl@adjust@lua{linebreak}{sea_enabled=false}}
4169 \@namedef{bbl@ADJ@linebreak.cjk@on}{%
4170 \bbl@adjust@lua{linebreak}{cjk_enabled=true}}
4171 \@namedef{bbl@ADJ@linebreak.cjk@off}{%
4172 \bbl@adjust@lua{linebreak}{cjk_enabled=false}}
4173 \@namedef{bbl@ADJ@justify.arabic@on}{%
4174 \bbl@adjust@lua{linebreak}{arabic.justify_enabled=true}}
4175 \@namedef{bbl@ADJ@justify.arabic@off}{%
4176 \bbl@adjust@lua{linebreak}{arabic.justify_enabled=false}}
4177 %
4178 \def\bbl@adjust@layout#1{%
4179 \ifvmode
4180 #1%
4181 \expandafter\@gobble
4182 \fi
4183 {\bbl@error % The error is gobbled if everything went ok.
4184 {Currently, layout related features can be adjusted only\\%
4185 in vertical mode.}%
4186 {Maybe things change in the future, but this is what it is.}}
4187 \@namedef{bbl@ADJ@layout.tabular@on}{%
4188 \bbl@adjust@layout{\let\@tabular\bbl@NL@tabular}}
4189 \@namedef{bbl@ADJ@layout.tabular@off}{%
4190 \bbl@adjust@layout{\let\@tabular\bbl@OL@tabular}}
4191 \@namedef{bbl@ADJ@layout.lists@on}{%
4192 \bbl@adjust@layout{\let\list\bbl@NL@list}}
4193 \@namedef{bbl@ADJ@layout.lists@off}{%
4194 \bbl@adjust@layout{\let\list\bbl@OL@list}}
4195 \@namedef{bbl@ADJ@hyphenation.extra@on}{%
4196 \bbl@activateposthyphen}
4197 %
4198 \@namedef{bbl@ADJ@autoload.bcp47@on}{%
4199 \bbl@bcpallowedtrue}
4200 \@namedef{bbl@ADJ@autoload.bcp47@off}{%
4201 \bbl@bcpallowedfalse}
4202 \@namedef{bbl@ADJ@autoload.bcp47.prefix#1}{%
4203 \def\bbl@bcp@prefix{#1}}
4204 \def\bbl@bcp@prefix{bcp47-}
4205 \@namedef{bbl@ADJ@autoload.options#1}{%
4206 \def\bbl@autoload@options{#1}}
4207 \let\bbl@autoload@bcptoptions\@empty
4208 \@namedef{bbl@ADJ@autoload.bcp47.options#1}{%
4209 \def\bbl@autoload@bcptoptions{#1}}
4210 \newif\ifbbl@bcptoname
4211 \@namedef{bbl@ADJ@bcp47.toname@on}{%
4212 \bbl@bcptonametrue}
4213 \BabelEnsureInfo}
4214 \@namedef{bbl@ADJ@bcp47.toname@off}{%
4215 \bbl@bcptonamefalse}
4216 \@namedef{bbl@ADJ@prehyphenation.disable@nohyphenation}{%
4217 \directlua{ Babel.ignore_pre_char = function(node)

```

```

4218     return (node.lang == \the\csname l@nohyphenation\endcsname)
4219   end }}
4220 \@namedef{bbl@ADJ@prehyphenation.disable@off}{%
4221   \directlua{ Babel.ignore_pre_char = function(node)
4222     return false
4223   end }}
4224 % TODO: use babel name, override
4225 %
4226 % As the final task, load the code for lua.
4227 %
4228 \ifx\directlua\@undefined\else
4229   \ifx\bbl@luapatterns\@undefined
4230     \input luababel.def
4231   \fi
4232 \fi
4233 </core>

A proxy file for switch.def

4234 <*kernel>
4235 \let\bbl@onlyswitch\@empty
4236 \input babel.def
4237 \let\bbl@onlyswitch\@undefined
4238 </kernel>
4239 <*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  $\text{\texttt{M\TeX}}$  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.

```

4240 <<Make sure ProvidesFile is defined>>
4241 \ProvidesFile{hyphen.cfg}[<<date>> <<version>> Babel hyphens]
4242 \xdef\bbl@format{\jobname}
4243 \def\bbl@version{<<version>>}
4244 \def\bbl@date{<<date>>}
4245 \ifx\AtBeginDocument\@undefined
4246   \def\@empty{}
4247   \let\orig@dump\dump
4248   \def\dump{%
4249     \ifx\@ztryfc\@undefined
4250     \else
4251       \toks0=\expandafter{\@preamblecmds}%
4252       \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
4253       \def\@begindocumenthook{}%
4254     \fi
4255     \let\dump\orig@dump\let\orig@dump\@undefined\dump}
4256 \fi
4257 <<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.

```
4258 \def\process@line#1#2 #3 #4 {%
4259   \ifx=#1%
4260     \process@synonym{#2}%
4261   \else
4262     \process@language{#1#2}{#3}{#4}%
4263   \fi
4264   \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.

```
4265 \toks@{}
4266 \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.

```
4267 \def\process@synonym#1{%
4268   \ifnum\last@language=\m@ne
4269     \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
4270   \else
4271     \expandafter\chardef\csname l@#1\endcsname\last@language
4272     \wlog{\string\l@#1=\string\language\the\last@language}%
4273     \expandafter\let\csname #1hyphenmins\endcsname\expandafter\endcsname
4274     \csname\language\endcsname hyphenmins\endcsname
4275     \let\bbl@elt\relax
4276     \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}}}%
4277   \fi}
```

`\process@language` The macro `\process@language` is used to process a non-empty line from the ‘configuration file’. It has three arguments, each delimited by white space. The first argument is the ‘name’ of a language; the second is the name of the file that contains the patterns. The optional third argument is the name of a file containing hyphenation exceptions.

The first thing to do is call `\addlanguage` to allocate a pattern register and to make that register ‘active’. Then the pattern file is read.

For some hyphenation patterns it is needed to load them with a specific font encoding selected. This can be specified in the file `language.dat` by adding for instance ‘:T1’ to the name of the language.

The macro `\bbl@get@enc` extracts the font encoding from the language name and stores it in `\bbl@hyph@enc`. The latter can be used in hyphenation files if you need to set a behavior depending on the given encoding (it is set to empty if no encoding is given).

Pattern files may contain assignments to `\lefthyphenmin` and `\righthyphenmin`.  $\TeX$  does not keep track of these assignments. Therefore we try to detect such assignments and store them in the `\(lang)hyphenmins` macro. When no assignments were made we provide a default setting.

Some pattern files contain changes to the `\lccode` and `\uccode` arrays. Such changes should remain local to the language; therefore we process the pattern file in a group; the `\patterns` command acts globally so its effect will be remembered.

Then we globally store the settings of `\lefthyphenmin` and `\righthyphenmin` and close the group.

When the hyphenation patterns have been processed we need to see if a file with hyphenation exceptions needs to be read. This is the case when the third argument is not empty and when it does not contain a space token. (Note however there is no need to save hyphenation exceptions into the format.)

`\bbl@languages` saves a snapshot of the loaded languages in the form

`\bbl@elt{<language-name>}{<number>}{<patterns-file>}{<exceptions-file>}`. Note the last 2 arguments are empty in ‘dialects’ defined in `language.dat` with `=`. Note also the language name can have encoding info.

Finally, if the counter `\language` is equal to zero we execute the synonyms stored.

```

4278 \def\process@language#1#2#3{%
4279   \expandafter\addlanguage\csname l@#1\endcsname
4280   \expandafter\language\csname l@#1\endcsname
4281   \edef\language{#1}%
4282   \bbl@hook@everylanguage{#1}%
4283   % > luatex
4284   \bbl@get@enc#1::\@@@
4285   \begingroup
4286     \lefthyphenmin\m@ne
4287     \bbl@hook@loadpatterns{#2}%
4288     % > luatex
4289     \ifnum\lefthyphenmin=\m@ne
4290     \else
4291       \expandafter\xdef\csname #1hyphenmins\endcsname{%
4292         \the\lefthyphenmin\the\righthyphenmin}%
4293       \fi
4294     \endgroup
4295   \def\bbl@tempa{#3}%
4296   \ifx\bbl@tempa\@empty\else
4297     \bbl@hook@loadexceptions{#3}%
4298     % > luatex
4299   \fi
4300   \let\bbl@elt\relax
4301   \edef\bbl@languages{%
4302     \bbl@languages\bbl@elt{#1}\the\language}{#2}{\bbl@tempa}}%
4303   \ifnum\the\language=\z@
4304     \expandafter\ifx\csname #1hyphenmins\endcsname\relax
4305       \set@hyphenmins\tw@\thr@@\relax
4306     \else
4307       \expandafter\expandafter\expandafter\set@hyphenmins
4308       \csname #1hyphenmins\endcsname
4309     \fi
4310     \the\toks@
4311     \toks@{}}%
4312   \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.

```

4313 \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.

```

4314 \def\bbl@hook@everylanguage#1{}
4315 \def\bbl@hook@loadpatterns#1{\input #1\relax}
4316 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
4317 \def\bbl@hook@loadkernel#1{%
4318   \def\addlanguage{\csname newlanguage\endcsname}%
4319   \def\adddialect##1##2{%
4320     \global\chardef##1##2\relax
4321     \wlog{\string##1 = a dialect from \string\language##2}}%
4322   \def\iflanguage##1{%
4323     \expandafter\ifx\csname l@##1\endcsname\relax
4324       \@nolanerr{##1}%
4325     \else
4326       \ifnum\csname l@##1\endcsname=\language
4327         \expandafter\expandafter\expandafter\@firstoftwo
4328       \else

```



```

4329     \expandafter\expandafter\expandafter\@secondoftwo
4330     \fi
4331     \fi}%
4332 \def\providehyphenmins##1##2{%
4333     \expandafter\ifx\csname ##1hyphenmins\endcsname\relax
4334     \@namedef{##1hyphenmins}{##2}%
4335     \fi}%
4336 \def\set@hyphenmins##1##2{%
4337     \lefthyphenmin##1\relax
4338     \righthyphenmin##2\relax}%
4339 \def\selectlanguage{%
4340     \errhelp{Selecting a language requires a package supporting it}%
4341     \errmessage{Not loaded}}%
4342 \let\foreignlanguage\selectlanguage
4343 \let\otherlanguage\selectlanguage
4344 \expandafter\let\csname otherlanguage*\endcsname\selectlanguage
4345 \def\bbl@usehooks##1##2{% TODO. Temporary!!
4346 \def\setlocale{%
4347     \errhelp{Find an armchair, sit down and wait}%
4348     \errmessage{Not yet available}}%
4349 \let\uselocale\setlocale
4350 \let\locale\setlocale
4351 \let\selectlocale\setlocale
4352 \let\localename\setlocale
4353 \let\textlocale\setlocale
4354 \let\textlanguage\setlocale
4355 \let\languagetext\setlocale}
4356 \begingroup
4357 \def\AddBabelHook#1#2{%
4358     \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
4359     \def\next{\toks1}%
4360     \else
4361     \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname####1}%
4362     \fi
4363     \next}
4364 \ifx\directlua@undefined
4365     \ifx\XeTeXinputencoding@undefined\else
4366     \input xebabel.def
4367     \fi
4368 \else
4369     \input luababel.def
4370 \fi
4371 \openin1 = babel-\bbl@format.cfg
4372 \ifeof1
4373 \else
4374     \input babel-\bbl@format.cfg\relax
4375 \fi
4376 \closein1
4377 \endgroup
4378 \bbl@hook@loadkernel{switch.def}

```

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

```

4379 \openin1 = language.dat

```

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

```

4380 \def\languagename{english}%
4381 \ifeof1
4382 \message{I couldn't find the file language.dat,\space

```

```

4383         I will try the file hyphen.tex}
4384 \input hyphen.tex\relax
4385 \chardef\l@english\z@
4386 \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`.

```

4387 \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.

```

4388 \loop
4389   \endlinechar\m@ne
4390   \read1 to \bbl@line
4391   \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.

```

4392   \if T\ifeof1F\fi T\relax
4393   \ifx\bbl@line\empty\else
4394     \edef\bbl@line{\bbl@line\space\space\space}%
4395     \expandafter\process@line\bbl@line\relax
4396   \fi
4397 \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.

```

4398 \begingroup
4399   \def\bbl@elt#1#2#3#4{%
4400     \global\language=#2\relax
4401     \gdef\language#1}%
4402   \def\bbl@elt##1##2##3##4{}}%
4403   \bbl@languages
4404 \endgroup
4405 \fi
4406 \closein1

```

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

```

4407 \if\the\toks@\else
4408   \errhelp{language.dat loads no language, only synonyms}
4409   \errmessage{Orphan language synonym}
4410 \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.

```

4411 \let\bbl@line\undefined
4412 \let\process@line\undefined
4413 \let\process@synonym\undefined
4414 \let\process@language\undefined
4415 \let\bbl@get@enc\undefined
4416 \let\bbl@hyph@enc\undefined
4417 \let\bbl@tempa\undefined
4418 \let\bbl@hook@loadkernel\undefined
4419 \let\bbl@hook@everylanguage\undefined
4420 \let\bbl@hook@loadpatterns\undefined

```

```

4421 \let\bbl@hook@loadexceptions\@undefined
4422 \end{patterns}

```

Here the code for `initTeX` 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].

```

4423 \langle\langle *More package options \rangle\rangle \equiv
4424 \chardef\bbl@bidimode\z@
4425 \DeclareOption{bidi=default}{\chardef\bbl@bidimode=\@ne}
4426 \DeclareOption{bidi=basic}{\chardef\bbl@bidimode=101 }
4427 \DeclareOption{bidi=basic-r}{\chardef\bbl@bidimode=102 }
4428 \DeclareOption{bidi=bidi}{\chardef\bbl@bidimode=201 }
4429 \DeclareOption{bidi=bidi-r}{\chardef\bbl@bidimode=202 }
4430 \DeclareOption{bidi=bidi-l}{\chardef\bbl@bidimode=203 }
4431 \rangle\langle /More package options \rangle\rangle

```

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.

```

4432 \langle\langle *Font selection \rangle\rangle \equiv
4433 \bbl@trace{Font handling with fontspec}
4434 \ifx\ExplSyntaxOn\@undefined\else
4435   \ExplSyntaxOn
4436   \catcode\ =10
4437   \def\bbl@loadfontspec{%
4438     \usepackage{fontspec}%
4439     \expandafter
4440     \def\csname msg~text~>~fontspec/language-not-exist\endcsname##1##2##3##4{%
4441       Font '\l_fontspec_fontname_tl' is using the\\%
4442       default features for language '##1'.\\%
4443       That's usually fine, because many languages\\%
4444       require no specific features, but if the output is\\%
4445       not as expected, consider selecting another font.}
4446     \expandafter
4447     \def\csname msg~text~>~fontspec/no-script\endcsname##1##2##3##4{%
4448       Font '\l_fontspec_fontname_tl' is using the\\%
4449       default features for script '##2'.\\%
4450       That's not always wrong, but if the output is\\%
4451       not as expected, consider selecting another font.}}
4452   \ExplSyntaxOff
4453 \fi
4454 \onlypreamble\babelfont
4455 \newcommand\babelfont[2][{}]{% 1=langs/scripts 2=fam
4456   \bbl@foreach{#1}{%
4457     \expandafter\ifx\csname date##1\endcsname\relax
4458       \IfFileExists{babel-##1.tex}%
4459       {\babelprovide{##1}}}%
4460     {}%
4461   \fi}%
4462 \edef\bbl@tempa{#1}%
4463 \def\bbl@tempb{#2}% Used by \bbl@bblfont
4464 \ifx\fontspec\@undefined

```

```

4465 \bbl@loadfontspec
4466 \fi
4467 \EnableBabelHook{babel-fontspec}% Just calls \bbl@switchfont
4468 \bbl@bblfont}
4469 \newcommand\bbl@bblfont[2][{}]{% 1=features 2=fontname, @font=rm|sf|tt
4470 \bbl@ifunset{\bbl@tempb family}%
4471 {\bbl@providfam{\bbl@tempb}}%
4472 {\bbl@exp{%
4473 \\\bbl@sreplace\<\bbl@tempb family >%
4474 {\@nameuse{\bbl@tempb default}}{\<\bbl@tempb default>}}}%
4475 % For the default font, just in case:
4476 \bbl@ifunset{\bbl@lsys@\language}\bbl@provide@lsys{\language}}{}%
4477 \expandafter\bbl@ifblank\expandafter{\bbl@tempa}%
4478 {\bbl@csarg\edef{\bbl@tempb dflt@}{<{#1}{#2}}% save bbl@rmdflt@
4479 \bbl@exp{%
4480 \let\<bbl@\bbl@tempb dflt@\language>\<bbl@\bbl@tempb dflt@>%
4481 \\\bbl@font@set\<bbl@\bbl@tempb dflt@\language>%
4482 \<\bbl@tempb default>\<\bbl@tempb family>}}%
4483 {\bbl@foreach\bbl@tempa{% ie bbl@rmdflt@lang / *scrt
4484 \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:

```

4485 \def\bbl@providfam#1{%
4486 \bbl@exp{%
4487 \\\newcommand\<#1default>{}% Just define it
4488 \\\bbl@add@list\\bbl@font@fams{#1}%
4489 \\\DeclareRobustCommand\<#1family>{%
4490 \\\not@math@alphabet\<#1family>\relax
4491 \\\fontfamily\<#1default>\selectfont}%
4492 \\\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.

```

4493 \def\bbl@nostdfont#1{%
4494 \bbl@ifunset{\bbl@WFF@\f@family}%
4495 {\bbl@csarg\gdef\WFF@\f@family}% Flag, to avoid dupl warns
4496 \bbl@infowarn{The current font is not a babel standard family:\%
4497 #1%
4498 \fontname\font\%
4499 There is nothing intrinsically wrong with this warning, and\%
4500 you can ignore it altogether if you do not need these\%
4501 families. But if they are used in the document, you should be\%
4502 aware 'babel' will no set Script and Language for them, so\%
4503 you may consider defining a new family with \string\babelfont.\%
4504 See the manual for further details about \string\babelfont.\%
4505 Reported}}
4506 {}}%
4507 \gdef\bbl@switchfont{%
4508 \bbl@ifunset{\bbl@lsys@\language}\bbl@provide@lsys{\language}}{}%
4509 \bbl@exp{% eg Arabic -> arabic
4510 \lowercase{\edef\\bbl@tempa{\bbl@cl{sname}}}%
4511 \bbl@foreach\bbl@font@fams{%
4512 \bbl@ifunset{\bbl@##1dflt@\language}% (1) language?
4513 {\bbl@ifunset{\bbl@##1dflt@*\bbl@tempa}% (2) from script?
4514 {\bbl@ifunset{\bbl@##1dflt@}% 2=F - (3) from generic?
4515 {}% 123=F - nothing!
4516 {\bbl@exp{% 3=T - from generic
4517 \global\let\<bbl@##1dflt@\language>%
4518 \<bbl@##1dflt@>}}}%

```

```

4519         {\bbl@exp{%                2=T - from script
4520         \global\let\<bbl@##1dflt@\languagename>%
4521         \<bbl@##1dflt@*\bbl@tempa>}}}%
4522     {}}%                1=T - language, already defined
4523 \def\bbl@tempa{\bbl@nostdfont{}}%
4524 \bbl@foreach\bbl@font@fams{%      don't gather with prev for
4525     \bbl@ifunset{\bbl@##1dflt@\languagename}%
4526     {\bbl@cs{famrst@##1}%
4527     \global\bbl@csarg\let{famrst@##1}\relax}%
4528     {\bbl@exp{% order is relevant. TODO: but sometimes wrong!
4529     \\\bbl@add\\\originalTeX{%
4530     \\\bbl@font@rst{\bbl@c1{##1dflt}}}%
4531     \<##1default>\<##1family>{##1}}}%
4532     \\\bbl@font@set\<bbl@##1dflt@\languagename>% the main part!
4533     \<##1default>\<##1family>}}}%
4534 \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.

```

4535 \ifx\f@family\undefined\else      % if latex
4536 \ifcase\bbl@engine                 % if pdftex
4537     \let\bbl@cckcckstdfonts\relax
4538 \else
4539     \def\bbl@cckcckstdfonts{%
4540         \begingroup
4541         \global\let\bbl@cckcckstdfonts\relax
4542         \let\bbl@tempa\@empty
4543         \bbl@foreach\bbl@font@fams{%
4544             \bbl@ifunset{\bbl@##1dflt@}%
4545             {\@nameuse{##1family}%
4546             \bbl@csarg\gdef{WFF@\f@family}}}% Flag
4547             \bbl@exp{\\\bbl@add\\\bbl@tempa{* \<##1family>= \f@family\\}%
4548             \space\space\fontname\font\\}%
4549             \bbl@csarg\xdef{##1dflt@}{\f@family}%
4550             \expandafter\xdef\csname ##1default\endcsname{\f@family}}}%
4551             {}}}%
4552 \ifx\bbl@tempa\@empty\else
4553     \bbl@infowarn{The following font families will use the default\\%
4554     settings for all or some languages:\\%
4555     \bbl@tempa
4556     There is nothing intrinsically wrong with it, but\\%
4557     'babel' will no set Script and Language, which could\\%
4558     be relevant in some languages. If your document uses\\%
4559     these families, consider redefining them with \string\babelfont.\\%
4560     Reported}%
4561 \fi
4562 \endgroup}
4563 \fi
4564 \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.

```

4565 \def\bbl@font@set#1#2#3{% eg \bbl@rmdflt@lang \rmdefault \rmfamily
4566     \bbl@xin@{<>}{#1}%
4567     \ifin@
4568     \bbl@exp{\\\bbl@fontspec@set\\#1\expandafter@gobbletwo#1\\#3}%
4569     \fi

```

```

4570 \bbl@exp{% 'Unprotected' macros return prev values
4571 \def\#2#1% eg, \rmdefault{\bbl@rmdflt@lang}
4572 \\\bbl@ifsamestring{#2}{\f@family}%
4573 {\#3%
4574 \\\bbl@ifsamestring{\f@series}{\bfdefault}{\bfseries}}}%
4575 \let\\bbl@tempa\relax}%
4576 {}}}
4577 % TODO - next should be global?, but even local does its job. I'm
4578 % still not sure -- must investigate:
4579 \def\bbl@fontspec@set#1#2#3#4{% eg \bbl@rmdflt@lang fnt-opt fnt-nme \xxfamily
4580 \let\bbl@tempe\bbl@mapselect
4581 \let\bbl@mapselect\relax
4582 \let\bbl@temp@fam#4% eg, '\rmfamily', to be restored below
4583 \let#4\empty % Make sure \renewfontfamily is valid
4584 \bbl@exp{%
4585 \let\\bbl@temp@pfam<\bbl@stripslash#4\space>% eg, '\rmfamily'
4586 <keys_if_exist:nnF>{fontspec-opentype}{Script/\bbl@cl{sname}}}%
4587 {\newfontscript{\bbl@cl{sname}}{\bbl@cl{sotf}}}%
4588 <keys_if_exist:nnF>{fontspec-opentype}{Language/\bbl@cl{lname}}}%
4589 {\newfontlanguage{\bbl@cl{lname}}{\bbl@cl{lotf}}}%
4590 \renewfontfamily\#4%
4591 [\bbl@cs{lsys@\languagename},#2]}{#3}% ie \bbl@exp{.}{#3}
4592 \begingroup
4593 #4%
4594 \xdef#1{\f@family}% eg, \bbl@rmdflt@lang{FreeSerif(0)}
4595 \endgroup
4596 \let#4\bbl@temp@fam
4597 \bbl@exp{\let<\bbl@stripslash#4\space>\bbl@temp@pfam
4598 \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.

```

4599 \def\bbl@font@rst#1#2#3#4{%
4600 \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.

```

4601 \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 :-).

```

4602 \newcommand\babelFSstore[2][{%
4603 \bbl@ifblank{#1}%
4604 {\bbl@csarg\def{sname@#2}{Latin}}%
4605 {\bbl@csarg\def{sname@#2}{#1}}%
4606 \bbl@provide@dirs{#2}%
4607 \bbl@csarg\ifnum{wdir@#2}>\z@
4608 \let\bbl@beforeforeign\leavevmode
4609 \EnableBabelHook{babel-bidi}%
4610 \fi
4611 \bbl@foreach{#2}{%
4612 \bbl@FSstore{##1}{rm}\rmdefault\bbl@save@rmdefault
4613 \bbl@FSstore{##1}{sf}\sfdefault\bbl@save@sfdefault
4614 \bbl@FSstore{##1}{tt}\ttdefault\bbl@save@ttdefault}}
4615 \def\bbl@FSstore#1#2#3#4{%
4616 \bbl@csarg\edef{#2default#1}{#3}%
4617 \expandafter\addto\csname extras#1\endcsname{%
4618 \let#4#3%
4619 \ifx#3\f@family

```

```

4620     \edef#3{\csname bbl@#2default#1\endcsname}%
4621     \fontfamily{#3}\selectfont
4622   \else
4623     \edef#3{\csname bbl@#2default#1\endcsname}%
4624     \fi}%
4625   \expandafter\addto\csname noextras#1\endcsname{%
4626     \ifx#3\f@family
4627       \fontfamily{#4}\selectfont
4628       \fi
4629     \let#3#4}}
4630 \let\bbl@langfeatures\@empty
4631 \def\babelFSfeatures{% make sure \fontspec is redefined once
4632   \let\bbl@ori@fontspec\fontspec
4633   \renewcommand\fontspec[1][{}]{%
4634     \bbl@ori@fontspec[\bbl@langfeatures##1]}
4635   \let\babelFSfeatures\bbl@FSfeatures
4636   \babelFSfeatures}
4637 \def\bbl@FSfeatures#1#2{%
4638   \expandafter\addto\csname extras#1\endcsname{%
4639     \babel@save\bbl@langfeatures
4640     \edef\bbl@langfeatures{#2,}}
4641 <</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.

```

4642 <<(*Footnote changes)>> ≡
4643 \bbl@trace{Bidi footnotes}
4644 \ifnum\bbl@bidimode>\z@
4645   \def\bbl@footnote#1#2#3{%
4646     \@ifnextchar[%
4647       {\bbl@footnote@o{#1}{#2}{#3}}%
4648       {\bbl@footnote@x{#1}{#2}{#3}}}
4649   \long\def\bbl@footnote@x#1#2#3#4{%
4650     \bgroup
4651       \select@language@x{\bbl@main@language}%
4652       \bbl@fn@footnote{#2#1{\ignorespaces#4}#3}%
4653     \egroup}
4654   \long\def\bbl@footnote@o#1#2#3[#4]#5{%
4655     \bgroup
4656       \select@language@x{\bbl@main@language}%
4657       \bbl@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}%
4658     \egroup}
4659   \def\bbl@footnotetext#1#2#3{%
4660     \@ifnextchar[%
4661       {\bbl@footnotetext@o{#1}{#2}{#3}}%
4662       {\bbl@footnotetext@x{#1}{#2}{#3}}}
4663   \long\def\bbl@footnotetext@x#1#2#3#4{%
4664     \bgroup
4665       \select@language@x{\bbl@main@language}%
4666       \bbl@fn@footnotetext{#2#1{\ignorespaces#4}#3}%
4667     \egroup}
4668   \long\def\bbl@footnotetext@o#1#2#3[#4]#5{%
4669     \bgroup
4670     \select@language@x{\bbl@main@language}%

```

```

4671 \bbl@fn@footnotetext[#4]{#2#1{\ignorespaces#5}#3}%
4672 \egroup}
4673 \def\BabelFootnote#1#2#3#4{%
4674 \ifx\bbl@fn@footnote\@undefined
4675 \let\bbl@fn@footnote\footnote
4676 \fi
4677 \ifx\bbl@fn@footnotetext\@undefined
4678 \let\bbl@fn@footnotetext\footnotetext
4679 \fi
4680 \bbl@ifblank{#2}%
4681 {\def#1{\bbl@footnote{\@firstofone}{#3}{#4}}
4682 \@namedef{\bbl@stripslash#1text}%
4683 {\bbl@footnotetext{\@firstofone}{#3}{#4}}}%
4684 {\def#1{\bbl@exp{\bbl@footnote{\foreignlanguage{#2}}}{#3}{#4}}%
4685 \@namedef{\bbl@stripslash#1text}%
4686 {\bbl@exp{\bbl@footnotetext{\foreignlanguage{#2}}}{#3}{#4}}}%
4687 \fi
4688 <</Footnote changes>>

```

Now, the code.

```

4689 (*xetex)
4690 \def\BabelStringsDefault{unicode}
4691 \let\xebbl@stop\relax
4692 \AddBabelHook{xetex}{encodedcommands}{%
4693 \def\bbl@tempa{#1}%
4694 \ifx\bbl@tempa\@empty
4695 \XeTeXinputencoding"bytes"%
4696 \else
4697 \XeTeXinputencoding"#1"%
4698 \fi
4699 \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
4700 \AddBabelHook{xetex}{stopcommands}{%
4701 \xebbl@stop
4702 \let\xebbl@stop\relax}
4703 \def\bbl@intraspace#1 #2 #3\@@{%
4704 \bbl@csarg\gdef{\xeisp@language}%
4705 {\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}}
4706 \def\bbl@intrapenalty#1\@@{%
4707 \bbl@csarg\gdef{\xeipn@language}%
4708 {\XeTeXlinebreakpenalty #1\relax}}
4709 \def\bbl@provide@intraspace{%
4710 \bbl@xin@{/s}{\bbl@cl{\lnbrk}}%
4711 \ifin@else\bbl@xin@{/c}{\bbl@cl{\lnbrk}}\fi
4712 \ifin@
4713 \bbl@ifunset{\bbl@intsp@language}{%
4714 {\expandafter\ifx\csname\bbl@intsp@language\endcsname\@empty\else
4715 \ifx\bbl@KVP@intraspace\@nil
4716 \bbl@exp{%
4717 \bbl@intraspace\bbl@cl{intsp}\@@}%
4718 \fi
4719 \ifx\bbl@KVP@intrapenalty\@nil
4720 \bbl@intrapenalty0\@@
4721 \fi
4722 \fi
4723 \ifx\bbl@KVP@intraspace\@nil\else % We may override the ini
4724 \expandafter\bbl@intraspace\bbl@KVP@intraspace\@@
4725 \fi
4726 \ifx\bbl@KVP@intrapenalty\@nil\else
4727 \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@@

```



```

4728 \fi
4729 \bbl@exp{%
4730   \\bbl@add\<extras\language>{%
4731     \XeTeXlinebreaklocale "\bbl@cl{tbc}"%
4732     \<bbl@xeisp@\language>%
4733     \<bbl@xeipn@\language>%
4734     \\bbl@toglobal\<extras\language>%
4735     \\bbl@add\<noextras\language>{%
4736       \XeTeXlinebreaklocale "en"%
4737       \\bbl@toglobal\<noextras\language>%
4738   \ifx\bbl@ispace\@undefined
4739     \gdef\bbl@ispace{\bbl@cl{xeisp}}%
4740   \ifx\AtBeginDocument\@notprerr
4741     \expandafter\@secondoftwo % to execute right now
4742   \fi
4743   \AtBeginDocument{%
4744     \expandafter\bbl@add
4745     \csname selectfont \endcsname{\bbl@ispace}%
4746     \expandafter\bbl@toglobal\csname selectfont \endcsname}%
4747   \fi}%
4748 \fi}
4749 \ifx\DisableBabelHook\@undefined\endinput\fi
4750 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
4751 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ckeckstdfonts}
4752 \DisableBabelHook{babel-fontspec}
4753 <<Font selection>>
4754 \input txtbabel.def
4755 </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>.

```

4756 (*texxet)
4757 \providecommand\bbl@provide@intraspace{}
4758 \bbl@trace{Redefinitions for bidi layout}
4759 \def\bbl@sspre@caption{%
4760   \bbl@exp{\everyhbox{\\bbl@textdir\bbl@cs{wdir@\bbl@main@language}}}}
4761 \ifx\bbl@opt@layout\@nnil\endinput\fi % No layout
4762 \def\bbl@startskip{\ifcase\bbl@thepardir\leftskip\else\rightskip\fi}
4763 \def\bbl@endskip{\ifcase\bbl@thepardir\rightskip\else\leftskip\fi}
4764 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
4765   \def\@hangfrom#1{%
4766     \setbox\@tempboxa\hbox{#1}%
4767     \hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
4768     \noindent\box\@tempboxa}
4769 \def\raggedright{%
4770   \let\\\@centercr
4771   \bbl@startskip\z@skip
4772   \@rightskip\@flushglue
4773   \bbl@endskip\@rightskip
4774   \parindent\z@
4775   \parfillskip\bbl@startskip}

```

```

4776 \def\raggedleft{%
4777   \let\\@centercr
4778   \bbl@startskip\@flushglue
4779   \bbl@endskip\z@skip
4780   \parindent\z@
4781   \parfillskip\bbl@endskip}
4782 \fi
4783 \IfBabelLayout{lists}
4784   {\bbl@sreplace\list
4785     {\@totalleftmargin\leftmargin}{\@totalleftmargin\bbl@listleftmargin}%
4786     \def\bbl@listleftmargin{%
4787       \ifcase\bbl@thepardir\leftmargin\else\rightmargin\fi}%
4788     \ifcase\bbl@engine
4789       \def\labelenumii{}\theenumii{}\pdfTeX doesn't reverse ()
4790       \def\p@enumii{\p@enumii}\theenumii}%
4791     \fi
4792     \bbl@sreplace\@verbatim
4793       {\leftskip\@totalleftmargin}%
4794       {\bbl@startskip\textwidth
4795         \advance\bbl@startskip-\linewidth}%
4796     \bbl@sreplace\@verbatim
4797       {\rightskip\z@skip}%
4798       {\bbl@endskip\z@skip}}%
4799   {}
4800 \IfBabelLayout{contents}
4801   {\bbl@sreplace\@dottedtocline{\leftskip}{\bbl@startskip}%
4802     \bbl@sreplace\@dottedtocline{\rightskip}{\bbl@endskip}}
4803   {}
4804 \IfBabelLayout{columns}
4805   {\bbl@sreplace\@outputdblcol{\hb@xt@\textwidth}{\bbl@outputbox}%
4806     \def\bbl@outputbox#1{%
4807       \hb@xt@\textwidth{%
4808         \hskip\columnwidth
4809         \hfil
4810         {\normalcolor\vrule \@width\columnseprule}%
4811         \hfil
4812         \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
4813         \hskip-\textwidth
4814         \hb@xt@\columnwidth{\box\@outputbox \hss}%
4815         \hskip\columnsep
4816         \hskip\columnwidth}}}%
4817   {}
4818 \langle\langle Footnote changes \rangle\rangle
4819 \IfBabelLayout{footnotes}%
4820   {\BabelFootnote\footnote\language\language{}{}}%
4821   \BabelFootnote\localfootnote\language\language{}{}}%
4822   \BabelFootnote\mainfootnote{}{}{}%
4823   {}

```

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.

```

4824 \IfBabelLayout{counters}%
4825   {\let\bbl@latinarabic=\@arabic
4826     \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
4827     \let\bbl@asciroman=\@roman
4828     \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
4829     \let\bbl@asciiRoman=\@Roman
4830     \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}}%
4831 \langle\textet\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).

```
4832 (*luatex)
4833 \ifx\AddBabelHook\undefined % When plain.def, babel.sty starts
4834 \bbl@trace{Read language.dat}
4835 \ifx\bbl@readstream\undefined
4836 \csname newread\endcsname\bbl@readstream
4837 \fi
4838 \begingroup
4839 \toks@{}
4840 \count@ \z@ % 0=start, 1=0th, 2=normal
4841 \def\bbl@process@line#1#2 #3 #4 {%
4842   \ifx=#1%
4843     \bbl@process@synonym{#2}%
4844   \else
4845     \bbl@process@language{#1#2}{#3}{#4}%
4846   \fi
4847   \ignorespaces}
4848 \def\bbl@manylang{%
4849   \ifnum\bbl@last>\@ne
4850     \bbl@info{Non-standard hyphenation setup}%
4851   \fi
4852   \let\bbl@manylang\relax}
4853 \def\bbl@process@language#1#2#3{%
4854   \ifcase\count@
4855     \@ifundefined{zth@#1}{\count@\tw@}{\count@\@ne}%
4856   \or
```

```

4857     \count@ \tw@
4858     \fi
4859     \ifnum \count@ = \tw@
4860         \expandafter \addlanguage \csname l@#1 \endcsname
4861         \language \allocationnumber
4862         \chardef \bbl@last \allocationnumber
4863         \bbl@many lang
4864         \let \bbl@elt \relax
4865         \xdef \bbl@languages {%
4866             \bbl@languages \bbl@elt{#1} {\the \language} {#2} {#3}} %
4867     \fi
4868     \the \toks@
4869     \toks@ {}
4870 \def \bbl@process@synonym@aux#1#2{%
4871     \global \expandafter \chardef \csname l@#1 \endcsname #2 \relax
4872     \let \bbl@elt \relax
4873     \xdef \bbl@languages {%
4874         \bbl@languages \bbl@elt{#1} {#2} {} {} } %
4875 \def \bbl@process@synonym#1{%
4876     \ifcase \count@
4877         \toks@ \expandafter {\the \toks@ \relax \bbl@process@synonym{#1}} %
4878     \or
4879         \ifundefined {zth#1} {\bbl@process@synonym@aux{#1} {0} {} } %
4880     \else
4881         \bbl@process@synonym@aux{#1} {\the \bbl@last} %
4882     \fi
4883 \ifx \bbl@languages \@undefined % Just a (sensible?) guess
4884     \chardef \l@english \z@
4885     \chardef \l@USenglish \z@
4886     \chardef \bbl@last \z@
4887     \global \@namedef {bbl@hyphendata@0} {{hyphen.tex} {}}
4888     \gdef \bbl@languages {%
4889         \bbl@elt {english} {0} {hyphen.tex} {} %
4890         \bbl@elt {USenglish} {0} {} {} }
4891 \else
4892     \global \let \bbl@languages @format \bbl@languages
4893     \def \bbl@elt#1#2#3#4% Remove all except language 0
4894         \ifnum #2 > \z@ \else
4895             \noexpand \bbl@elt{#1}{#2}{#3}{#4}%
4896         \fi}%
4897     \xdef \bbl@languages {\bbl@languages}%
4898 \fi
4899 \def \bbl@elt#1#2#3#4{\@namedef {zth#1} {}} % Define flags
4900 \bbl@languages
4901 \openin \bbl@readstream = language.dat
4902 \ifeof \bbl@readstream
4903     \bbl@warning {I couldn't find language.dat. No additional \\\%
4904         patterns loaded. Reported}%
4905 \else
4906     \loop
4907         \endlinechar \m@ne
4908         \read \bbl@readstream to \bbl@line
4909         \endlinechar ``^^M
4910         \if T \ifeof \bbl@readstream F \fi T \relax
4911         \ifx \bbl@line \empty \else
4912             \edef \bbl@line {\bbl@line \space \space \space}%
4913             \expandafter \bbl@process@line \bbl@line \relax
4914         \fi
4915     \repeat

```

```

4916 \fi
4917 \endgroup
4918 \bbl@trace{Macros for reading patterns files}
4919 \def\bbl@get@enc#1:#2:#3\@@{\def\bbl@hyph@enc{#2}}
4920 \ifx\babelcatcodetablenum\undefined
4921 \ifx\newcatcodetable\undefined
4922 \def\babelcatcodetablenum{5211}
4923 \def\bbl@pattcodes{\numexpr\babelcatcodetablenum+1\relax}
4924 \else
4925 \newcatcodetable\babelcatcodetablenum
4926 \newcatcodetable\bbl@pattcodes
4927 \fi
4928 \else
4929 \def\bbl@pattcodes{\numexpr\babelcatcodetablenum+1\relax}
4930 \fi
4931 \def\bbl@luapatterns#1#2{%
4932 \bbl@get@enc#1::\@@@
4933 \setbox\z@\hbox\bgroup
4934 \begingroup
4935 \savecatcodetable\babelcatcodetablenum\relax
4936 \initcatcodetable\bbl@pattcodes\relax
4937 \catcodetable\bbl@pattcodes\relax
4938 \catcode`\#=6 \catcode`\$=3 \catcode`\&=4 \catcode`\^=7
4939 \catcode`\_ =8 \catcode`\{=1 \catcode`\}=2 \catcode`\-=13
4940 \catcode`\@=11 \catcode`\^^I=10 \catcode`\^^J=12
4941 \catcode`\<=12 \catcode`\>=12 \catcode`\*=12 \catcode`\.=12
4942 \catcode`\-=12 \catcode`\/=12 \catcode`\[=12 \catcode`\]=12
4943 \catcode`\`=12 \catcode`\'=12 \catcode`\`=12
4944 \input #1\relax
4945 \catcodetable\babelcatcodetablenum\relax
4946 \endgroup
4947 \def\bbl@tempa{#2}%
4948 \ifx\bbl@tempa\empty\else
4949 \input #2\relax
4950 \fi
4951 \egroup}%
4952 \def\bbl@patterns@lua#1{%
4953 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
4954 \csname l@#1\endcsname
4955 \edef\bbl@tempa{#1}%
4956 \else
4957 \csname l@#1:\f@encoding\endcsname
4958 \edef\bbl@tempa{#1:\f@encoding}%
4959 \fi\relax
4960 \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
4961 \@ifundefined{bbl@hyphendata@the\language}%
4962 {\def\bbl@elt##1##2##3##4{%
4963 \ifnum##2=\csname l@#1:\f@encoding\endcsname % #2=spanish, dutch:OT1...
4964 \def\bbl@tempb{##3}%
4965 \ifx\bbl@tempb\empty\else % if not a synonymous
4966 \def\bbl@tempc{##3}{##4}%
4967 \fi
4968 \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4969 \fi}%
4970 \bbl@languages
4971 \@ifundefined{bbl@hyphendata@the\language}%
4972 {\bbl@info{No hyphenation patterns were set for\%
4973 language '\bbl@tempa'. Reported}}%
4974 {\expandafter\expandafter\expandafter\bbl@luapatterns

```

```

4975         \csname bbl@hyphendata@the\language\endcsname}}{}
4976 \endinput\fi
4977 % Here ends \ifx\AddBabelHook\undefined
4978 % A few lines are only read by hyphen.cfg
4979 \ifx\DisableBabelHook\undefined
4980   \AddBabelHook{luatex}{everylanguage}{%
4981     \def\process@language##1##2##3{%
4982       \def\process@line####1####2 ####3 ####4 {}}}
4983   \AddBabelHook{luatex}{loadpatterns}{%
4984     \input #1\relax
4985     \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
4986       {{#1}}}}
4987   \AddBabelHook{luatex}{loadexceptions}{%
4988     \input #1\relax
4989     \def\bbl@tempb##1##2{{##1}{#1}}%
4990     \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
4991       {\expandafter\expandafter\expandafter\bbl@tempb
4992         \csname bbl@hyphendata@the\language\endcsname}}
4993 \endinput\fi
4994 % Here stops reading code for hyphen.cfg
4995 % The following is read the 2nd time it's loaded
4996 \begingroup % TODO - to a lua file
4997 \catcode`\%=12
4998 \catcode`\'=12
4999 \catcode`\%=12
5000 \catcode`\:=12
5001 \directlua{
5002   Babel = Babel or {}
5003   function Babel.bytes(line)
5004     return line:gsub("(.)",
5005       function (chr) return unicode.utf8.char(string.byte(chr)) end)
5006   end
5007   function Babel.begin_process_input()
5008     if luatexbase and luatexbase.add_to_callback then
5009       luatexbase.add_to_callback('process_input_buffer',
5010         Babel.bytes, 'Babel.bytes')
5011     else
5012       Babel.callback = callback.find('process_input_buffer')
5013       callback.register('process_input_buffer', Babel.bytes)
5014     end
5015   end
5016   function Babel.end_process_input ()
5017     if luatexbase and luatexbase.remove_from_callback then
5018       luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
5019     else
5020       callback.register('process_input_buffer', Babel.callback)
5021     end
5022   end
5023   function Babel.addpatterns(pp, lg)
5024     local lg = lang.new(lg)
5025     local pats = lang.patterns(lg) or ''
5026     lang.clear_patterns(lg)
5027     for p in pp:gmatch('[^%s]+') do
5028       ss = ''
5029       for i in string.utfcharacters(p:gsub('%d', '')) do
5030         ss = ss .. '%d?' .. i
5031       end
5032       ss = ss:gsub('^%%d%?%', '%%.') .. '%d?'
5033       ss = ss:gsub('%.%d%?$', '%%.')

```

```

5034     pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
5035     if n == 0 then
5036         tex.sprint(
5037             [[\string\csname\space bbl@info\endcsname{New pattern: }]]
5038             .. p .. [[]]])
5039     pats = pats .. ' ' .. p
5040     else
5041         tex.sprint(
5042             [[\string\csname\space bbl@info\endcsname{Renew pattern: }]]
5043             .. p .. [[]]])
5044     end
5045 end
5046 lang.patterns(lg, pats)
5047 end
5048 }
5049 \endgroup
5050 \ifx\newattribute\@undefined\else
5051 \newattribute\bbl@attr@locale
5052 \directlua{ Babel.attr_locale = luatexbase.registernumber'bbl@attr@locale'}
5053 \AddBabelHook{luatex}{beforeextras}{%
5054     \setattribute\bbl@attr@locale\localeid}
5055 \fi
5056 \def\BabelStringsDefault{unicode}
5057 \let\luabbbl@stop\relax
5058 \AddBabelHook{luatex}{encodedcommands}{%
5059     \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
5060     \ifx\bbl@tempa\bbl@tempb\else
5061         \directlua{Babel.begin_process_input()}%
5062         \def\luabbbl@stop{%
5063             \directlua{Babel.end_process_input()}}%
5064     \fi}%
5065 \AddBabelHook{luatex}{stopcommands}{%
5066     \luabbbl@stop
5067     \let\luabbbl@stop\relax}
5068 \AddBabelHook{luatex}{patterns}{%
5069     \@ifundefined{bbl@hyphendata@the\language}%
5070     {\def\bbl@elt##1##2##3##4{%
5071         \ifnum##2=\csname l@##2\endcsname % #2=spanish, dutch:OT1...
5072         \def\bbl@tempb{##3}%
5073         \ifx\bbl@tempb@empty\else % if not a synonymous
5074             \def\bbl@tempc{##3}{##4}}%
5075         \fi
5076         \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
5077         \fi}%
5078     \bbl@languages
5079     \@ifundefined{bbl@hyphendata@the\language}%
5080     {\bbl@info{No hyphenation patterns were set for\%
5081         language '#2'. Reported}}%
5082     {\expandafter\expandafter\expandafter\bbl@luapatterns
5083         \csname bbl@hyphendata@the\language\endcsname}}}%
5084 \@ifundefined{bbl@patterns@}{}%
5085 \begingroup
5086 \bbl@xin@{,\number\language,}{,\bbl@pttnlist}%
5087 \ifin@else
5088     \ifx\bbl@patterns@empty\else
5089         \directlua{ Babel.addpatterns(
5090             [[\bbl@patterns@]], \number\language) }%
5091     \fi
5092     \@ifundefined{bbl@patterns@#1}%

```

```

5093         \@empty
5094         {\directlua{ Babel.addpatterns(
5095             [[\space\csname bbl@patterns@#1\endcsname]],
5096             \number\language) }}%
5097         \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
5098     \fi
5099 \endgroup}%
5100 \bbl@exp{%
5101     \bbl@ifunset{\bbl@prehc@\languagename}{}%
5102     {\bbl@ifblank{\bbl@cs{\prehc@\languagename}}{}}%
5103     {\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.

```

5104 \@onlypreamble\babelpatterns
5105 \AtEndOfPackage{%
5106     \newcommand\babelpatterns[2][\@empty]{%
5107         \ifx\bbl@patterns\relax
5108             \let\bbl@patterns@\@empty
5109         \fi
5110         \ifx\bbl@pttnlist\@empty\else
5111             \bbl@warning{%
5112                 You must not intermingle \string\selectlanguage\space and\%
5113                 \string\babelpatterns\space or some patterns will not\%
5114                 be taken into account. Reported}%
5115             \fi
5116             \ifx\@empty#1%
5117                 \protected@edef\bbl@patterns@{\bbl@patterns@\space#2}%
5118             \else
5119                 \edef\bbl@tempb{\zap@space#1 \@empty}%
5120                 \bbl@for\bbl@tempa\bbl@tempb{%
5121                     \bbl@fixname\bbl@tempa
5122                     \bbl@iflanguage\bbl@tempa{%
5123                         \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
5124                             \@ifundefined{\bbl@patterns@\bbl@tempa}%
5125                             \@empty
5126                             {\csname bbl@patterns@\bbl@tempa\endcsname\space}%
5127                             #2}}}%
5128             \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.

```

5129% TODO - to a lua file
5130 \directlua{
5131     Babel = Babel or {}
5132     Babel.linebreaking = Babel.linebreaking or {}
5133     Babel.linebreaking.before = {}
5134     Babel.linebreaking.after = {}
5135     Babel.locale = {} % Free to use, indexed by \localeid
5136     function Babel.linebreaking.add_before(func)
5137         tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
5138         table.insert(Babel.linebreaking.before, func)
5139     end

```



```

5140 function Babel.linebreaking.add_after(func)
5141     tex.print([[ \noexpand\csname bbl@luahyphenate\endcsname ]])
5142     table.insert(Babel.linebreaking.after, func)
5143 end
5144 }
5145 \def\bbl@intraspace#1 #2 #3\@@{%
5146     \directlua{
5147         Babel = Babel or {}
5148         Babel.intraspaces = Babel.intraspaces or {}
5149         Babel.intraspaces['\csname bbl@sbc@ \language\endcsname'] = %
5150             {b = #1, p = #2, m = #3}
5151         Babel.locale_props[\the\localeid].intraspace = %
5152             {b = #1, p = #2, m = #3}
5153     }}
5154 \def\bbl@intrapenalty#1\@@{%
5155     \directlua{
5156         Babel = Babel or {}
5157         Babel.intrapenalties = Babel.intrapenalties or {}
5158         Babel.intrapenalties['\csname bbl@sbc@ \language\endcsname'] = #1
5159         Babel.locale_props[\the\localeid].intrapenalty = #1
5160     }}
5161 \begingroup
5162 \catcode`\%=12
5163 \catcode`\^=14
5164 \catcode`\'=12
5165 \catcode`\~=12
5166 \gdef\bbl@seaintraspace{^
5167     \let\bbl@seaintraspace\relax
5168     \directlua{
5169         Babel = Babel or {}
5170         Babel.sea_enabled = true
5171         Babel.sea_ranges = Babel.sea_ranges or {}
5172         function Babel.set_chrngs (script, chrng)
5173             local c = 0
5174             for s, e in string.gmatch(chrng..' ', '(.-%.%.(-)%s') do
5175                 Babel.sea_ranges[script..c]={tonumber(s,16), tonumber(e,16)}
5176                 c = c + 1
5177             end
5178         end
5179         function Babel.sea_disc_to_space (head)
5180             local sea_ranges = Babel.sea_ranges
5181             local last_char = nil
5182             local quad = 655360      ^% 10 pt = 655360 = 10 * 65536
5183             for item in node.traverse(head) do
5184                 local i = item.id
5185                 if i == node.id'glyph' then
5186                     last_char = item
5187                 elseif i == 7 and item.subtype == 3 and last_char
5188                     and last_char.char > 0x0C99 then
5189                     quad = font.getfont(last_char.font).size
5190                     for lg, rg in pairs(sea_ranges) do
5191                         if last_char.char > rg[1] and last_char.char < rg[2] then
5192                             lg = lg:sub(1, 4)  ^% Remove trailing number of, eg, Cyril1
5193                             local intraspace = Babel.intraspaces[lg]
5194                             local intrapenalty = Babel.intrapenalties[lg]
5195                             local n
5196                             if intrapenalty ~= 0 then
5197                                 n = node.new(14, 0)      ^% penalty
5198                                 n.penalty = intrapenalty

```

```

5199         node.insert_before(head, item, n)
5200     end
5201     n = node.new(12, 13)      ^% (glue, spaceskip)
5202     node.setglue(n, intraspace.b * quad,
5203                 intraspace.p * quad,
5204                 intraspace.m * quad)
5205     node.insert_before(head, item, n)
5206     node.remove(head, item)
5207 end
5208 end
5209 end
5210 end
5211 end
5212 }^^
5213 \bbl@luahyphenate}

```

### 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.

```

5214 \catcode`\%=14
5215 \gdef\bbl@cjkintraspacespace{%
5216   \let\bbl@cjkintraspacespace\relax
5217   \directlua{
5218     Babel = Babel or {}
5219     require('babel-data-cjk.lua')
5220     Babel.cjk_enabled = true
5221     function Babel.cjk_linebreak(head)
5222       local GLYPH = node.id'glyph'
5223       local last_char = nil
5224       local quad = 655360      % 10 pt = 655360 = 10 * 65536
5225       local last_class = nil
5226       local last_lang = nil
5227
5228       for item in node.traverse(head) do
5229         if item.id == GLYPH then
5230
5231           local lang = item.lang
5232
5233           local LOCALE = node.get_attribute(item,
5234             luatexbase.registernumber'bbl@attr@locale')
5235           local props = Babel.locale_props[LOCALE]
5236
5237           local class = Babel.cjk_class[item.char].c
5238
5239           if class == 'cp' then class = 'cl' end % )] as CL
5240           if class == 'id' then class = 'I' end
5241
5242           local br = 0
5243           if class and last_class and Babel.cjk_breaks[last_class][class] then
5244             br = Babel.cjk_breaks[last_class][class]
5245           end
5246
5247           if br == 1 and props.linebreak == 'c' and

```

```

5248         lang ~= \the\l@nohyphenation\space and
5249         last_lang ~= \the\l@nohyphenation then
5250         local intrapenalty = props.intrapenalty
5251         if intrapenalty ~= 0 then
5252             local n = node.new(14, 0)      % penalty
5253             n.penalty = intrapenalty
5254             node.insert_before(head, item, n)
5255         end
5256         local intraspace = props.intraspace
5257         local n = node.new(12, 13)        % (glue, spaceskip)
5258         node.setglue(n, intraspace.b * quad,
5259                       intraspace.p * quad,
5260                       intraspace.m * quad)
5261         node.insert_before(head, item, n)
5262     end
5263
5264     if font.getfont(item.font) then
5265         quad = font.getfont(item.font).size
5266     end
5267     last_class = class
5268     last_lang = lang
5269     else % if penalty, glue or anything else
5270         last_class = nil
5271     end
5272 end
5273 lang.hyphenate(head)
5274 end
5275 }%
5276 \bbl@luahyphenate}
5277 \gdef\bbl@luahyphenate{%
5278 \let\bbl@luahyphenate\relax
5279 \directlua{
5280     luatexbase.add_to_callback('hyphenate',
5281     function (head, tail)
5282         if Babel.linebreaking.before then
5283             for k, func in ipairs(Babel.linebreaking.before) do
5284                 func(head)
5285             end
5286         end
5287         if Babel.cjk_enabled then
5288             Babel.cjk_linebreak(head)
5289         end
5290         lang.hyphenate(head)
5291         if Babel.linebreaking.after then
5292             for k, func in ipairs(Babel.linebreaking.after) do
5293                 func(head)
5294             end
5295         end
5296         if Babel.sea_enabled then
5297             Babel.sea_disc_to_space(head)
5298         end
5299     end,
5300     'Babel.hyphenate')
5301 }
5302 }
5303 \endgroup
5304 \def\bbl@provide@intraspace{%
5305 \bbl@ifunset{\bbl@intsp@languagename}{}%
5306 {\expandafter\ifx\curname\bbl@intsp@languagename\endcsname\@empty\else

```

```

5307 \bbl@xin@{/c}{/\bbl@cl{lnbrk}}%
5308 \ifin@ % cjk
5309 \bbl@cjkintraspacespace
5310 \directlua{
5311     Babel = Babel or {}
5312     Babel.locale_props = Babel.locale_props or {}
5313     Babel.locale_props[\the\localeid].linebreak = 'c'
5314 }%
5315 \bbl@exp{\\bbl@intraspacespace\bbl@cl{intsp}\\@@}%
5316 \ifx\bbl@KVP@intrapenalty\@nil
5317     \bbl@intrapenalty0\@@
5318 \fi
5319 \else % sea
5320     \bbl@seaintraspacespace
5321     \bbl@exp{\\bbl@intraspacespace\bbl@cl{intsp}\\@@}%
5322     \directlua{
5323         Babel = Babel or {}
5324         Babel.sea_ranges = Babel.sea_ranges or {}
5325         Babel.set_chranges('\bbl@cl{sbcpr}',
5326                             '\bbl@cl{chrng}')
5327     }%
5328     \ifx\bbl@KVP@intrapenalty\@nil
5329         \bbl@intrapenalty0\@@
5330     \fi
5331 \fi
5332 \fi
5333 \ifx\bbl@KVP@intrapenalty\@nil\else
5334     \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@@
5335 \fi}}

```

## 13.6 Arabic justification

```

5336 \ifnum\bbl@bidimode>100 \ifnum\bbl@bidimode<200
5337 \def\bblar@chars{%
5338     0628,0629,062A,062B,062C,062D,062E,062F,0630,0631,0632,0633,%
5339     0634,0635,0636,0637,0638,0639,063A,063B,063C,063D,063E,063F,%
5340     0640,0641,0642,0643,0644,0645,0646,0647,0649}
5341 \def\bblar@elongated{%
5342     0626,0628,062A,062B,0633,0634,0635,0636,063B,%
5343     063C,063D,063E,063F,0641,0642,0643,0644,0646,%
5344     0649,064A}
5345 \begingroup
5346 \catcode`\_ =11 \catcode`\:=11
5347 \gdef\bblar@nofswarn{\gdef\msg_warning:nx##1##2##3{}}
5348 \endgroup
5349 \gdef\bbl@arabicjust{%
5350     \let\bbl@arabicjust\relax
5351     \newattribute\bblar@kashida
5352     \bblar@kashida=\z@
5353     \expandafter\bbl@add\csname selectfont \endcsname{\bbl@parsejalt}}%
5354     \directlua{
5355         Babel.arabic.elong_map = Babel.arabic.elong_map or {}
5356         Babel.arabic.elong_map[\the\localeid] = {}
5357         luatexbase.add_to_callback('post_linebreak_filter',
5358             Babel.arabic.justify, 'Babel.arabic.justify')
5359         luatexbase.add_to_callback('hpack_filter',
5360             Babel.arabic.justify_hbox, 'Babel.arabic.justify_hbox')
5361     }}%
5362 % Save both node lists to make replacement. TODO. Save also widths to

```

```

5363 % make computations
5364 \def\bblar@fetchjalt#1#2#3#4{%
5365   \bbl@exp{\bbl@foreach{#1}}{%
5366     \bbl@ifunset\bblar@JE@##1}%
5367     {\setbox\z@\hbox{^^^200d\char"##1#2}}%
5368     {\setbox\z@\hbox{^^^200d\char"@nameuse\bblar@JE@##1#2}}%
5369   \directlua{%
5370     local last = nil
5371     for item in node.traverse(tex.box[0].head) do
5372       if item.id == node.id'glyph' and item.char > 0x600 and
5373         not (item.char == 0x200D) then
5374         last = item
5375       end
5376     end
5377     Babel.arabic.#3['##1#4'] = last.char
5378   }}
5379 % Brute force. No rules at all, yet. The ideal: look at jalt table. And
5380 % perhaps other tables (falt?, csw?). What about kaf? And diacritic
5381 % positioning?
5382 \gdef\bbl@parsejalt{%
5383   \ifx\addfontfeature\undefined\else
5384     \bbl@xin@{/e}{/\bbl@cl{lbrk}}%
5385     \ifin@
5386       \directlua{%
5387         if Babel.arabic.elong_map[\the\localeid][\fontid\font] == nil then
5388           Babel.arabic.elong_map[\the\localeid][\fontid\font] = {}
5389           tex.print([[string\csname\space\bbl@parsejalt\endcsname]])
5390         end
5391       }%
5392     \fi
5393   \fi}
5394 \gdef\bbl@parsejalti{%
5395   \begingroup
5396     \let\bbl@parsejalt\relax % To avoid infinite loop
5397     \edef\bbl@tempb{\fontid\font}%
5398     \bblar@nofswarn
5399     \bblar@fetchjalt\bblar@elongated{}{from}{}%
5400     \bblar@fetchjalt\bblar@chars{^^^064a}{from}{a}% Alef maksura
5401     \bblar@fetchjalt\bblar@chars{^^^0649}{from}{y}% Yeh
5402     \addfontfeature{RawFeature+=jalt}%
5403     % \namedef\bblar@JE@0643{06AA}% todo: catch medial kaf
5404     \bblar@fetchjalt\bblar@elongated{}{dest}{}%
5405     \bblar@fetchjalt\bblar@chars{^^^064a}{dest}{a}%
5406     \bblar@fetchjalt\bblar@chars{^^^0649}{dest}{y}%
5407     \directlua{%
5408       for k, v in pairs(Babel.arabic.from) do
5409         if Babel.arabic.dest[k] and
5410           not (Babel.arabic.from[k] == Babel.arabic.dest[k]) then
5411           Babel.arabic.elong_map[\the\localeid][\bbl@tempb]
5412             [Babel.arabic.from[k]] = Babel.arabic.dest[k]
5413         end
5414       end
5415     }%
5416   \endgroup}
5417 %
5418 \begingroup
5419 \catcode`#=11
5420 \catcode`~ =11
5421 \directlua{

```

```

5422
5423 Babel.arabic = Babel.arabic or {}
5424 Babel.arabic.from = {}
5425 Babel.arabic.dest = {}
5426 Babel.arabic.justify_factor = 0.95
5427 Babel.arabic.justify_enabled = true
5428
5429 function Babel.arabic.justify(head)
5430   if not Babel.arabic.justify_enabled then return head end
5431   for line in node.traverse_id(node.id'hlist', head) do
5432     Babel.arabic.justify_hlist(head, line)
5433   end
5434   return head
5435 end
5436
5437 function Babel.arabic.justify_hbox(head, gc, size, pack)
5438   local has_inf = false
5439   if Babel.arabic.justify_enabled and pack == 'exactly' then
5440     for n in node.traverse_id(12, head) do
5441       if n.stretch_order > 0 then has_inf = true end
5442     end
5443     if not has_inf then
5444       Babel.arabic.justify_hlist(head, nil, gc, size, pack)
5445     end
5446   end
5447   return head
5448 end
5449
5450 function Babel.arabic.justify_hlist(head, line, gc, size, pack)
5451   local d, new
5452   local k_list, k_item, pos_inline
5453   local width, width_new, full, k_curr, wt_pos, goal, shift
5454   local subst_done = false
5455   local elong_map = Babel.arabic.elong_map
5456   local last_line
5457   local GLYPH = node.id'glyph'
5458   local KASHIDA = luatexbase.registernumber'bblar@kashida'
5459   local LOCALE = luatexbase.registernumber'bbl@attr@locale'
5460
5461   if line == nil then
5462     line = {}
5463     line.glue_sign = 1
5464     line.glue_order = 0
5465     line.head = head
5466     line.shift = 0
5467     line.width = size
5468   end
5469
5470   % Exclude last line. todo. But-- it discards one-word lines, too!
5471   % ? Look for glue = 12:15
5472   if (line.glue_sign == 1 and line.glue_order == 0) then
5473     elongs = {} % Stores elongated candidates of each line
5474     k_list = {} % And all letters with kashida
5475     pos_inline = 0 % Not yet used
5476
5477     for n in node.traverse_id(GLYPH, line.head) do
5478       pos_inline = pos_inline + 1 % To find where it is. Not used.
5479
5480       % Elongated glyphs

```

```

5481     if elong_map then
5482         local locale = node.get_attribute(n, LOCALE)
5483         if elong_map[locale] and elong_map[locale][n.font] and
5484             elong_map[locale][n.font][n.char] then
5485             table.insert(elongs, {node = n, locale = locale} )
5486             node.set_attribute(n.prev, KASHIDA, 0)
5487         end
5488     end
5489
5490     % Tatwil
5491     if Babel.kashida_wts then
5492         local k_wt = node.get_attribute(n, KASHIDA)
5493         if k_wt > 0 then % todo. parameter for multi inserts
5494             table.insert(k_list, {node = n, weight = k_wt, pos = pos_inline})
5495         end
5496     end
5497
5498     end % of node.traverse_id
5499
5500     if #elongs == 0 and #k_list == 0 then goto next_line end
5501     full = line.width
5502     shift = line.shift
5503     goal = full * Babel.arabic.justify_factor % A bit crude
5504     width = node.dimensions(line.head) % The 'natural' width
5505
5506     % == Elongated ==
5507     % Original idea taken from 'chickenize'
5508     while (#elongs > 0 and width < goal) do
5509         subst_done = true
5510         local x = #elongs
5511         local curr = elongs[x].node
5512         local oldchar = curr.char
5513         curr.char = elong_map[elongs[x].locale][curr.font][curr.char]
5514         width = node.dimensions(line.head) % Check if the line is too wide
5515         % Substitute back if the line would be too wide and break:
5516         if width > goal then
5517             curr.char = oldchar
5518             break
5519         end
5520         % If continue, pop the just substituted node from the list:
5521         table.remove(elongs, x)
5522     end
5523
5524     % == Tatwil ==
5525     if #k_list == 0 then goto next_line end
5526
5527     width = node.dimensions(line.head) % The 'natural' width
5528     k_curr = #k_list
5529     wt_pos = 1
5530
5531     while width < goal do
5532         subst_done = true
5533         k_item = k_list[k_curr].node
5534         if k_list[k_curr].weight == Babel.kashida_wts[wt_pos] then
5535             d = node.copy(k_item)
5536             d.char = 0x0640
5537             line.head, new = node.insert_after(line.head, k_item, d)
5538             width_new = node.dimensions(line.head)
5539             if width > goal or width == width_new then

```

```

5540         node.remove(line.head, new) % Better compute before
5541         break
5542     end
5543     width = width_new
5544 end
5545 if k_curr == 1 then
5546     k_curr = #k_list
5547     wt_pos = (wt_pos >= table.getn(Babel.kashida_wts)) and 1 or wt_pos+1
5548 else
5549     k_curr = k_curr - 1
5550 end
5551 end
5552
5553 ::next_line::
5554
5555 % Must take into account marks and ins, see luatex manual.
5556 % Have to be executed only if there are changes. Investigate
5557 % what's going on exactly.
5558 if subst_done and not gc then
5559     d = node.hpack(line.head, full, 'exactly')
5560     d.shift = shift
5561     node.insert_before(head, line, d)
5562     node.remove(head, line)
5563 end
5564 end % if process line
5565 end
5566 }
5567 \endgroup
5568 \fi\fi % Arabic just block

```

### 13.7 Common stuff

```

5569 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
5570 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ckeckstdfonts}
5571 \DisableBabelHook{babel-fontspec}
5572 <<Font selection>>

```

### 13.8 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 form 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.

```

5573 % TODO - to a lua file
5574 \directlua{
5575 Babel.script_blocks = {
5576   ['dflt'] = {},
5577   ['Arab'] = {{0x0600, 0x06FF}, {0x08A0, 0x08FF}, {0x0750, 0x077F},
5578             {0xFE70, 0xFEFF}, {0xFB50, 0xFDFF}, {0x1EE00, 0x1EEFF}},
5579   ['Armn'] = {{0x0530, 0x058F}},
5580   ['Beng'] = {{0x0980, 0x09FF}},
5581   ['Cher'] = {{0x13A0, 0x13FF}, {0xAB70, 0xABBF}},
5582   ['Copt'] = {{0x03E2, 0x03EF}, {0x2C80, 0x2CFF}, {0x102E0, 0x102FF}},
5583   ['Cyr1'] = {{0x0400, 0x04FF}, {0x0500, 0x052F}, {0x1C80, 0x1C8F},
5584             {0x2DE0, 0x2DFF}, {0xA640, 0xA69F}},
5585   ['Deva'] = {{0x0900, 0x097F}, {0xA8E0, 0xA8FF}},
5586   ['Ethi'] = {{0x1200, 0x137F}, {0x1380, 0x139F}, {0x2D80, 0x2DDF}},

```



```

5587         {0xAB00, 0xAB2F}},
5588 ['Geor'] = {{0x10A0, 0x10FF}, {0x2D00, 0x2D2F}},
5589 % Don't follow strictly Unicode, which places some Coptic letters in
5590 % the 'Greek and Coptic' block
5591 ['Grek'] = {{0x0370, 0x03E1}, {0x03F0, 0x03FF}, {0x1F00, 0x1FFF}},
5592 ['Hans'] = {{0x2E80, 0x2EFF}, {0x3000, 0x303F}, {0x31C0, 0x31EF},
5593             {0x3300, 0x33FF}, {0x3400, 0x4DBF}, {0x4E00, 0x9FFF},
5594             {0xF900, 0xFAFF}, {0xFE30, 0xFE4F}, {0xFF00, 0xFFEF},
5595             {0x20000, 0x2A6DF}, {0x2A700, 0x2B73F},
5596             {0x2B740, 0x2B81F}, {0x2B820, 0x2CEAF},
5597             {0x2CEB0, 0x2EBEF}, {0x2F800, 0x2FA1F}},
5598 ['Hebr'] = {{0x0590, 0x05FF}},
5599 ['Jpan'] = {{0x3000, 0x303F}, {0x3040, 0x309F}, {0x30A0, 0x30FF},
5600             {0x4E00, 0x9FAF}, {0xFF00, 0xFFEF}},
5601 ['Khmr'] = {{0x1780, 0x17FF}, {0x19E0, 0x19FF}},
5602 ['Knda'] = {{0x0C80, 0x0CFF}},
5603 ['Kore'] = {{0x1100, 0x11FF}, {0x3000, 0x303F}, {0x3130, 0x318F},
5604             {0x4E00, 0x9FAF}, {0xA960, 0xA97F}, {0xAC00, 0xD7AF},
5605             {0xD7B0, 0xD7FF}, {0xFF00, 0xFFEF}},
5606 ['Lao'] = {{0x0E80, 0x0EFF}},
5607 ['Latn'] = {{0x0000, 0x007F}, {0x0080, 0x00FF}, {0x0100, 0x017F},
5608             {0x0180, 0x024F}, {0x1E00, 0x1EFF}, {0x2C60, 0x2C7F},
5609             {0xA720, 0xA7FF}, {0xAB30, 0xAB6F}},
5610 ['Mahj'] = {{0x11150, 0x1117F}},
5611 ['Mlym'] = {{0x0D00, 0x0D7F}},
5612 ['Mymr'] = {{0x1000, 0x109F}, {0xAA60, 0xAA7F}, {0xA9E0, 0xA9FF}},
5613 ['Orya'] = {{0x0B00, 0x0B7F}},
5614 ['Sinh'] = {{0x0D80, 0x0DFF}, {0x111E0, 0x111FF}},
5615 ['Syr'] = {{0x0700, 0x074F}, {0x0860, 0x086F}},
5616 ['Taml'] = {{0x0B80, 0x0BFF}},
5617 ['Telu'] = {{0x0C00, 0x0C7F}},
5618 ['Tfng'] = {{0x2D30, 0x2D7F}},
5619 ['Thai'] = {{0x0E00, 0x0E7F}},
5620 ['Tibt'] = {{0x0F00, 0x0FFF}},
5621 ['Vaii'] = {{0xA500, 0xA63F}},
5622 ['Yiii'] = {{0xA000, 0xA48F}, {0xA490, 0xA4CF}}
5623 }
5624
5625 Babel.script_blocks.Cyrs = Babel.script_blocks.Cyrl
5626 Babel.script_blocks.Hant = Babel.script_blocks.Hans
5627 Babel.script_blocks.Kana = Babel.script_blocks.Jpan
5628
5629 function Babel.locale_map(head)
5630   if not Babel.locale_mapped then return head end
5631
5632   local LOCALE = luatexbase.registernumber'bbl@attr@locale'
5633   local GLYPH = node.id('glyph')
5634   local inmath = false
5635   local toloc_save
5636   for item in node.traverse(head) do
5637     local toloc
5638     if not inmath and item.id == GLYPH then
5639       % Optimization: build a table with the chars found
5640       if Babel.chr_to_loc[item.char] then
5641         toloc = Babel.chr_to_loc[item.char]
5642       else
5643         for lc, maps in pairs(Babel.loc_to_scr) do
5644           for _, rg in pairs(maps) do
5645             if item.char >= rg[1] and item.char <= rg[2] then

```

```

5646         Babel.chr_to_loc[item.char] = lc
5647         toloc = lc
5648         break
5649     end
5650 end
5651 end
5652 end
5653 % Now, take action, but treat composite chars in a different
5654 % fashion, because they 'inherit' the previous locale. Not yet
5655 % optimized.
5656 if not toloc and
5657     (item.char >= 0x0300 and item.char <= 0x036F) or
5658     (item.char >= 0x1AB0 and item.char <= 0x1AFF) or
5659     (item.char >= 0x1DC0 and item.char <= 0x1DFF) then
5660     toloc = toloc_save
5661 end
5662 if toloc and toloc > -1 then
5663     if Babel.locale_props[toloc].lg then
5664         item.lang = Babel.locale_props[toloc].lg
5665         node.set_attribute(item, LOCALE, toloc)
5666     end
5667     if Babel.locale_props[toloc]['/'..item.font] then
5668         item.font = Babel.locale_props[toloc]['/'..item.font]
5669     end
5670     toloc_save = toloc
5671 end
5672 elseif not inmath and item.id == 7 then
5673     item.replace = item.replace and Babel.locale_map(item.replace)
5674     item.pre      = item.pre and Babel.locale_map(item.pre)
5675     item.post     = item.post and Babel.locale_map(item.post)
5676 elseif item.id == node.id'math' then
5677     inmath = (item.subtype == 0)
5678 end
5679 end
5680 return head
5681 end
5682 }

```

The code for \babelcharproperty is straightforward. Just note the modified lua table can be different.

```

5683 \newcommand\babelcharproperty[1]{%
5684   \count@=#1\relax
5685   \ifvmode
5686     \expandafter\bbl@chprop
5687   \else
5688     \bbl@error{\string\babelcharproperty\space can be used only in\\%
5689       vertical mode (preamble or between paragraphs)}%
5690     {See the manual for futher info}%
5691   \fi}
5692 \newcommand\bbl@chprop[3][\the\count@]{%
5693   \@tempcnta=#1\relax
5694   \bbl@ifunset{\bbl@chprop@#2}%
5695   {\bbl@error{No property named '#2'. Allowed values are\\%
5696     direction (bc), mirror (bmg), and linebreak (lb)}%
5697     {See the manual for futher info}}%
5698   }%
5699   \loop
5700     \bbl@cs{chprop@#2}{#3}%
5701     \ifnum\count@<\@tempcnta

```

```

5702 \advance\count@\ne
5703 \repeat}
5704 \def\bbbl@chprop@direction#1{%
5705 \directlua{
5706   Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
5707   Babel.characters[\the\count@]['d'] = '#1'
5708 }}
5709 \let\bbbl@chprop@bc\bbbl@chprop@direction
5710 \def\bbbl@chprop@mirror#1{%
5711 \directlua{
5712   Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
5713   Babel.characters[\the\count@]['m'] = '\number#1'
5714 }}
5715 \let\bbbl@chprop@bmg\bbbl@chprop@mirror
5716 \def\bbbl@chprop@linebreak#1{%
5717 \directlua{
5718   Babel.cjk_characters[\the\count@] = Babel.cjk_characters[\the\count@] or {}
5719   Babel.cjk_characters[\the\count@]['c'] = '#1'
5720 }}
5721 \let\bbbl@chprop@lb\bbbl@chprop@linebreak
5722 \def\bbbl@chprop@locale#1{%
5723 \directlua{
5724   Babel.chr_to_loc = Babel.chr_to_loc or {}
5725   Babel.chr_to_loc[\the\count@] =
5726     \bbbl@ifblank{#1}{-1000}{\the\bbbl@cs{id@#1}}\space
5727 }}

```

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.

```

5728 \begingroup % TODO - to a lua file
5729 \catcode`\~=12
5730 \catcode`\#=12
5731 \catcode`\%=12
5732 \catcode`\&=14
5733 \directlua{
5734   Babel.linebreaking.replacements = {}
5735   Babel.linebreaking.replacements[0] = {} &% pre
5736   Babel.linebreaking.replacements[1] = {} &% post
5737
5738   &% Discretionaries contain strings as nodes
5739   function Babel.str_to_nodes(fn, matches, base)
5740     local n, head, last
5741     if fn == nil then return nil end
5742     for s in string.utfvalues(fn(matches)) do
5743       if base.id == 7 then
5744         base = base.replace
5745       end
5746       n = node.copy(base)

```

```

5747     n.char    = s
5748     if not head then
5749         head = n
5750     else
5751         last.next = n
5752     end
5753     last = n
5754 end
5755 return head
5756 end
5757
5758 Babel.fetch_subtext = {}
5759
5760 Babel.ignore_pre_char = function(node)
5761     return (node.lang == \the\l@nohyphenation)
5762 end
5763
5764 %% Merging both functions doesn't seem feasible, because there are too
5765 %% many differences.
5766 Babel.fetch_subtext[0] = function(head)
5767     local word_string = ''
5768     local word_nodes = {}
5769     local lang
5770     local item = head
5771     local inmath = false
5772
5773     while item do
5774
5775         if item.id == 11 then
5776             inmath = (item.subtype == 0)
5777         end
5778
5779         if inmath then
5780             %% pass
5781
5782         elseif item.id == 29 then
5783             local locale = node.get_attribute(item, Babel.attr_locale)
5784
5785             if lang == locale or lang == nil then
5786                 lang = lang or locale
5787                 if Babel.ignore_pre_char(item) then
5788                     word_string = word_string .. Babel.us_char
5789                 else
5790                     word_string = word_string .. unicode.utf8.char(item.char)
5791                 end
5792                 word_nodes[#word_nodes+1] = item
5793             else
5794                 break
5795             end
5796
5797         elseif item.id == 12 and item.subtype == 13 then
5798             word_string = word_string .. ' '
5799             word_nodes[#word_nodes+1] = item
5800
5801             %% Ignore leading unrecognized nodes, too.
5802         elseif word_string ~= '' then
5803             word_string = word_string .. Babel.us_char
5804             word_nodes[#word_nodes+1] = item %% Will be ignored
5805         end

```

```

5806
5807     item = item.next
5808 end
5809
5810     %% Here and above we remove some trailing chars but not the
5811     %% corresponding nodes. But they aren't accessed.
5812     if word_string:sub(-1) == ' ' then
5813         word_string = word_string:sub(1,-2)
5814     end
5815     word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
5816     return word_string, word_nodes, item, lang
5817 end
5818
5819 Babel.fetch_subtext[1] = function(head)
5820     local word_string = ''
5821     local word_nodes = {}
5822     local lang
5823     local item = head
5824     local inmath = false
5825
5826     while item do
5827
5828         if item.id == 11 then
5829             inmath = (item.subtype == 0)
5830         end
5831
5832         if inmath then
5833             %% pass
5834
5835         elseif item.id == 29 then
5836             if item.lang == lang or lang == nil then
5837                 if (item.char ~= 124) and (item.char ~= 61) then %% not =, not |
5838                     lang = lang or item.lang
5839                     word_string = word_string .. unicode.utf8.char(item.char)
5840                     word_nodes[#word_nodes+1] = item
5841                 end
5842             else
5843                 break
5844             end
5845
5846         elseif item.id == 7 and item.subtype == 2 then
5847             word_string = word_string .. '='
5848             word_nodes[#word_nodes+1] = item
5849
5850         elseif item.id == 7 and item.subtype == 3 then
5851             word_string = word_string .. '|'
5852             word_nodes[#word_nodes+1] = item
5853
5854             %% (1) Go to next word if nothing was found, and (2) implicitly
5855             %% remove leading USs.
5856             elseif word_string == '' then
5857                 %% pass
5858
5859             %% This is the responsible for splitting by words.
5860             elseif (item.id == 12 and item.subtype == 13) then
5861                 break
5862
5863             else
5864                 word_string = word_string .. Babel.us_char

```

```

5865         word_nodes[#word_nodes+1] = item  &% Will be ignored
5866     end
5867
5868     item = item.next
5869 end
5870
5871     word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
5872     return word_string, word_nodes, item, lang
5873 end
5874
5875 function Babel.pre_hyphenate_replace(head)
5876     Babel.hyphenate_replace(head, 0)
5877 end
5878
5879 function Babel.post_hyphenate_replace(head)
5880     Babel.hyphenate_replace(head, 1)
5881 end
5882
5883 function Babel.debug_hyph(w, wn, sc, first, last, last_match)
5884     local ss = ''
5885     for pp = 1, 40 do
5886         if wn[pp] then
5887             if wn[pp].id == 29 then
5888                 ss = ss .. unicode.utf8.char(wn[pp].char)
5889             else
5890                 ss = ss .. '{' .. wn[pp].id .. '}'
5891             end
5892         end
5893     end
5894     print('nod', ss)
5895     print('lst_m',
5896         string.rep(' ', unicode.utf8.len(
5897             string.sub(w, 1, last_match))-1) .. '>')
5898     print('str', w)
5899     print('sc', string.rep(' ', sc-1) .. '^')
5900     if first == last then
5901         print('f=l', string.rep(' ', first-1) .. '!')
5902     else
5903         print('f/l', string.rep(' ', first-1) .. '[' ..
5904             string.rep(' ', last-first-1) .. ']')
5905     end
5906 end
5907
5908 Babel.us_char = string.char(31)
5909
5910 function Babel.hyphenate_replace(head, mode)
5911     local u = unicode.utf8
5912     local lbkr = Babel.linebreaking.replacements[mode]
5913
5914     local word_head = head
5915
5916     while true do  &% for each subtext block
5917
5918         local w, w_nodes, nw, lang = Babel.fetch_subtext[mode](word_head)
5919
5920         if Babel.debug then
5921             print()
5922             print((mode == 0) and '@@@<' or '@@@>', w)
5923         end

```

```

5924
5925     if nw == nil and w == '' then break end
5926
5927     if not lang then goto next end
5928     if not lbkr[lang] then goto next end
5929
5930     %% For each saved (pre|post)hyphenation. TODO. Reconsider how
5931     %% loops are nested.
5932     for k=1, #lbkr[lang] do
5933         local p = lbkr[lang][k].pattern
5934         local r = lbkr[lang][k].replace
5935
5936         if Babel.debug then
5937             print('*****', p, mode)
5938         end
5939
5940         %% This variable is set in some cases below to the first *byte*
5941         %% after the match, either as found by u.match (faster) or the
5942         %% computed position based on sc if w has changed.
5943         local last_match = 0
5944         local step = 0
5945
5946         %% For every match.
5947         while true do
5948             if Babel.debug then
5949                 print('====')
5950             end
5951             local new  %% used when inserting and removing nodes
5952
5953             local matches = { u.match(w, p, last_match) }
5954
5955             if #matches < 2 then break end
5956
5957             %% Get and remove empty captures (with ())'s, which return a
5958             %% number with the position), and keep actual captures
5959             %% (from (...)), if any, in matches.
5960             local first = table.remove(matches, 1)
5961             local last  = table.remove(matches, #matches)
5962             %% Non re-fetched substrings may contain \31, which separates
5963             %% subsubstrings.
5964             if string.find(w:sub(first, last-1), Babel.us_char) then break end
5965
5966             local save_last = last  %% with A()BC()D, points to D
5967
5968             %% Fix offsets, from bytes to unicode. Explained above.
5969             first = u.len(w:sub(1, first-1)) + 1
5970             last  = u.len(w:sub(1, last-1))  %% now last points to C
5971
5972             %% This loop stores in n small table the nodes
5973             %% corresponding to the pattern. Used by 'data' to provide a
5974             %% predictable behavior with 'insert' (now w_nodes is modified on
5975             %% the fly), and also access to 'remove'd nodes.
5976             local sc = first-1          %% Used below, too
5977             local data_nodes = {}
5978
5979             for q = 1, last-first+1 do
5980                 data_nodes[q] = w_nodes[sc+q]
5981             end
5982

```

```

5983      %% This loop traverses the matched substring and takes the
5984      %% corresponding action stored in the replacement list.
5985      %% sc = the position in substr nodes / string
5986      %% rc = the replacement table index
5987      local rc = 0
5988
5989      while rc < last-first+1 do %% for each replacement
5990          if Babel.debug then
5991              print('.....', rc + 1)
5992          end
5993          sc = sc + 1
5994          rc = rc + 1
5995
5996          if Babel.debug then
5997              Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
5998              local ss = ''
5999              for itt in node.traverse(head) do
6000                  if itt.id == 29 then
6001                      ss = ss .. unicode.utf8.char(itt.char)
6002                  else
6003                      ss = ss .. '{' .. itt.id .. '}'
6004                  end
6005              end
6006              print('*****', ss)
6007          end
6008
6009          local crep = r[rc]
6010          local item = w_nodes[sc]
6011          local item_base = item
6012          local placeholder = Babel.us_char
6013          local d
6014
6015          if crep and crep.data then
6016              item_base = data_nodes[crep.data]
6017          end
6018
6019          if crep then
6020              step = crep.step or 0
6021          end
6022
6023          if crep and next(crep) == nil then %% = {}
6024              last_match = save_last    %% Optimization
6025              goto next
6026          end
6027
6028          elseif crep == nil or crep.remove then
6029              node.remove(head, item)
6030              table.remove(w_nodes, sc)
6031              w = u.sub(w, 1, sc-1) .. u.sub(w, sc+1)
6032              sc = sc - 1  %% Nothing has been inserted.
6033              last_match = utf8.offset(w, sc+1+step)
6034              goto next
6035          end
6036
6037          elseif crep and crep.kashida then %% Experimental
6038              node.set_attribute(item,
6039                  luatexbase.registernumber'bblar@kashida',
6040                  crep.kashida)
6041              last_match = utf8.offset(w, sc+1+step)
6042              goto next

```



```

6042
6043 elseif crep and crep.string then
6044     local str = crep.string(matches)
6045     if str == '' then %% Gather with nil
6046         node.remove(head, item)
6047         table.remove(w_nodes, sc)
6048         w = u.sub(w, 1, sc-1) .. u.sub(w, sc+1)
6049         sc = sc - 1 %% Nothing has been inserted.
6050     else
6051         local loop_first = true
6052         for s in string.utfvalues(str) do
6053             d = node.copy(item_base)
6054             d.char = s
6055             if loop_first then
6056                 loop_first = false
6057                 head, new = node.insert_before(head, item, d)
6058                 if sc == 1 then
6059                     word_head = head
6060                 end
6061                 w_nodes[sc] = d
6062                 w = u.sub(w, 1, sc-1) .. u.char(s) .. u.sub(w, sc+1)
6063             else
6064                 sc = sc + 1
6065                 head, new = node.insert_before(head, item, d)
6066                 table.insert(w_nodes, sc, new)
6067                 w = u.sub(w, 1, sc-1) .. u.char(s) .. u.sub(w, sc)
6068             end
6069             if Babel.debug then
6070                 print('.....', 'str')
6071                 Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
6072             end
6073             end %% for
6074             node.remove(head, item)
6075         end %% if ''
6076         last_match = utf8.offset(w, sc+1+step)
6077         goto next
6078
6079 elseif mode == 1 and crep and (crep.pre or crep.no or crep.post) then
6080     d = node.new(7, 0) %% (disc, discretionary)
6081     d.pre = Babel.str_to_nodes(crep.pre, matches, item_base)
6082     d.post = Babel.str_to_nodes(crep.post, matches, item_base)
6083     d.replace = Babel.str_to_nodes(crep.no, matches, item_base)
6084     d.attr = item_base.attr
6085     if crep.pre == nil then %% TeXbook p96
6086         d.penalty = crep.penalty or tex.hyphenpenalty
6087     else
6088         d.penalty = crep.penalty or tex.exhyphenpenalty
6089     end
6090     placeholder = '|'
6091     head, new = node.insert_before(head, item, d)
6092
6093 elseif mode == 0 and crep and (crep.pre or crep.no or crep.post) then
6094     %% ERROR
6095
6096 elseif crep and crep.penalty then
6097     d = node.new(14, 0) %% (penalty, userpenalty)
6098     d.attr = item_base.attr
6099     d.penalty = crep.penalty
6100     head, new = node.insert_before(head, item, d)

```

```

6101
6102 elseif crep and crep.space then
6103     %% 655360 = 10 pt = 10 * 65536 sp
6104     d = node.new(12, 13)      %% (glue, spaceskip)
6105     local quad = font.getfont(item_base.font).size or 655360
6106     node.setglue(d, crep.space[1] * quad,
6107                  crep.space[2] * quad,
6108                  crep.space[3] * quad)
6109     if mode == 0 then
6110         placeholder = ' '
6111     end
6112     head, new = node.insert_before(head, item, d)
6113
6114 elseif crep and crep.spacefactor then
6115     d = node.new(12, 13)      %% (glue, spaceskip)
6116     local base_font = font.getfont(item_base.font)
6117     node.setglue(d,
6118                  crep.spacefactor[1] * base_font.parameters['space'],
6119                  crep.spacefactor[2] * base_font.parameters['space_stretch'],
6120                  crep.spacefactor[3] * base_font.parameters['space_shrink'])
6121     if mode == 0 then
6122         placeholder = ' '
6123     end
6124     head, new = node.insert_before(head, item, d)
6125
6126 elseif mode == 0 and crep and crep.space then
6127     %% ERROR
6128
6129 end    %% ie replacement cases
6130
6131 %% Shared by disc, space and penalty.
6132 if sc == 1 then
6133     word_head = head
6134 end
6135 if crep.insert then
6136     w = u.sub(w, 1, sc-1) .. placeholder .. u.sub(w, sc)
6137     table.insert(w_nodes, sc, new)
6138     last = last + 1
6139 else
6140     w_nodes[sc] = d
6141     node.remove(head, item)
6142     w = u.sub(w, 1, sc-1) .. placeholder .. u.sub(w, sc+1)
6143 end
6144
6145 last_match = utf8.offset(w, sc+1+step)
6146
6147 ::next::
6148
6149 end    %% for each replacement
6150
6151 if Babel.debug then
6152     print('.....', '/')
6153     Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
6154 end
6155
6156 end    %% for match
6157
6158 end    %% for patterns
6159

```

```

6160         ::next::
6161         word_head = nw
6162     end &% for substring
6163     return head
6164 end
6165
6166 &% This table stores capture maps, numbered consecutively
6167 Babel.capture_maps = {}
6168
6169 &% The following functions belong to the next macro
6170 function Babel.capture_func(key, cap)
6171     local ret = "[" .. cap:gsub('{{[0-9]}}', ")]..m[%1]..["] .. "]"
6172     local cnt
6173     local u = unicode.utf8
6174     ret, cnt = ret:gsub('{{[0-9]}|([^|]+)|(.-)}', Babel.capture_func_map)
6175     if cnt == 0 then
6176         ret = u.gsub(ret, '{{(%x%x%x%x+)}',
6177             function (n)
6178                 return u.char(tonumber(n, 16))
6179             end)
6180     end
6181     ret = ret:gsub("%[%[%]]%.", '')
6182     ret = ret:gsub("%.%[%[%]]%", '')
6183     return key .. "[=function(m) return ]] .. ret .. [[ end]]
6184 end
6185
6186 function Babel.capt_map(from, mapno)
6187     return Babel.capture_maps[mapno][from] or from
6188 end
6189
6190 &% Handle the {n|abc|ABC} syntax in captures
6191 function Babel.capture_func_map(capno, from, to)
6192     local u = unicode.utf8
6193     from = u.gsub(from, '{{(%x%x%x%x+)}',
6194         function (n)
6195             return u.char(tonumber(n, 16))
6196         end)
6197     to = u.gsub(to, '{{(%x%x%x%x+)}',
6198         function (n)
6199             return u.char(tonumber(n, 16))
6200         end)
6201     local froms = {}
6202     for s in string.utfcharacters(from) do
6203         table.insert(froms, s)
6204     end
6205     local cnt = 1
6206     table.insert(Babel.capture_maps, {})
6207     local mlen = table.getn(Babel.capture_maps)
6208     for s in string.utfcharacters(to) do
6209         Babel.capture_maps[mlen][froms[cnt]] = s
6210         cnt = cnt + 1
6211     end
6212     return "]"..Babel.capt_map(m[" .. capno .. "], " ..
6213         (mlen) .. ").." .. "["
6214 end
6215
6216 &% Create/Extend reversed sorted list of kashida weights:
6217 function Babel.capture_kashida(key, wt)
6218     wt = tonumber(wt)

```

```

6219   if Babel.kashida_wts then
6220     for p, q in ipairs(Babel.kashida_wts) do
6221       if wt == q then
6222         break
6223       elseif wt > q then
6224         table.insert(Babel.kashida_wts, p, wt)
6225         break
6226       elseif table.getn(Babel.kashida_wts) == p then
6227         table.insert(Babel.kashida_wts, wt)
6228       end
6229     end
6230   else
6231     Babel.kashida_wts = { wt }
6232   end
6233   return 'kashida = ' .. wt
6234 end
6235 }

```

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 `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 `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).

```

6236 \catcode`\# = 6
6237 \gdef\babelposthyphenation#1#2#3{&&
6238   \bbl@activateposthyphen
6239   \begingroup
6240     \def\babeltempa{\bbl@add@list\babeltempb}&&
6241     \let\babeltempb\@empty
6242     \def\bbl@tempa{#3}&& TODO. Ugly trick to preserve {}:
6243     \bbl@replace\bbl@tempa{,}{ , }&&
6244     \expandafter\bbl@foreach\expandafter{\bbl@tempa}{&&
6245       \bbl@ifsamestring{##1}{remove}&&
6246       {\bbl@add@list\babeltempb{nil}}&&
6247       {\directlua{
6248         local rep = {[#1]=]
6249         rep = rep:gsub('^%s*(remove)%s*$', 'remove = true')
6250         rep = rep:gsub('^%s*(insert)%s*', 'insert = true, ')
6251         rep = rep:gsub(' (no)%s*=%s*([^\s,]*)', Babel.capture_func)
6252         rep = rep:gsub(' (pre)%s*=%s*([^\s,]*)', Babel.capture_func)
6253         rep = rep:gsub(' (post)%s*=%s*([^\s,]*)', Babel.capture_func)
6254         rep = rep:gsub(' (string)%s*=%s*([^\s,]*)', Babel.capture_func)
6255         tex.print([[ \string\babeltempa{[]] .. rep .. [[]]])
6256       }}}&&
6257   \directlua{
6258     local lbkr = Babel.linebreaking.replacements[1]
6259     local u = unicode.utf8
6260     local id = \the\csname l@#1\endcsname
6261     && Convert pattern:
6262     local patt = string.gsub([=[#2]=], '%s', '')
6263     if not u.find(patt, '()', nil, true) then
6264       patt = '()' .. patt .. '()'
6265     end
6266     patt = string.gsub(patt, '%(%)^', '^()')
6267     patt = string.gsub(patt, '%$(%)', '()$')

```

```

6268     patt = u.gsub(patt, '{(.)}',
6269         function (n)
6270             return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
6271         end)
6272     patt = u.gsub(patt, '{(%x%x%x%x+)}',
6273         function (n)
6274             return u.gsub(u.char(tonumber(n, 16)), '(%p)', '%%1')
6275         end)
6276     lbkr[id] = lbkr[id] or {}
6277     table.insert(lbkr[id], { pattern = patt, replace = { \babeltempb } })
6278 }&%
6279 \endgroup}
6280 % TODO. Copy paste pattern.
6281 \gdef\babelprehyphenation#1#2#3{&%
6282     \bbl@activateprehyphen
6283     \begingroup
6284     \def\babeltempa{\bbl@add@list\babeltempb}&%
6285     \let\babeltempb\@empty
6286     \def\bbl@tempa{#3}&% TODO. Ugly trick to preserve {}:
6287     \bbl@replace\bbl@tempa{,}{ ,}&%
6288     \expandafter\bbl@foreach\expandafter{\bbl@tempa}{&%
6289         \bbl@ifsamestring{##1}{remove}&%
6290         {\bbl@add@list\babeltempb{nil}}&%
6291         {\directlua{
6292             local rep = {[##1]=}
6293             rep = rep.gsub('^%s*(remove)%s*$', 'remove = true')
6294             rep = rep.gsub('^%s*(insert)%s*', 'insert = true, ')
6295             rep = rep.gsub('(string)%s*=%s*([^\s,]*)', Babel.capture_func)
6296             rep = rep.gsub('(space)%s*=%s*([%d%.]+)%s+([%d%.]+)%s+([%d%.]+)',
6297                 'space = {' .. '%2, %3, %4' .. '}')
6298             rep = rep.gsub('(spacefactor)%s*=%s*([%d%.]+)%s+([%d%.]+)%s+([%d%.]+)',
6299                 'spacefactor = {' .. '%2, %3, %4' .. '}')
6300             rep = rep.gsub('(kashida)%s*=%s*([^\s,]*)', Babel.capture_kashida)
6301             tex.print([[ \string\babeltempa{[]} .. rep .. []]])
6302         }}&%
6303     \directlua{
6304         local lbkr = Babel.linebreaking.replacements[0]
6305         local u = unicode.utf8
6306         local id = \the\csname bbl@id@@#1\endcsname
6307         &% Convert pattern:
6308         local patt = string.gsub(==[#2]==, '%s', '')
6309         local patt = string.gsub(patt, '|', ' ')
6310         if not u.find(patt, '()', nil, true) then
6311             patt = '()' .. patt .. '()'
6312         end
6313         &% patt = string.gsub(patt, '%(%)%', '^()')
6314         &% patt = string.gsub(patt, '([^\%])%$%', '%1()$')
6315         patt = u.gsub(patt, '{(.)}',
6316             function (n)
6317                 return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
6318             end)
6319         patt = u.gsub(patt, '{(%x%x%x%x+)}',
6320             function (n)
6321                 return u.gsub(u.char(tonumber(n, 16)), '(%p)', '%%1')
6322             end)
6323         lbkr[id] = lbkr[id] or {}
6324         table.insert(lbkr[id], { pattern = patt, replace = { \babeltempb } })
6325     }&%
6326 \endgroup}

```

```

6327 \endgroup
6328 \def\bbl@activateposthyphen{%
6329   \let\bbl@activateposthyphen\relax
6330   \directlua{
6331     Babel.linebreaking.add_after(Babel.post_hyphenate_replace)
6332   }}
6333 \def\bbl@activateprehyphen{%
6334   \let\bbl@activateprehyphen\relax
6335   \directlua{
6336     Babel.linebreaking.add_before(Babel.pre_hyphenate_replace)
6337   }}

```

### 13.9 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`, `hline`, `colortbl`, `longtable`, `booktabs`, etc. However, `dcolumn` still fails.

```

6338 \bbl@trace{Redefinitions for bidi layout}
6339 \ifx\@eqnnum\undefined\else
6340   \ifx\bbl@attr@dir\undefined\else
6341     \edef\@eqnnum{%
6342       \unexpanded{\ifcase\bbl@attr@dir\else\bbl@textdir\@ne\fi}%
6343       \unexpanded\expandafter{\@eqnnum}}}%
6344   \fi
6345 \fi
6346 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
6347 \ifnum\bbl@bidimode>\z@
6348   \def\bbl@nextfake#1{% non-local changes, use always inside a group!
6349     \bbl@exp{%
6350       \mathdir\the\bodydir
6351       #1%           Once entered in math, set boxes to restore values
6352       \<ifmmode>%
6353         \everyvbox{%
6354           \the\everyvbox
6355           \bodydir\the\bodydir
6356           \mathdir\the\mathdir
6357           \everyhbox{\the\everyhbox}%
6358           \everyvbox{\the\everyvbox}}%
6359         \everyhbox{%
6360           \the\everyhbox
6361           \bodydir\the\bodydir
6362           \mathdir\the\mathdir
6363           \everyhbox{\the\everyhbox}%
6364           \everyvbox{\the\everyvbox}}%
6365       \<fi>}}%
6366   \def\@hangfrom#1{%
6367     \setbox\@tempboxa\hbox{{#1}}%
6368     \hangindent\wd\@tempboxa
6369     \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else

```

```

6370     \shapemode\@ne
6371     \fi
6372     \noindent\box\@tempboxa}
6373 \fi
6374 \IfBabelLayout{tabular}
6375   {\let\bbl@OL@tabular\@tabular
6376     \bbl@replace\@tabular{$}\bbl@nextfake$}%
6377     \let\bbl@NL@tabular\@tabular
6378     \AtBeginDocument{%
6379       \ifx\bbl@NL@tabular\@tabular\else
6380         \bbl@replace\@tabular{$}\bbl@nextfake$}%
6381         \let\bbl@NL@tabular\@tabular
6382       \fi}}
6383   {}
6384 \IfBabelLayout{lists}
6385   {\let\bbl@OL@list\list
6386     \bbl@sreplace\list{\parshape}\bbl@listparshape}%
6387     \let\bbl@NL@list\list
6388     \def\bbl@listparshape#1#2#3{%
6389       \parshape #1 #2 #3 %
6390       \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
6391         \shapemode\tw@
6392       \fi}}
6393   {}
6394 \IfBabelLayout{graphics}
6395   {\let\bbl@pictresetdir\relax
6396     \def\bbl@pictsetdir#1{%
6397       \ifcase\bbl@thetextdir
6398         \let\bbl@pictresetdir\relax
6399       \else
6400         \ifcase#1\bodydir TLT % Remember this sets the inner boxes
6401           \or\textdir TLT
6402           \else\bodydir TLT \textdir TLT
6403         \fi
6404         % \(\text|par)dir required in pgf:
6405         \def\bbl@pictresetdir{\bodydir TRT\pardir TRT\textdir TRT\relax}%
6406       \fi}%
6407   \ifx\AddToHook\@undefined\else
6408     \AddToHook{env/picture/begin}{\bbl@pictsetdir\tw@}%
6409     \directlua{
6410       Babel.get_picture_dir = true
6411       Babel.picture_has_bidi = 0
6412       function Babel.picture_dir (head)
6413         if not Babel.get_picture_dir then return head end
6414         for item in node.traverse(head) do
6415           if item.id == node.id'glyph' then
6416             local itemchar = item.char
6417             % TODO. Copy paste pattern from Babel.bidi (-r)
6418             local chardata = Babel.characters[itemchar]
6419             local dir = chardata and chardata.d or nil
6420             if not dir then
6421               for nn, et in ipairs(Babel.ranges) do
6422                 if itemchar < et[1] then
6423                   break
6424                 elseif itemchar <= et[2] then
6425                   dir = et[3]
6426                   break
6427                 end
6428               end

```

```

6429         end
6430         if dir and (dir == 'al' or dir == 'r') then
6431             Babel.picture_has_bidi = 1
6432         end
6433     end
6434 end
6435     return head
6436 end
6437     luatexbase.add_to_callback("hpack_filter", Babel.picture_dir,
6438     "Babel.picture_dir")
6439 }%
6440 \AtBeginDocument{%
6441     \long\def\put(#1,#2)#3{%
6442         \@killglue
6443         % Try:
6444         \ifx\bbl@pictresetdir\relax
6445             \def\bbl@tempc{0}%
6446         \else
6447             \directlua{
6448                 Babel.get_picture_dir = true
6449                 Babel.picture_has_bidi = 0
6450             }%
6451             \setbox\z@\hb@xt@\z@{%
6452                 \@defaultunitsset\@tempdimc{#1}\unitlength
6453                 \kern\@tempdimc
6454                 #3\hss}%
6455             \edef\bbl@tempc{\directlua{tex.print(Babel.picture_has_bidi)}}%
6456         \fi
6457         % Do:
6458         \@defaultunitsset\@tempdimc{#2}\unitlength
6459         \raise\@tempdimc\hb@xt@\z@{%
6460             \@defaultunitsset\@tempdimc{#1}\unitlength
6461             \kern\@tempdimc
6462             {\ifnum\bbl@tempc>\z@\bbl@pictresetdir\fi#3}\hss}%
6463         \ignorespaces}%
6464         \MakeRobust\put}%
6465 \fi
6466 \AtBeginDocument
6467 {
6468     \ifx\tikz@atbegin@node\undefined\else
6469         \ifx\AddToHook\undefined\else % TODO. Still tentative.
6470             \AddToHook{env/pgfpicture/begin}{\bbl@pictsetdir\@ne}%
6471             \bbl@add\pgfinterruptpicture{\bbl@pictresetdir}%
6472         \fi
6473         \let\bbl@OL@pgfpicture\pgfpicture
6474         \bbl@sreplace\pgfpicture{\pgfpicturetrue}%
6475         {\bbl@pictsetdir\z@\pgfpicturetrue}%
6476         \bbl@add\pgfsys@beginpicture{\bbl@pictsetdir\z@}%
6477         \bbl@add\tikz@atbegin@node{\bbl@pictresetdir}%
6478         \bbl@sreplace\tikz{\beginpgroup}%
6479         {\beginpgroup\bbl@pictsetdir\tw@}%
6480     \fi
6481     \ifx\AddToHook\undefined\else
6482         \AddToHook{env/tcolorbox/begin}{\bbl@pictsetdir\@ne}%
6483     \fi
6484 }%

```

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.

```

6485 \IfBabelLayout{counters}%
6486   {\let\bb1@OL@@textsuperscript\@textsuperscript
6487     \bb1@sreplace\@textsuperscript{\m@th}{\m@th\mathdir\pagedir}%
6488     \let\bb1@latinarabic=\@arabic
6489     \let\bb1@OL@@arabic\@arabic
6490     \def\@arabic#1{\babelsublr{\bb1@latinarabic#1}}}%
6491   \@ifpackagewith{babel}{bidi=default}%
6492     {\let\bb1@asciroman=\@roman
6493       \let\bb1@OL@@roman\@roman
6494       \def\@roman#1{\babelsublr{\ensureascii{\bb1@asciroman#1}}}%
6495       \let\bb1@asciiRoman=\@Roman
6496       \let\bb1@OL@@roman\@Roman
6497       \def\@Roman#1{\babelsublr{\ensureascii{\bb1@asciiRoman#1}}}%
6498       \let\bb1@OL@labelenumii\labelenumii
6499       \def\labelenumii{}\theenumii{}}%
6500     \let\bb1@OL@p@enumiii\p@enumiii
6501     \def\p@enumiii{\p@enumii}\theenumii{}}{}{}
6502 <<Footnote changes>>
6503 \IfBabelLayout{footnotes}%
6504   {\let\bb1@OL@footnote\footnote
6505     \BabelFootnote\footnote\languagename{}}{}%
6506     \BabelFootnote\localfootnote\languagename{}}{}%
6507     \BabelFootnote\mainfootnote{}}{}{}
6508   {}

```

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.

```

6509 \IfBabelLayout{extras}%
6510   {\let\bb1@OL@underline\underline
6511     \bb1@sreplace\underline{\$ \@underline}{\bb1@nextfake\$ \@underline}%
6512     \let\bb1@OL@LaTeX2e\LaTeX2e
6513     \DeclareRobustCommand{\LaTeXe}{\mbox{\m@th
6514       \if b\expandafter\@car\@series\@nil\boldmath\fi
6515       \babelsublr{%
6516         \LaTeX\kern.15em2\bb1@nextfake$_{\textstyle\varepsilon}$}}}%
6517   {}
6518 </luatex>

```

### 13.10 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):

Arrrgh!! 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.

```

6519 (*basic-r)
6520 Babel = Babel or {}
6521
6522 Babel.bidi_enabled = true
6523
6524 require('babel-data-bidi.lua')
6525
6526 local characters = Babel.characters
6527 local ranges = Babel.ranges
6528
6529 local DIR = node.id("dir")
6530
6531 local function dir_mark(head, from, to, outer)
6532   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
6533   local d = node.new(DIR)
6534   d.dir = '+' .. dir
6535   node.insert_before(head, from, d)
6536   d = node.new(DIR)
6537   d.dir = '-' .. dir
6538   node.insert_after(head, to, d)
6539 end
6540
6541 function Babel.bidi(head, ispar)
6542   local first_n, last_n          -- first and last char with nums
6543   local last_es                  -- an auxiliary 'last' used with nums
6544   local first_d, last_d          -- first and last char in L/R block
6545   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):

```

6546   local strong = ('TRT' == tex.pardir) and 'r' or 'l'
6547   local strong_lr = (strong == 'l') and 'l' or 'r'
6548   local outer = strong
6549
6550   local new_dir = false
6551   local first_dir = false
6552   local inmath = false
6553
6554   local last_lr
6555

```

```

6556 local type_n = ''
6557
6558 for item in node.traverse(head) do
6559     -- three cases: glyph, dir, otherwise
6560     if item.id == node.id'glyph'
6561     or (item.id == 7 and item.subtype == 2) then
6562         local itemchar
6563         if item.id == 7 and item.subtype == 2 then
6564             itemchar = item.replace.char
6565         else
6566             itemchar = item.char
6567         end
6568         local chardata = characters[itemchar]
6569         dir = chardata and chardata.d or nil
6570         if not dir then
6571             for nn, et in ipairs(ranges) do
6572                 if itemchar < et[1] then
6573                     break
6574                 elseif itemchar <= et[2] then
6575                     dir = et[3]
6576                     break
6577                 end
6578             end
6579         end
6580         dir = dir or 'l'
6581         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).

```

6584     if new_dir then
6585         attr_dir = 0
6586         for at in node.traverse(item.attr) do
6587             if at.number == luatexbase.registernumber'bbl@attr@dir' then
6588                 attr_dir = at.value % 3
6589             end
6590         end
6591         if attr_dir == 1 then
6592             strong = 'r'
6593         elseif attr_dir == 2 then
6594             strong = 'al'
6595         else
6596             strong = 'l'
6597         end
6598         strong_lr = (strong == 'l') and 'l' or 'r'
6599         outer = strong_lr
6600         new_dir = false
6601     end
6602
6603     if dir == 'nsm' then dir = strong end -- W1

```

**Numbers.** The dual <al>/<r> system for R is somewhat cumbersome.

```

6604     dir_real = dir -- We need dir_real to set strong below
6605     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:

```

6606     if strong == 'al' then
6607         if dir == 'en' then dir = 'an' end           -- W2
6608         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
6609         strong_lr = 'r'                               -- W3
6610     end

```

Once finished the basic setup for glyphs, consider the two other cases: dir node and the rest.

```

6611     elseif item.id == node.id'dir' and not inmath then
6612         new_dir = true
6613         dir = nil
6614     elseif item.id == node.id'math' then
6615         inmath = (item.subtype == 0)
6616     else
6617         dir = nil           -- Not a char
6618     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>.

```

6619     if dir == 'en' or dir == 'an' or dir == 'et' then
6620         if dir ~= 'et' then
6621             type_n = dir
6622         end
6623         first_n = first_n or item
6624         last_n = last_n or item
6625         last_es = nil
6626     elseif dir == 'es' and last_n then -- W3+W6
6627         last_es = item
6628     elseif dir == 'cs' then           -- it's right - do nothing
6629     elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
6630         if strong_lr == 'r' and type_n ~= '' then
6631             dir_mark(head, first_n, last_n, 'r')
6632         elseif strong_lr == 'l' and first_d and type_n == 'an' then
6633             dir_mark(head, first_n, last_n, 'r')
6634             dir_mark(head, first_d, last_d, outer)
6635             first_d, last_d = nil, nil
6636         elseif strong_lr == 'l' and type_n ~= '' then
6637             last_d = last_n
6638         end
6639         type_n = ''
6640         first_n, last_n = nil, nil
6641     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:

```

6642     if dir == 'l' or dir == 'r' then
6643         if dir ~= outer then
6644             first_d = first_d or item
6645             last_d = item
6646         elseif first_d and dir ~= strong_lr then
6647             dir_mark(head, first_d, last_d, outer)
6648             first_d, last_d = nil, nil
6649         end

```

```
6650     end
```

**Mirroring.** Each chunk of text in a certain language is considered a “closed” sequence. If <r on r> and <l on l>, it's clearly <r> and <l>, resp'tly, but with other combinations depends on outer. From all these, we select only those resolving <on> → <r>. At the beginning (when last\_lr is nil) of an R text, they are mirrored directly.

TODO - numbers in R mode are processed. It doesn't hurt, but should not be done.

```
6651     if dir and not last_lr and dir ~= 'l' and outer == 'r' then
6652         item.char = characters[item.char] and
6653             characters[item.char].m or item.char
6654     elseif (dir or new_dir) and last_lr ~= item then
6655         local mir = outer .. strong_lr .. (dir or outer)
6656         if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
6657             for ch in node.traverse(node.next(last_lr)) do
6658                 if ch == item then break end
6659                 if ch.id == node.id'glyph' and characters[ch.char] then
6660                     ch.char = characters[ch.char].m or ch.char
6661                 end
6662             end
6663         end
6664     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).

```
6665     if dir == 'l' or dir == 'r' then
6666         last_lr = item
6667         strong = dir_real          -- Don't search back - best save now
6668         strong_lr = (strong == 'l') and 'l' or 'r'
6669     elseif new_dir then
6670         last_lr = nil
6671     end
6672 end
```

Mirror the last chars if they are no directed. And make sure any open block is closed, too.

```
6673 if last_lr and outer == 'r' then
6674     for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
6675         if characters[ch.char] then
6676             ch.char = characters[ch.char].m or ch.char
6677         end
6678     end
6679 end
6680 if first_n then
6681     dir_mark(head, first_n, last_n, outer)
6682 end
6683 if first_d then
6684     dir_mark(head, first_d, last_d, outer)
6685 end
```

In boxes, the dir node could be added before the original head, so the actual head is the previous node.

```
6686 return node.prev(head) or head
6687 end
6688 </basic-r>
```

And here the Lua code for bidi=basic:

```
6689 <(*basic>
6690 Babel = Babel or {}
6691
6692 -- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
6693
```

```

6694 Babel.fontmap = Babel.fontmap or {}
6695 Babel.fontmap[0] = {}      -- l
6696 Babel.fontmap[1] = {}      -- r
6697 Babel.fontmap[2] = {}      -- al/an
6698
6699 Babel.bidi_enabled = true
6700 Babel.mirroring_enabled = true
6701
6702 require('babel-data-bidi.lua')
6703
6704 local characters = Babel.characters
6705 local ranges = Babel.ranges
6706
6707 local DIR = node.id('dir')
6708 local GLYPH = node.id('glyph')
6709
6710 local function insert_implicit(head, state, outer)
6711   local new_state = state
6712   if state.sim and state.eim and state.sim ~= state.eim then
6713     dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
6714     local d = node.new(DIR)
6715     d.dir = '+' .. dir
6716     node.insert_before(head, state.sim, d)
6717     local d = node.new(DIR)
6718     d.dir = '-' .. dir
6719     node.insert_after(head, state.eim, d)
6720   end
6721   new_state.sim, new_state.eim = nil, nil
6722   return head, new_state
6723 end
6724
6725 local function insert_numeric(head, state)
6726   local new
6727   local new_state = state
6728   if state.san and state.ean and state.san ~= state.ean then
6729     local d = node.new(DIR)
6730     d.dir = '+TLT'
6731     _, new = node.insert_before(head, state.san, d)
6732     if state.san == state.sim then state.sim = new end
6733     local d = node.new(DIR)
6734     d.dir = '-TLT'
6735     _, new = node.insert_after(head, state.ean, d)
6736     if state.ean == state.eim then state.eim = new end
6737   end
6738   new_state.san, new_state.ean = nil, nil
6739   return head, new_state
6740 end
6741
6742 -- TODO - \hbox with an explicit dir can lead to wrong results
6743 -- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>. A small attempt
6744 -- was s made to improve the situation, but the problem is the 3-dir
6745 -- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
6746 -- well.
6747
6748 function Babel.bidi(head, ispar, hdir)
6749   local d -- d is used mainly for computations in a loop
6750   local prev_d = ''
6751   local new_d = false
6752

```

```

6753 local nodes = {}
6754 local outer_first = nil
6755 local inmath = false
6756
6757 local glue_d = nil
6758 local glue_i = nil
6759
6760 local has_en = false
6761 local first_et = nil
6762
6763 local ATDIR = luatexbase.registernumber'bbl@attr@dir'
6764
6765 local save_outer
6766 local temp = node.get_attribute(head, ATDIR)
6767 if temp then
6768     temp = temp % 3
6769     save_outer = (temp == 0 and 'l') or
6770                 (temp == 1 and 'r') or
6771                 (temp == 2 and 'al')
6772 elseif ispar then -- Or error? Shouldn't happen
6773     save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
6774 else -- Or error? Shouldn't happen
6775     save_outer = ('TRT' == hdir) and 'r' or 'l'
6776 end
6777 -- when the callback is called, we are just _after_ the box,
6778 -- and the textdir is that of the surrounding text
6779 -- if not ispar and hdir ~= tex.textdir then
6780 --     save_outer = ('TRT' == hdir) and 'r' or 'l'
6781 -- end
6782 local outer = save_outer
6783 local last = outer
6784 -- 'al' is only taken into account in the first, current loop
6785 if save_outer == 'al' then save_outer = 'r' end
6786
6787 local fontmap = Babel.fontmap
6788
6789 for item in node.traverse(head) do
6790
6791     -- In what follows, #node is the last (previous) node, because the
6792     -- current one is not added until we start processing the neutrals.
6793
6794     -- three cases: glyph, dir, otherwise
6795     if item.id == GLYPH
6796         or (item.id == 7 and item.subtype == 2) then
6797
6798         local d_font = nil
6799         local item_r
6800         if item.id == 7 and item.subtype == 2 then
6801             item_r = item.replace -- automatic discs have just 1 glyph
6802         else
6803             item_r = item
6804         end
6805         local chardata = characters[item_r.char]
6806         d = chardata and chardata.d or nil
6807         if not d or d == 'nsm' then
6808             for nn, et in ipairs(ranges) do
6809                 if item_r.char < et[1] then
6810                     break
6811                 elseif item_r.char <= et[2] then

```

```

6812         if not d then d = et[3]
6813         elseif d == 'nsm' then d_font = et[3]
6814         end
6815         break
6816     end
6817 end
6818 end
6819 d = d or 'l'
6820
6821 -- A short 'pause' in bidi for mapfont
6822 d_font = d_font or d
6823 d_font = (d_font == 'l' and 0) or
6824           (d_font == 'nsm' and 0) or
6825           (d_font == 'r' and 1) or
6826           (d_font == 'al' and 2) or
6827           (d_font == 'an' and 2) or nil
6828 if d_font and fontmap and fontmap[d_font][item_r.font] then
6829     item_r.font = fontmap[d_font][item_r.font]
6830 end
6831
6832 if new_d then
6833     table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
6834     if inmath then
6835         attr_d = 0
6836     else
6837         attr_d = node.get_attribute(item, ATDIR)
6838         attr_d = attr_d % 3
6839     end
6840     if attr_d == 1 then
6841         outer_first = 'r'
6842         last = 'r'
6843     elseif attr_d == 2 then
6844         outer_first = 'r'
6845         last = 'al'
6846     else
6847         outer_first = 'l'
6848         last = 'l'
6849     end
6850     outer = last
6851     has_en = false
6852     first_et = nil
6853     new_d = false
6854 end
6855
6856 if glue_d then
6857     if (d == 'l' and 'l' or 'r') ~= glue_d then
6858         table.insert(nodes, {glue_i, 'on', nil})
6859     end
6860     glue_d = nil
6861     glue_i = nil
6862 end
6863
6864 elseif item.id == DIR then
6865     d = nil
6866     new_d = true
6867
6868 elseif item.id == node.id'glue' and item.subtype == 13 then
6869     glue_d = d
6870     glue_i = item

```



```

6871     d = nil
6872
6873 elseif item.id == node.id'math' then
6874     inmath = (item.subtype == 0)
6875
6876 else
6877     d = nil
6878 end
6879
6880 -- AL <= EN/ET/ES      -- W2 + W3 + W6
6881 if last == 'al' and d == 'en' then
6882     d = 'an'           -- W3
6883 elseif last == 'al' and (d == 'et' or d == 'es') then
6884     d = 'on'           -- W6
6885 end
6886
6887 -- EN + CS/ES + EN      -- W4
6888 if d == 'en' and #nodes >= 2 then
6889     if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
6890         and nodes[#nodes-1][2] == 'en' then
6891         nodes[#nodes][2] = 'en'
6892     end
6893 end
6894
6895 -- AN + CS + AN          -- W4 too, because uax9 mixes both cases
6896 if d == 'an' and #nodes >= 2 then
6897     if (nodes[#nodes][2] == 'cs')
6898         and nodes[#nodes-1][2] == 'an' then
6899         nodes[#nodes][2] = 'an'
6900     end
6901 end
6902
6903 -- ET/EN                  -- W5 + W7->l / W6->on
6904 if d == 'et' then
6905     first_et = first_et or (#nodes + 1)
6906 elseif d == 'en' then
6907     has_en = true
6908     first_et = first_et or (#nodes + 1)
6909 elseif first_et then      -- d may be nil here !
6910     if has_en then
6911         if last == 'l' then
6912             temp = 'l'    -- W7
6913         else
6914             temp = 'en'   -- W5
6915         end
6916     else
6917         temp = 'on'      -- W6
6918     end
6919     for e = first_et, #nodes do
6920         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
6921     end
6922     first_et = nil
6923     has_en = false
6924 end
6925
6926 -- Force mathdir in math if ON (currently works as expected only
6927 -- with 'l')
6928 if inmath and d == 'on' then
6929     d = ('TRT' == tex.mathdir) and 'r' or 'l'

```

```

6930     end
6931
6932     if d then
6933         if d == 'al' then
6934             d = 'r'
6935             last = 'al'
6936         elseif d == 'l' or d == 'r' then
6937             last = d
6938         end
6939         prev_d = d
6940         table.insert(nodes, {item, d, outer_first})
6941     end
6942
6943     outer_first = nil
6944
6945 end
6946
6947 -- TODO -- repeated here in case EN/ET is the last node. Find a
6948 -- better way of doing things:
6949 if first_et then      -- dir may be nil here !
6950     if has_en then
6951         if last == 'l' then
6952             temp = 'l'    -- W7
6953         else
6954             temp = 'en'   -- W5
6955         end
6956     else
6957         temp = 'on'      -- W6
6958     end
6959     for e = first_et, #nodes do
6960         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
6961     end
6962 end
6963
6964 -- dummy node, to close things
6965 table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
6966
6967 ----- NEUTRAL -----
6968
6969 outer = save_outer
6970 last = outer
6971
6972 local first_on = nil
6973
6974 for q = 1, #nodes do
6975     local item
6976
6977     local outer_first = nodes[q][3]
6978     outer = outer_first or outer
6979     last = outer_first or last
6980
6981     local d = nodes[q][2]
6982     if d == 'an' or d == 'en' then d = 'r' end
6983     if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
6984
6985     if d == 'on' then
6986         first_on = first_on or q
6987     elseif first_on then
6988         if last == d then

```

```

6989         temp = d
6990     else
6991         temp = outer
6992     end
6993     for r = first_on, q - 1 do
6994         nodes[r][2] = temp
6995         item = nodes[r][1]    -- MIRRORING
6996         if Babel.mirroring_enabled and item.id == GLYPH
6997             and temp == 'r' and characters[item.char] then
6998             local font_mode = font.fonts[item.font].properties.mode
6999             if font_mode ~= 'harf' and font_mode ~= 'plug' then
7000                 item.char = characters[item.char].m or item.char
7001             end
7002         end
7003     end
7004     first_on = nil
7005 end
7006
7007 if d == 'r' or d == 'l' then last = d end
7008 end
7009
7010 ----- IMPLICIT, REORDER -----
7011
7012 outer = save_outer
7013 last = outer
7014
7015 local state = {}
7016 state.has_r = false
7017
7018 for q = 1, #nodes do
7019
7020     local item = nodes[q][1]
7021
7022     outer = nodes[q][3] or outer
7023
7024     local d = nodes[q][2]
7025
7026     if d == 'nsm' then d = last end          -- W1
7027     if d == 'en' then d = 'an' end
7028     local isdir = (d == 'r' or d == 'l')
7029
7030     if outer == 'l' and d == 'an' then
7031         state.san = state.san or item
7032         state.ean = item
7033     elseif state.san then
7034         head, state = insert_numeric(head, state)
7035     end
7036
7037     if outer == 'l' then
7038         if d == 'an' or d == 'r' then      -- im -> implicit
7039             if d == 'r' then state.has_r = true end
7040             state.sim = state.sim or item
7041             state.eim = item
7042         elseif d == 'l' and state.sim and state.has_r then
7043             head, state = insert_implicit(head, state, outer)
7044         elseif d == 'l' then
7045             state.sim, state.eim, state.has_r = nil, nil, false
7046         end
7047     else

```

```

7048     if d == 'an' or d == 'l' then
7049         if nodes[q][3] then -- nil except after an explicit dir
7050             state.sim = item -- so we move sim 'inside' the group
7051         else
7052             state.sim = state.sim or item
7053         end
7054         state.eim = item
7055     elseif d == 'r' and state.sim then
7056         head, state = insert_implicit(head, state, outer)
7057     elseif d == 'r' then
7058         state.sim, state.eim = nil, nil
7059     end
7060 end
7061
7062 if isdir then
7063     last = d -- Don't search back - best save now
7064 elseif d == 'on' and state.san then
7065     state.san = state.san or item
7066     state.ean = item
7067 end
7068
7069 end
7070
7071 return node.prev(head) or head
7072 end
7073 </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.

```

7074 <{*nil}>
7075 \ProvidesLanguage{nil}[<<date>>] <<version>> Nil language]
7076 \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.

```

7077 \ifx\l@nil\@undefined
7078   \newlanguage\l@nil
7079   \@namedef{bbl@hyphendata@the\l@nil}{\{}}% Remove warning
7080   \let\bbl@elt\relax

```

```

7081 \edef\bbl@languages{% Add it to the list of languages
7082 \bbl@languages\bbl@elt{nil}{\the\l@nil}{}}
7083 \fi

```

This macro is used to store the values of the hyphenation parameters `\lefthyphenmin` and `\righthyphenmin`.

```

7084 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}

```

The next step consists of defining commands to switch to (and from) the ‘nil’ language.

```

\captionnil
\datenil
7085 \let\captionnil\@empty
7086 \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.

```

7087 \ldf@finish{nil}
7088 </nil>

```

## 16 Support for Plain $\text{\TeX}$ (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  $\text{\TeX}$ -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 `ini $\text{\TeX}$` , 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 `ini $\text{\TeX}$`  sees, we need to set some category codes just to be able to change the definition of `\input`.

```

7089 <(*bplain | blplain)
7090 \catcode`\{=1 % left brace is begin-group character
7091 \catcode`\}=2 % right brace is end-group character
7092 \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).

```

7093 \openin 0 hyphen.cfg
7094 \ifeof0
7095 \else
7096 \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.

```

7097 \def\input #1 {%
7098 \let\input\input
7099 \a hyphen.cfg
7100 \let\input\undefined
7101 }
7102 \fi
7103 </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`.

```
7104 \bplain\la plain.tex
7105 \bplain\la 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.

```
7106 \bplain\def\fmtname{babel-plain}
7107 \bplain\def\fmtname{babel-lplain}
```

When you are using a different format, based on `plain.tex` you can make a copy of `blplain.tex`, rename it and replace `plain.tex` with the name of your format file.

## 16.2 Emulating some $\text{\LaTeX}$ features

The following code duplicates or emulates parts of  $\text{\LaTeX} 2_{\epsilon}$  that are needed for `babel`.

```
7108 \langle *Emulate LaTeX\rangle \equiv
7109 % == Code for plain ==
7110 \def\@empty{}
7111 \def\loadlocalcfg#1{%
7112   \openin0#1.cfg
7113   \ifeof0
7114     \closein0
7115   \else
7116     \closein0
7117     {\immediate\write16{*****}%
7118      \immediate\write16{* Local config file #1.cfg used}%
7119      \immediate\write16{*}%
7120     }
7121     \input #1.cfg\relax
7122     \fi
7123     \@endofldf}
```

## 16.3 General tools

A number of  $\text{\LaTeX}$  macro's that are needed later on.

```
7124 \long\def\@firstofone#1{#1}
7125 \long\def\@firstoftwo#1#2{#1}
7126 \long\def\@secondoftwo#1#2{#2}
7127 \def\@nnil{\@nil}
7128 \def\@gobbletwo#1#2{}
7129 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
7130 \def\@star@or@long#1{%
7131   \@ifstar
7132   {\let\l@ngrel@x\relax#1}%
7133   {\let\l@ngrel@x\long#1}}
7134 \let\l@ngrel@x\relax
7135 \def\@car#1#2\@nil{#1}
7136 \def\@cdr#1#2\@nil{#2}
7137 \let\@typeset@protect\relax
7138 \let\protected@edef\edef
7139 \long\def\@gobble#1{}
7140 \edef\@backslashchar{\expandafter\@gobble\string\}
7141 \def\strip@prefix#1>{}
7142 \def\g@addto@macro#1#2{%
7143   \toks@\expandafter{#1#2}%
7144   \xdef#1{\the\toks@}}
7145 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
```

```

7146 \def\@nameuse#1{\csname #1\endcsname}
7147 \def\@ifundefined#1{%
7148   \expandafter\ifx\csname#1\endcsname\relax
7149     \expandafter\@firstoftwo
7150   \else
7151     \expandafter\@secondoftwo
7152   \fi}
7153 \def\@expandtwoargs#1#2#3{%
7154   \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
7155 \def\zap@space#1 #2{%
7156   #1%
7157   \ifx#2\@empty\else\expandafter\zap@space\fi
7158   #2}
7159 \let\bbl@trace\@gobble

```

$\text{\LaTeX}$  has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

7160 \ifx\@preamblecmds\undefined
7161   \def\@preamblecmds{}
7162 \fi
7163 \def\@onlypreamble#1{%
7164   \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
7165     \@preamblecmds\do#1}}
7166 \@onlypreamble\@onlypreamble

```

Mimick  $\text{\LaTeX}$ 's `\AtBeginDocument`; for this to work the user needs to add `\begin{document}` to his file.

```

7167 \def\begin{document}{%
7168   \@begin{document}hook
7169   \global\let\@begin{document}hook\undefined
7170   \def\do##1{\global\let##1\undefined}%
7171   \@preamblecmds
7172   \global\let\do\noexpand}
7173 \ifx\@begin{document}hook\undefined
7174   \def\@begin{document}hook{}
7175 \fi
7176 \@onlypreamble\@begin{document}hook
7177 \def\AtBeginDocument{\g@addto@macro\@begin{document}hook}

```

We also have to mimick  $\text{\LaTeX}$ 's `\AtEndOfPackage`. Our replacement macro is much simpler; it stores its argument in `\@endoflfd`.

```

7178 \def\AtEndOfPackage#1{\g@addto@macro\@endoflfd{#1}}
7179 \@onlypreamble\AtEndOfPackage
7180 \def\@endoflfd{}
7181 \@onlypreamble\@endoflfd
7182 \let\bbl@afterlang\@empty
7183 \chardef\bbl@opt@hyphenmap\z@

```

$\text{\LaTeX}$  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.

```

7184 \catcode`\&=\z@
7185 \ifx&\if@files\@undefined
7186   \expandafter\let\csname if@files\endcsname\expandafter\endcsname
7187   \csname iffalse\endcsname
7188 \fi
7189 \catcode`\&=4

```

Mimick  $\text{\LaTeX}$ 's commands to define control sequences.

```

7190 \def\newcommand{\@star@or@long\newcommand}

```

```

7191 \def\new@command#1{%
7192   \@testopt{\@newcommand#1}0}
7193 \def\@newcommand#1[#2]{%
7194   \@ifnextchar [{\@xargdef#1[#2]}%
7195                 {\@argdef#1[#2]}}
7196 \long\def\@argdef#1[#2]#3{%
7197   \@yargdef#1\@ne{#2}{#3}}
7198 \long\def\@xargdef#1[#2][#3]#4{%
7199   \expandafter\def\expandafter#1\expandafter{%
7200     \expandafter\@protected@testopt\expandafter #1%
7201     \csname\string#1\expandafter\endcsname{#3}}}%
7202   \expandafter\@yargdef \csname\string#1\endcsname
7203   \tw@{#2}{#4}}
7204 \long\def\@yargdef#1#2#3{%
7205   \@tempcnta#3\relax
7206   \advance \@tempcnta \@ne
7207   \let\@hash@\relax
7208   \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
7209   \@tempcntb #2%
7210   \@whilenum\@tempcntb <\@tempcnta
7211   \do{%
7212     \edef\reserved@a{\reserved@a\@hash@the\@tempcntb}%
7213     \advance\@tempcntb \@ne}%
7214   \let\@hash@###
7215   \l@ngrel@x\expandafter\def\expandafter#1\reserved@a}
7216 \def\providecommand{\@star@or@long\provide@command}
7217 \def\provide@command#1{%
7218   \begingroup
7219   \escapechar\m@ne\xdef\@gtempa{{\string#1}}%
7220   \endgroup
7221   \expandafter\@ifundefined\@gtempa
7222     {\def\reserved@a{\new@command#1}}%
7223     {\let\reserved@a\relax
7224     \def\reserved@a{\new@command\reserved@a}}%
7225   \reserved@a}%

7226 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
7227 \def\declare@robustcommand#1{%
7228   \edef\reserved@a{\string#1}%
7229   \def\reserved@b{#1}%
7230   \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
7231   \edef#1{%
7232     \ifx\reserved@a\reserved@b
7233       \noexpand\x@protect
7234       \noexpand#1%
7235     \fi
7236     \noexpand\protect
7237     \expandafter\noexpand\csname
7238       \expandafter\@gobble\string#1 \endcsname
7239   }%
7240   \expandafter\new@command\csname
7241     \expandafter\@gobble\string#1 \endcsname
7242 }
7243 \def\x@protect#1{%
7244   \ifx\protect\@typeset@protect\else
7245     \@x@protect#1%
7246   \fi
7247 }
7248 \catcode`\&=\z@ % Trick to hide conditionals

```



```
7249 \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`.

```
7250 \def\bbl@tempa{\csname newif\endcsname&ifin@}
7251 \catcode`\&=4
7252 \ifx\in@\@undefined
7253 \def\in@#1#2{%
7254 \def\in@@##1##2##3\in@{%
7255 \ifx\in@@##2\in@false\else\in@true\fi}%
7256 \in@@#2#1\in@\in@@}
7257 \else
7258 \let\bbl@tempa\@empty
7259 \fi
7260 \bbl@tempa
```

$\text{\LaTeX}$  has a macro to check whether a certain package was loaded with specific options. The command has two extra arguments which are code to be executed in either the true or false case. This is used to detect whether the document needs one of the accents to be activated (activegrave and activeacute). For plain  $\text{\TeX}$  we assume that the user wants them to be active by default. Therefore the only thing we do is execute the third argument (the code for the true case).

```
7261 \def\ifpackagewith#1#2#3#4{#3}
```

The  $\text{\LaTeX}$  macro `\ifl@aded` checks whether a file was loaded. This functionality is not needed for plain  $\text{\TeX}$  but we need the macro to be defined as a no-op.

```
7262 \def\ifl@aded#1#2#3#4{}
```

For the following code we need to make sure that the commands `\newcommand` and `\providecommand` exist with some sensible definition. They are not fully equivalent to their  $\text{\LaTeX 2}_\epsilon$  versions; just enough to make things work in plain  $\text{\TeX}$  environments.

```
7263 \ifx\@tempcnta\@undefined
7264 \csname newcount\endcsname\@tempcnta\relax
7265 \fi
7266 \ifx\@tempcntb\@undefined
7267 \csname newcount\endcsname\@tempcntb\relax
7268 \fi
```

To prevent wasting two counters in  $\text{\LaTeX 2.09}$  (because counters with the same name are allocated later by it) we reset the counter that holds the next free counter (`\count10`).

```
7269 \ifx\bye\@undefined
7270 \advance\count10 by -2\relax
7271 \fi
7272 \ifx\@ifnextchar\@undefined
7273 \def\@ifnextchar#1#2#3{%
7274 \let\reserved@d=#1%
7275 \def\reserved@a{#2}\def\reserved@b{#3}%
7276 \futurelet\@let@token\@ifnch}
7277 \def\@ifnch{%
7278 \ifx\@let@token\@sptoken
7279 \let\reserved@c\@xifnch
7280 \else
7281 \ifx\@let@token\reserved@d
7282 \let\reserved@c\reserved@a
7283 \else
7284 \let\reserved@c\reserved@b
7285 \fi
7286 \fi
7287 \reserved@c}
7288 \def\:{\let\@sptoken= } \: % this makes \@sptoken a space token
```

```

7289 \def\:{\@xifnch} \expandafter\def\:{\futurelet\@let@token\@ifnch}
7290 \fi
7291 \def\@testopt#1#2{%
7292   \@ifnextchar[{\#1}{\#1[\#2]}}
7293 \def\@protected@testopt#1{%
7294   \ifx\protect\@typeset@protect
7295     \expandafter\@testopt
7296   \else
7297     \@x@protect#1%
7298   \fi}
7299 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{\#1\relax
7300   #2\relax}\fi}
7301 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
7302   \else\expandafter\@gobble\fi{\#1}}

```

## 16.4 Encoding related macros

Code from `ltoutenc.dtx`, adapted for use in the plain  $\TeX$  environment.

```

7303 \def\DeclareTextCommand{%
7304   \@dec@text@cmd\providecommand
7305 }
7306 \def\ProvideTextCommand{%
7307   \@dec@text@cmd\providecommand
7308 }
7309 \def\DeclareTextSymbol#1#2#3{%
7310   \@dec@text@cmd\chardef#1{\#2}\#3\relax
7311 }
7312 \def\@dec@text@cmd#1#2#3{%
7313   \expandafter\def\expandafter#2%
7314     \expandafter{%
7315       \csname#3-cmd\expandafter\endcsname
7316       \expandafter#2%
7317       \csname#3\string#2\endcsname
7318     }%
7319 %   \let\@ifdefinable\@rc@ifdefinable
7320   \expandafter#1\csname#3\string#2\endcsname
7321 }
7322 \def\@current@cmd#1{%
7323   \ifx\protect\@typeset@protect\else
7324     \noexpand#1\expandafter\@gobble
7325   \fi
7326 }
7327 \def\@changed@cmd#1#2{%
7328   \ifx\protect\@typeset@protect
7329     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
7330       \expandafter\ifx\csname ?\string#1\endcsname\relax
7331         \expandafter\def\csname ?\string#1\endcsname{%
7332           \@changed@x@err{\#1}%
7333         }%
7334       \fi
7335       \global\expandafter\let
7336         \csname\cf@encoding\string#1\expandafter\endcsname
7337         \csname ?\string#1\endcsname
7338     \fi
7339     \csname\cf@encoding\string#1%
7340       \expandafter\endcsname
7341   \else
7342     \noexpand#1%

```

```

7343 \fi
7344 }
7345 \def\@changed@x@err#1{%
7346 \errhelp{Your command will be ignored, type <return> to proceed}%
7347 \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
7348 \def\DeclareTextCommandDefault#1{%
7349 \DeclareTextCommand#1?%
7350 }
7351 \def\ProvideTextCommandDefault#1{%
7352 \ProvideTextCommand#1?%
7353 }
7354 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
7355 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
7356 \def\DeclareTextAccent#1#2#3{%
7357 \DeclareTextCommand#1{#2}[1]{\accent#3 #1}
7358 }
7359 \def\DeclareTextCompositeCommand#1#2#3#4{%
7360 \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
7361 \edef\reserved@b{\string##1}%
7362 \edef\reserved@c{%
7363 \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
7364 \ifx\reserved@b\reserved@c
7365 \expandafter\expandafter\expandafter\ifx
7366 \expandafter\@car\reserved@a\relax\relax\@nil
7367 \@text@composite
7368 \else
7369 \edef\reserved@b##1{%
7370 \def\expandafter\noexpand
7371 \csname#2\string#1\endcsname####1{%
7372 \noexpand\@text@composite
7373 \expandafter\noexpand\csname#2\string#1\endcsname
7374 ####1\noexpand\@empty\noexpand\@text@composite
7375 {##1}%
7376 }%
7377 }%
7378 \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
7379 \fi
7380 \expandafter\def\csname\expandafter\string\csname
7381 #2\endcsname\string#1-\string#3\endcsname{#4}
7382 \else
7383 \errhelp{Your command will be ignored, type <return> to proceed}%
7384 \errmessage{\string\DeclareTextCompositeCommand\space used on
7385 inappropriate command \protect#1}
7386 \fi
7387 }
7388 \def\@text@composite#1#2#3\@text@composite{%
7389 \expandafter\@text@composite@x
7390 \csname\string#1-\string#2\endcsname
7391 }
7392 \def\@text@composite@x#1#2{%
7393 \ifx#1\relax
7394 #2%
7395 \else
7396 #1%
7397 \fi
7398 }
7399 %
7400 \def\@strip@args#1:#2-#3\@strip@args{#2}
7401 \def\DeclareTextComposite#1#2#3#4{%

```

```

7402 \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
7403 \bgroup
7404 \lccode` \@=#4%
7405 \lowercase{%
7406 \egroup
7407 \reserved@a @%
7408 }%
7409 }
7410 %
7411 \def\UseTextSymbol#1#2{#2}
7412 \def\UseTextAccent#1#2#3{}
7413 \def\@use@text@encoding#1{}
7414 \def\DeclareTextSymbolDefault#1#2{%
7415 \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
7416 }
7417 \def\DeclareTextAccentDefault#1#2{%
7418 \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
7419 }
7420 \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.

```

7421 \DeclareTextAccent{"}{OT1}{127}
7422 \DeclareTextAccent{'}{OT1}{19}
7423 \DeclareTextAccent{^}{OT1}{94}
7424 \DeclareTextAccent{`}{OT1}{18}
7425 \DeclareTextAccent{~}{OT1}{126}

```

The following control sequences are used in `babel.def` but are not defined for `PLAIN  $\text{\TeX}$` .

```

7426 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
7427 \DeclareTextSymbol{\textquotedblright}{OT1}{`"}
7428 \DeclareTextSymbol{\textquoteleft}{OT1}{``}
7429 \DeclareTextSymbol{\textquoteright}{OT1}{''}
7430 \DeclareTextSymbol{\i}{OT1}{16}
7431 \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`.

```

7432 \ifx\scriptsize\@undefined
7433 \let\scriptsize\sevenrm
7434 \fi
7435 % End of code for plain
7436 <</Emulate LaTeX>>

```

A proxy file:

```

7437 <{*plain}>
7438 \input babel.def
7439 </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).