

# Babel

Localization and  
internationalization

Unicode

T<sub>E</sub>X

pdfT<sub>E</sub>X

LuaT<sub>E</sub>X

XeT<sub>E</sub>X

Version 3.75.2737  
2022/05/14

Javier Bezos  
Current maintainer

Johannes L. Braams  
Original author

# 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 . . . . .	7
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 . . . . .	10
1.10	Shorthands . . . . .	12
1.11	Package options . . . . .	15
1.12	The base option . . . . .	17
1.13	ini files . . . . .	17
1.14	Selecting fonts . . . . .	25
1.15	Modifying a language . . . . .	27
1.16	Creating a language . . . . .	28
1.17	Digits and counters . . . . .	31
1.18	Dates . . . . .	33
1.19	Accessing language info . . . . .	33
1.20	Hyphenation and line breaking . . . . .	34
1.21	Transforms . . . . .	36
1.22	Selection based on BCP 47 tags . . . . .	39
1.23	Selecting scripts . . . . .	40
1.24	Selecting directions . . . . .	41
1.25	Language attributes . . . . .	44
1.26	Hooks . . . . .	45
1.27	Languages supported by babel with ldf files . . . . .	46
1.28	Unicode character properties in luatex . . . . .	47
1.29	Tweaking some features . . . . .	48
1.30	Tips, workarounds, known issues and notes . . . . .	48
1.31	Current and future work . . . . .	49
1.32	Tentative and experimental code . . . . .	49
<b>2</b>	<b>Loading languages with language.dat</b>	<b>50</b>
2.1	Format . . . . .	50
<b>3</b>	<b>The interface between the core of babel and the language definition files</b>	<b>51</b>
3.1	Guidelines for contributed languages . . . . .	52
3.2	Basic macros . . . . .	52
3.3	Skeleton . . . . .	53
3.4	Support for active characters . . . . .	54
3.5	Support for saving macro definitions . . . . .	55
3.6	Support for extending macros . . . . .	55
3.7	Macros common to a number of languages . . . . .	55
3.8	Encoding-dependent strings . . . . .	56
3.9	Executing code based on the selector . . . . .	59
<b>II</b>	<b>Source code</b>	<b>59</b>
<b>4</b>	<b>Identification and loading of required files</b>	<b>59</b>
<b>5</b>	<b>locale directory</b>	<b>60</b>

<b>6</b>	<b>Tools</b>	<b>60</b>
6.1	Multiple languages	64
6.2	The Package File ( <code>\LaTeX</code> , <code>babel.sty</code> )	65
6.3	<code>base</code>	66
6.4	<code>key=value</code> options and other general option	67
6.5	Conditional loading of shorthands	68
6.6	Interlude for Plain	70
<b>7</b>	<b>Multiple languages</b>	<b>70</b>
7.1	Selecting the language	72
7.2	Errors	80
7.3	Hooks	83
7.4	Setting up language files	84
7.5	Shorthands	86
7.6	Language attributes	95
7.7	Support for saving macro definitions	97
7.8	Short tags	98
7.9	Hyphens	98
7.10	Multiencoding strings	100
7.11	Macros common to a number of languages	106
7.12	Making glyphs available	106
7.12.1	Quotation marks	106
7.12.2	Letters	108
7.12.3	Shorthands for quotation marks	109
7.12.4	Umlauts and tremas	109
7.13	Layout	110
7.14	Load engine specific macros	111
7.15	Creating and modifying languages	111
<b>8</b>	<b>Adjusting the Babel behavior</b>	<b>132</b>
8.1	Cross referencing macros	133
8.2	Marks	136
8.3	Preventing clashes with other packages	137
8.3.1	<code>ifthen</code>	137
8.3.2	<code>varioref</code>	138
8.3.3	<code>hhline</code>	138
8.4	Encoding and fonts	139
8.5	Basic bidi support	140
8.6	Local Language Configuration	144
8.7	Language options	144
<b>9</b>	<b>The kernel of Babel (<code>babel.def</code>, <code>common</code>)</b>	<b>147</b>
<b>10</b>	<b>Loading hyphenation patterns</b>	<b>147</b>
<b>11</b>	<b>Font handling with <code>fontspec</code></b>	<b>151</b>
<b>12</b>	<b>Hooks for XeTeX and LuaTeX</b>	<b>156</b>
12.1	XeTeX	156
12.2	Layout	157
12.3	LuaTeX	159
12.4	Southeast Asian scripts	165
12.5	CJK line breaking	166
12.6	Arabic justification	168
12.7	Common stuff	172
12.8	Automatic fonts and ids switching	172
12.9	Bidi	177
12.10	Layout	179
12.11	Lua: transforms	184

12.12 Lua: Auto bidi with basic and basic-r . . . . .	192
<b>13 Data for CJK</b>	<b>202</b>
<b>14 The ‘nil’ language</b>	<b>202</b>
<b>15 Support for Plain T<sub>E</sub>X (plain.def)</b>	<b>203</b>
15.1 Not renaming hyphen.tex . . . . .	203
15.2 Emulating some L <sup>A</sup> T <sub>E</sub> X features . . . . .	203
15.3 General tools . . . . .	204
15.4 Encoding related macros . . . . .	207
<b>16 Acknowledgements</b>	<b>210</b>

## Troubleshooting

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

# Part I

## User guide

**What is this document about?** This user guide focuses on internationalization and localization with  $\LaTeX$  and pdf $\TeX$ , 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:

PDF $\TeX$

```
\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, T<sub>E</sub>XLive, 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 is 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 pdf<sub>tex</sub> follows. The main language is french, which is activated when the document begins. It assumes UTF-8:

PDF<sub>TEX</sub>

```
\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:



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

\babelfont[russian]{rm}{FreeSerif}

\begin{document}

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

\end{document}
```

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

## 1.4 Modifiers

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

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

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

## 1.5 Troubleshooting

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

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

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

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

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

<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.

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

## 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.

**NOTE** Bear in mind `\selectlanguage` can be automatically executed, in some cases, in the auxiliary files, at heads and foots, and after the environment `otherlanguage*`.

**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.

**WARNING** There are a couple of issues related to the way the language information is written to the auxiliary files:

- `\selectlanguage` should not be used inside some boxed environments (like floats or `minipage`) to switch the language if you need the information written to the aux be correctly synchronized. This rarely happens, but if it were the case, you must use `otherlanguage` instead.
- In addition, this macro inserts a `\write` in vertical mode, which may break the vertical spacing in some cases (for example, between lists). **New 3.64** The behavior can be adjusted with `\babeladjust{select.write=<mode>}`, where `<mode>` is `shift` (which shifts the skips down and adds a `\penalty`); `keep` (the default – with it the `\write` and the skips are kept in the order they are written), and `omit` (which may seem a too drastic solution, because nothing is written, but more often than not this command is applied to more or less short texts with no sectioning or similar commands and therefore no language synchronization is necessary).

**`\foreignlanguage`** [*⟨option-list⟩*]{*⟨language⟩*}{*⟨text⟩*}

The command `\foreignlanguage` takes two arguments; the second argument is a phrase to be typeset according to the rules of the language named in its first one.

This command (1) only switches the extra definitions and the hyphenation rules for the language, *not* the names and dates, (2) does not send information about the language to auxiliary files (i.e., the surrounding language is still in force), and (3) it works even if the language has not been set as package option (but in such a case it only sets the hyphenation patterns and a warning is shown). With the `bidi` option, it also enters in horizontal mode (this is not done always for backwards compatibility), and since it is meant for phrases only the text direction (and not the paragraph one) is set.

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

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

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

## 1.8 Auxiliary language selectors

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

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

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

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

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

Spaces after the environment are ignored.

**`\begin{otherlanguage*}`** [*⟨option-list⟩*]{*⟨language⟩*} ... **`\end{otherlanguage*}`**

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

This environment was originally intended for intermixing left-to-right typesetting with right-to-left typesetting in engines not supporting a change in the writing direction inside a line. However, by default it never complied with the documented behavior and it is just a version as environment of `\foreignlanguage`, except when the option `bidi` is set – in this case, `\foreignlanguage` emits a `\leavevmode`, while `otherlanguage*` does not.

## 1.9 More on selection

**`\babeltags`** {*⟨tag1⟩* = *⟨language1⟩*, *⟨tag2⟩* = *⟨language2⟩*, ...}

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

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

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

**EXAMPLE** With

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

you can write

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

and

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

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

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

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

**New 3.9i** Except in a few languages, like `ruussian`, 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{ruussian}{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,  $\text{\TeX}$  of `\dag`). With `ini` files (see below), captions are ensured by default.

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

## 1.10 Shorthands

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

**NOTE** Keep in mind the following:

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

**TROUBLESHOOTING** A typical error when using shorthands is the following:

```
! Argument of \language@active@arg" has an extra }.
```

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

`\shorthandon`  $\{\langle shorthands-list \rangle\}$   
`\shorthandoff`  $\ast\{\langle shorthands-list \rangle\}$

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

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

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

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

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

**WARNING** It is worth emphasizing these macros are meant for temporary changes. Whenever possible and if there are not conflicts with other packages, shorthands must be always enabled (or disabled).

`\useshortands` `*{\langle char \rangle}`

The command `\useshortands` 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 `\useshortands*{\langle char \rangle}` is provided, which makes sure shorthands are always activated.

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

`\defineshortand` `[\langle language \rangle, \langle language \rangle, ...]{\langle shorthand \rangle}{\langle code \rangle}`

The command `\defineshortand` 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 `\languageshortands{\langle lang \rangle}` to the corresponding `\extras{\langle lang \rangle}`, as explained below). By default, user shorthands are (re)defined.

User shorthands override language ones, which in turn override system shorthands.

Language-dependent user shorthands (new in 3.9) take precedence over “normal” user shorthands.

**EXAMPLE** Let’s assume you want a unified set of shorthand for dictionaries (languages do not define shorthands consistently, and “-”, “\”, “=” have different meanings). You can start with, say:

```
\useshortands*{"}
\defineshortand{"*}{\babelhyphen{soft}}
\defineshortand{"-}{\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:

```
\defineshortand[*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.

`\languageshortands` `{\langle language \rangle}`

The command `\languageshortands` 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{\languageshortands{ngerman}}
```

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

<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.

**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]{\{\language shorthands{none}\tipaencoding#1}}
```

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

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

**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  
**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** `{\langle character \rangle}{\langle true \rangle}{\langle false \rangle}`

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

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

The command `\aliasshorthand` can be used to let another character perform the same functions as the default shorthand character. If one prefers for example to use the

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

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

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.

- KeepShorthandsActive** Tells babel not to deactivate shorthands after loading a language file, so that they are also available in the preamble.
- activeacute** For some languages babel supports this options to set ' as a shorthand in case it is not done by default.
- activegrave** Same for `.
- shorthands=** `<char><char>... | off`  
The only language shorthands activated are those given, like, eg:

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

If ' is included, `activeacute` is set; if ` is included, `activegrave` is set. Active characters (like `~`) should be preceded by `\string` (otherwise they will be expanded by  $\TeX$  before they are passed to the package and therefore they will not be recognized); however, `t` is provided for the common case of `~` (as well as `c` for not so common case of the comma). With `shorthands=off` no language shorthands are defined. As some languages use this mechanism for tools not available otherwise, a macro `\babelshorthand` is defined, which allows using them; see above.

- safe=** `none | ref | bib`  
Some  $\TeX$  macros are redefined so that using shorthands is safe. With `safe=bib` only `\nocite`, `\bibcite` and `\bibitem` are redefined. With `safe=ref` only `\newlabel`, `\ref` and `\pageref` are redefined (as well as a few macros from `varioref` and `ifthen`). With `safe=none` no macro is redefined. This option is strongly recommended, because a good deal of incompatibilities and errors are related to these redefinitions. As of **New 3.34**, in  $\epsilon\TeX$  based engines (ie, almost every engine except the oldest ones) shorthands can be used in these macros (formerly you could not).
- math=** `active | normal`  
Shorthands are mainly intended for text, not for math. By setting this option with the value `normal` they are deactivated in math mode (default is `active`) and things like `#{a'}` (a closing brace after a shorthand) are not a source of trouble anymore.



- config=** *<file>*  
Load *<file>*.cfg instead of the default config file `bblopts.cfg` (the file is loaded even with `noconfigs`).
- main=** *<language>*  
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.
- headfoot=** *<language>*  
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.
- noconfigs** Global and language default config files are not loaded, so you can make sure your document is not spoilt by an unexpected .cfg file. However, if the key `config` is set, this file is loaded.
- 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, LICR and ASCII strings), `unicode` (for engines like xetex and luatex) and `encoded` (for special cases requiring mixed encodings). Other allowed values are font encoding codes (T1, T2A, LGR, L7X...), but only in languages supporting them. Be aware with encoded captions are protected, but they work in `\MakeUpper case` 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 `other language`;
  - other\*** also sets it at `other language*` 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>

<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.

**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.

**provide=** \*

**New 3.49** An alternative to `\babelprovide` for languages passed as options. See section 1.13, which describes also the variants `provide+=` and `provide*=`.

## 1.12 The base option

With this package option `babel` just loads some basic macros (those in `switch.def`), defines `\AfterBabelLanguage` and exits. It also selects the hyphenation patterns for the last language passed as option (by its name in `language.dat`). There are two main uses: classes and packages, and as a last resort in case there are, for some reason, incompatible languages. It can be used if you just want to select the hyphenation patterns of a single language, too.

**\AfterBabelLanguage** `{<option-name>}{<code>}`

This command is currently the only provided by `base`. Executes `<code>` 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 `<option-name>` 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}
```

**NOTE** With a recent version of  $\text{\TeX}$ , an alternative method to execute some code just after an `ldf` file is loaded is with `\AddToHook` and the hook `file/<language>.ldf/after`. `Babel` does not predeclare it, and you have to do it yourself with `\ActivateGenericHook`.

**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 250 of these files containing the basic data required for a locale, plus basic templates for 500 about locales.

`ini` files are not meant only for `babel`, and they have been devised as a resource for other packages. To easy interoperability between  $\text{\TeX}$  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 math and 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).

**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 are hard-coded in xetex, but they can be modified in luatex). 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]{lṇ lṃ lṂ lṁ lṅ lṇ lṅ} % Random
```

**East Asia scripts** Settings for either Simplified or Traditional should work out of the box, with basic line breaking with any renderer. Although for a few words and shorts texts the ini files should be fine, CJK texts are best set with a dedicated framework (CJK, luatexja, kotex, CTeX, etc.). This is what the class ltjbook does with luatex, which can be used in conjunction with the ldf for japanese, because the following piece of code loads luatexja:

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

**Latin, Greek, Cyrillic** Combining chars with the default luatex font renderer might be wrong; on the other hand, with the Harfbuzz renderer diacritics are stacked correctly, but many hyphenations points are discarded (this bug is 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>	bem	Bemba
agq	Aghem	bez	Bena
ak	Akan	bg	Bulgarian <sup>ul</sup>
am	Amharic <sup>ul</sup>	bm	Bambara
ar	Arabic <sup>ul</sup>	bn	Bangla <sup>ul</sup>
ar-DZ	Arabic <sup>ul</sup>	bo	Tibetan <sup>u</sup>
ar-MA	Arabic <sup>ul</sup>	brx	Bodo
ar-SY	Arabic <sup>ul</sup>	bs-Cyrl	Bosnian
as	Assamese	bs-Latn	Bosnian <sup>ul</sup>
asa	Asu	bs	Bosnian <sup>ul</sup>
ast	Asturian <sup>ul</sup>	ca	Catalan <sup>ul</sup>
az-Cyrl	Azerbaijani	ce	Chechen
az-Latn	Azerbaijani	cgg	Chiga
az	Azerbaijani <sup>ul</sup>	chr	Cherokee
bas	Basaa	ckb	Central Kurdish
be	Belarusian <sup>ul</sup>	cop	Coptic

cs	Czech <sup>ul</sup>	hsb	Upper Sorbian <sup>ul</sup>
cu	Church Slavic	hu	Hungarian <sup>ul</sup>
cu-Cyrs	Church Slavic	hy	Armenian <sup>u</sup>
cu-Glag	Church Slavic	ia	Interlingua <sup>ul</sup>
cy	Welsh <sup>ul</sup>	id	Indonesian <sup>ul</sup>
da	Danish <sup>ul</sup>	ig	Igbo
dav	Taita	ii	Sichuan Yi
de-AT	German <sup>ul</sup>	is	Icelandic <sup>ul</sup>
de-CH	German <sup>ul</sup>	it	Italian <sup>ul</sup>
de	German <sup>ul</sup>	ja	Japanese
dje	Zarma	jgo	Ngomba
dsb	Lower Sorbian <sup>ul</sup>	jmc	Machame
dua	Duala	ka	Georgian <sup>ul</sup>
dyo	Jola-Fonyi	kab	Kabyle
dz	Dzongkha	kam	Kamba
ebu	Embu	kde	Makonde
ee	Ewe	kea	Kabuverdianu
el	Greek <sup>ul</sup>	khq	Koyra Chiini
el-polyton	Polytonic Greek <sup>ul</sup>	ki	Kikuyu
en-AU	English <sup>ul</sup>	kk	Kazakh
en-CA	English <sup>ul</sup>	kkj	Kako
en-GB	English <sup>ul</sup>	kl	Kalaallisut
en-NZ	English <sup>ul</sup>	kln	Kalenjin
en-US	English <sup>ul</sup>	km	Khmer
en	English <sup>ul</sup>	kn	Kannada <sup>ul</sup>
eo	Esperanto <sup>ul</sup>	ko	Korean
es-MX	Spanish <sup>ul</sup>	kok	Konkani
es	Spanish <sup>ul</sup>	ks	Kashmiri
et	Estonian <sup>ul</sup>	ksb	Shambala
eu	Basque <sup>ul</sup>	ksf	Bafia
ewo	Ewondo	ksh	Colognian
fa	Persian <sup>ul</sup>	kw	Cornish
ff	Fulah	ky	Kyrgyz
fi	Finnish <sup>ul</sup>	lag	Langi
fil	Filipino	lb	Luxembourgish
fo	Faroese	lg	Ganda
fr	French <sup>ul</sup>	lkt	Lakota
fr-BE	French <sup>ul</sup>	ln	Lingala
fr-CA	French <sup>ul</sup>	lo	Lao <sup>ul</sup>
fr-CH	French <sup>ul</sup>	lrc	Northern Luri
fr-LU	French <sup>ul</sup>	lt	Lithuanian <sup>ul</sup>
fur	Friulian <sup>ul</sup>	lu	Luba-Katanga
fy	Western Frisian	luo	Luo
ga	Irish <sup>ul</sup>	luy	Luyia
gd	Scottish Gaelic <sup>ul</sup>	lv	Latvian <sup>ul</sup>
gl	Galician <sup>ul</sup>	mas	Masai
grc	Ancient Greek <sup>ul</sup>	mer	Meru
gsw	Swiss German	mfe	Morisyen
gu	Gujarati	mg	Malagasy
guz	Gusii	mgh	Makhuwa-Meetto
gv	Manx	mgo	Meta'
ha-GH	Hausa	mk	Macedonian <sup>ul</sup>
ha-NE	Hausa <sup>l</sup>	ml	Malayalam <sup>ul</sup>
ha	Hausa	mn	Mongolian
haw	Hawaiian	mr	Marathi <sup>ul</sup>
he	Hebrew <sup>ul</sup>	ms-BN	Malay <sup>l</sup>
hi	Hindi <sup>u</sup>	ms-SG	Malay <sup>l</sup>
hr	Croatian <sup>ul</sup>	ms	Malay <sup>ul</sup>

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

In some contexts (currently `\babel font`) an ini file may be loaded by its name. Here is the list of the names currently supported. With these languages, `\babel font` loads (if not done before) the language and script names (even if the language is defined as a package option

with an ldf file). These are also the names recognized by `\babelprovide` with a valueless `import`.

---

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

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

french-luxembourg	lowersorbian
french-switzerland	lsorbian
french	lubakatanga
friulian	luo
fulah	luxembourgish
galician	luyia
ganda	macedonian
georgian	machame
german-at	makhuwameetto
german-austria	makonde
german-ch	malagasy
german-switzerland	malay-bn
german	malay-brunei
greek	malay-sg
gujarati	malay-singapore
gusii	malay
hausa-gh	malayalam
hausa-ghana	maltese
hausa-ne	manx
hausa-niger	marathi
hausa	masai
hawaiian	mazanderani
hebrew	meru
hindi	meta
hungarian	mexican
icelandic	mongolian
igbo	morisyen
inarisami	mundang
indonesian	nama
interlingua	nepali
irish	newzealand
italian	ngiemboon
japanese	ngomba
jolafonyi	norsk
kabuverdianu	northernluri
kabyle	northernsami
kako	northndebele
kalaallisut	norwegianbokmal
kalenjin	norwegiannynorsk
kamba	nswissgerman
kannada	nuer
kashmiri	nyankole
kazakh	nynorsk
khmer	occitan
kikuyu	oriya
kinyarwanda	oromo
konkani	ossetic
korean	pashto
koyraborosenni	persian
koyrachiini	piedmontese
kwasio	polish
kyrgyz	polytonicgreek
lakota	portuguese-br
langi	portuguese-brazil
lao	portuguese-portugal
latvian	portuguese-pt
lingala	portuguese
lithuanian	punjabi-arab



punjabi-arabic	soga
punjabi-gurmukhi	somali
punjabi-guru	spanish-mexico
punjabi	spanish-mx
quechua	spanish
romanian	standardmoroccantamazight
romansh	swahili
rombo	swedish
rundi	swissgerman
russian	tachelhit-latin
rwa	tachelhit-latn
sakha	tachelhit-tfng
samburu	tachelhit-tifinagh
samin	tachelhit
sango	taita
sangu	tamil
sanskrit-beng	tasawaq
sanskrit-bengali	telugu
sanskrit-deva	teso
sanskrit-devanagari	thai
sanskrit-gujarati	tibetan
sanskrit-gujr	tigrinya
sanskrit-kannada	tongan
sanskrit-knda	turkish
sanskrit-malayalam	turkmen
sanskrit-mlym	ukenglish
sanskrit-telu	ukrainian
sanskrit-telugu	uppersorbian
sanskrit	urdu
scottishgaelic	usenglish
sena	usorbian
serbian-cyrillic-bosniaherzegovina	uyghur
serbian-cyrillic-kosovo	uzbek-arab
serbian-cyrillic-montenegro	uzbek-arabic
serbian-cyrillic	uzbek-cyrillic
serbian-cyrl-ba	uzbek-cyrl
serbian-cyrl-me	uzbek-latin
serbian-cyrl-xk	uzbek-latn
serbian-cyrl	uzbek
serbian-latin-bosniaherzegovina	vai-latin
serbian-latin-kosovo	vai-latn
serbian-latin-montenegro	vai-vai
serbian-latin	vai-vaii
serbian-latn-ba	vai
serbian-latn-me	vietnam
serbian-latn-xk	vietnamese
serbian-latn	vunjo
serbian	walser
shambala	welsh
shona	westernfrisian
sichuanyi	yangben
sinhala	yiddish
slovak	yoruba
slovene	zarma
slovenian	zulu afrikaans

---

### Modifying and adding values to ini files

**New 3.39** There is a way to modify the values of ini files when they get loaded with

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

## 1.14 Selecting fonts

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

LUATEX/XETEX

```
\documentclass{article}

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

\babelprovide[import]{hebrew}

\babelfont{rm}{FreeSerif}

\begin{document}

Svenska \foreignlanguage{hebrew}{עִבְרִית} 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}
```

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

`\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* an 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* an 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.

**NOTE** `\babelfont` is a high level interface to `fontspec`, and therefore in `xetex` you can apply Mappings. For example, there is a set of [transliterations for Brahmic scripts](#) by Davis M. Jones. After installing them in your distribution, just set the map as you would do with `fontspec`.

## 1.15 Modifying a language

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

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

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

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

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

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

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

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

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

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

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

As of 3.15, there is no need to hide spaces with % (babel removes them), but it is advisable to do so. This redefinition is not activated until the language is selected.

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

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

This redefinition is immediate.

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

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

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

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

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

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

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

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

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

first loads `danish.ldf`, and then redefines the captions for danish (as provided by the `ini` file) and prevents hyphenation. The rest of the language definitions are not touched. Without the optional argument it just loads some additional tools if provided by the `ini` file, like extra counters.

## 1.16 Creating a language

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

`\babelprovide` [*options*]{*language-name*}

If the language *language-name* has not been loaded as class or package option and there are no *options*, it creates an “empty” one with some defaults in its internal structure: the hyphen rules, if not available, are set to the current ones, left and right hyphen mins are set to 2 and 3. In either case, caption, date and language system are not defined. If no `ini` file is imported with `import`, *language-name* is still relevant because in such a case the hyphenation and like breaking rules (including those for South East Asian and CJK) are based on it as provided in the `ini` file corresponding to that name; the same applies to OpenType language and script. Conveniently, some options allow to fill the language, and `babel` warns you about what to do if there is a missing string. Very likely you will find alerts like that in the log file:

```
Package babel Warning: \chaptername not set for 'mylang'. Please,
(babel)                define it after the language has been loaded
(babel)                (typically in the preamble) with:
(babel)                \setlocalecaption{mylang}{chapter}{..}
(babel)                Reported on input line 26.
```

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

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

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

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

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

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

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

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

**New 3.13** Imports data from an ini file, including captions and date (also line breaking rules in newly defined languages). For example:

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

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

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

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

There are about 250 ini files, with data taken from the ldf files and the CLDR provided by Unicode. Not all languages in the latter are complete, and therefore neither are the ini files. A few languages may show a warning about the current lack of suitability of some features.

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

\<language>date{\the\year}{\the\month}{\the\day}. **New 3.44** More convenient is usually \localedate, which prints the date for the current locale.

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

Loads only the strings. For example:

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

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

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

```
\babelprovide[hyphenrules=chavacano spanish italian]{chavacano}
```

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

A special value is +, which allocates a new language (in the 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 (xetex or luatex) is mainly in Polytonic Greek with 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, polytonicgreek, provide=*]{babel}
```

Remember there is an alternative syntax for the latter:

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

Finally, also remember you might not need to load `italian` at all if there are only a few word in this language (see 1.3).

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

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

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

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

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

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

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

Same for `\Alph`.

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

**onchar=** ids | fonts

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

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

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

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

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

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

**transforms=**  $\langle transform-list \rangle$

See section 1.21.

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

Languages providing native digits in all or some variants are:



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

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

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

`\localenumeral`  $\{\langle style \rangle\}\{\langle number \rangle\}$   
`\localecounter`  $\{\langle style \rangle\}\{\langle counter \rangle\}$

**New 3.41** 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):

- `\localenumeral{\langle style \rangle}\{\langle number \rangle\}`, like `\localenumeral{abjad}{15}`
- `\localecounter{\langle style \rangle}\{\langle counter \rangle\}`, 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`  
**Armenian** `lower.letter`, `upper.letter`  
**Belarusan, Bulgarian, Church Slavic, Macedonian, Serbian** `lower`, `upper`  
**Bengali** `alphabetic`  
**Central Kurdish** `alphabetic`  
**Chinese** `cjk-earthly-branch`, `cjk-heavenly-stem`, `circled.ideograph`, `parenthesized.ideograph`, `fullwidth.lower.alpha`, `fullwidth.upper.alpha`  
**Church Slavic (Glagolitic)** `letters`  
**Coptic** `epact`, `lower.letters`  
**French** `date.day` (mainly for internal use).  
**Georgian** `letters`  
**Greek** `lower.modern`, `upper.modern`, `lower.ancient`, `upper.ancient` (all with `keraia`)  
**Hebrew** `letters` (neither `geresh` nor `gershayim yet`)  
**Hindi** `alphabetic`  
**Italian** `lower.legal`, `upper.legal`  
**Japanese** `hiragana`, `hiragana.iroha`, `katakana`, `katakana.iroha`, `circled.katakana`, `informal`, `formal`, `cjk-earthly-branch`, `cjk-heavenly-stem`, `circled.ideograph`, `parenthesized.ideograph`, `fullwidth.lower.alpha`, `fullwidth.upper.alpha`

**Khmer** consonant  
**Korean** consonant, syllable, hanja.informal, hanja.formal, hangul.formal, cjk-earthly-branch, cjk-heavenly-stem, circled.ideograph, parenthesized.ideograph, 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

**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 an ini file 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=izafa it prints 31'ê *Çileyä Pêşînê 2019*.

## 1.19 Accessing language info

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

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

**\iflanguage** {*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 פרק.

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

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

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 (when it makes sense) as an attribute, too.

**`\LocaleForEach`** `{\code}`

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.

**`\BabelEnsureInfo`** ini files are loaded with `\babelprovide` and also when languages are selected if there is a `\babelfont` or they have not been explicitly declared. 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.

## 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. With `luatex` there are also tools for non-standard hyphenation rules, explained in the next section.

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

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

**New 3.9a** Sets hyphenation exceptions for the languages given or, without the optional argument, for *all* languages (eg, proper nouns or common loan words, and of course monolingual documents). Multiple declarations work much like `\hyphenation` (last wins), but 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 `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}
```

**New 3.67** Transforms predefined in the ini locale files can be made attribute-dependent, too. When an attribute between parenthesis is inserted subsequent transforms will be assigned to it (up to the list end or another attribute). For example, and provided an attribute called `\withsigmafinal` has been declared:

```
transforms = transliteration.omega (\withsigmafinal) sigma.final
```

<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.

<sup>15</sup>They are similar in concept, but not the same, as those in Unicode. The main inspiration for this feature is the Omega transformation processes.

This applies `transliteration.omega` always, but `sigma.final` only when `\withsigmafinal` is set.  
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.
Finnish	<code>prehyphen.nobreak</code>	Line breaks just after hyphens prepended to words are prevented, like in “pakastekaapit ja -arkut”.
Greek	<code>diaeresis.hyphen</code>	Removes the diaeresis above iota and upsilon if hyphenated just before. It works with the three variants.
Greek	<code>transliteration.omega</code>	Although the provided combinations are not the full set, this transform follows the syntax of Omega: = for the circumflex, v for digamma, and so on. For better compatibility with Levy’s system, ~ (as ‘string’) is an alternative to =. ' is tonos in Monotonic Greek, but oxia in Polytonic and Ancient Greek.
Greek	<code>sigma.final</code>	The transliteration system above does not convert the sigma at the end of a word (on purpose). This transform does it. To prevent the conversion (an abbreviation, for example), write “s.
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: !?;.
Hungarian	<code>digraphs.hyphen</code>	Hyphenates the long digraphs <i>ccs, ddz, ggy, lly, nny, ssz, tty</i> and <i>zzs</i> as <i>cs-cs, dz-dz</i> , etc.
Indic scripts	<code>danda.nobreak</code>	Prevents a line break before a danda or double danda if there is a space. For Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Oriya, Tamil, Telugu.
Latin	<code>digraphs.ligatures</code>	Replaces the groups <i>ae, AE, oe, OE</i> with <i>æ, Æ, œ, Æ</i> .
Latin	<code>letters.noj</code>	Replaces <i>j, J</i> with <i>i, I</i> .
Latin	<code>letters.uv</code>	Replaces <i>v, U</i> with <i>u, V</i> .
Sanskrit	<code>transliteration.iast</code>	The IAST system to romanize Devanagari. <sup>16</sup>

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.
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.

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

**New 3.37-3.39** With *luatex* it is 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. A few rules are currently provided (see above), 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`.

**New 3.67** With the optional argument you can associate a user defined transform to an attribute, so that it’s active only when it’s set (currently its attribute value is ignored). With this mechanism transforms can be set or unset even in the middle of paragraphs, and applied to single words. To define, set and unset the attribute, the LaTeX kernel provides the macros `\newattribute`, `\setattribute` and `\unsetattribute`. The following example shows how to use it, provided an attribute named `\latinnoj` has been declared:

```
\babelprehyphenation[attribute=\latinnoj]{latin}{ J }{ string = I }
```

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** [*<options>*]{*<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.

See the description above for the optional argument.

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>17</sup>

Some languages sharing the same script define macros to switch it (eg, `\textcyrillic`), but be aware they may also set the language to a certain default. Even the `babel` core defined `\textlatin`, but 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>18</sup>

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

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

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

<sup>17</sup>The so-called Unicode fonts do not improve the situation either. So, a font suited for Vietnamese is not necessarily suited for, say, the romanization of Indic languages, and the fact it contains glyphs for Modern Greek does not mean it includes them for Classic Greek.

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

The symbol encodings TS1, T3, and TS3 are not taken into account, since they are not used for “ordinary” text (they are stored in `\BabelNonText`, used in some special cases when no Latin encoding is explicitly set).

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

## 1.24 Selecting directions

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

**WARNING** The current code for `text` in `luatex` should be considered essentially stable, but, of course, it is not bug-free and there can be improvements in the future, because setting bidi text has many subtleties (see for example <https://www.w3.org/TR/html-bidi/>). A basic stable version for other engines must wait. This applies to text; there is a basic support for **graphical** elements, including the picture environment (with `pict2e`) and `pfg/tikz`. Also, indexes and the like are under study, as well as math (there are progresses in the latter, including `amsmath` and `mathtools` too, but for example gathered 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 (الاجريقي) بـ
    Arabia أو Aravia (بالاغريقية Αραβία)، استخدم الرومان ثلاث
    بادئات بـ“Arabia” على ثلاث مناطق من شبه الجزيرة العربية، إلا أنها
    حقيقةً كانت أكبر مما تعرف عليه اليوم.

\end{document}

```

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

```

\documentclass{book}

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

\babelprovide[onchar=ids fonts]{arabic}

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

\begin{document}

    Most Arabic speakers consider the two varieties to be two registers
    of one language, although the two registers can be referred to in
    Arabic as \textit{fuṣḥā l-‘aṣr} (MSA) and
\textit{fuṣḥā t-turāth} (CA).

\end{document}

```

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

**NOTE** Boxes are “black boxes”. Numbers inside an `\hbox` (for example in a `\ref`) do not know anything about the surrounding chars. So, `\ref{A}-\ref{B}` are not rendered in the visual order A-B, but in the wrong one B-A (because the hyphen does not “see” the digits inside the `\hbox`’es). If you need `\ref` ranges, the best option is to define a dedicated macro like this (to avoid explicit direction changes in the body; here `\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>19</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 T<sub>E</sub>X 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}`

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

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

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

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

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

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

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

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

**New 3.17** Something like:

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

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

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

(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.

**New 3.64** This is not the only way to inject code at those points. The events listed below can be used as a hook name in `\AddToHook` in the form `babel/⟨language-name⟩/⟨event-name⟩` (with \* it's applied to all languages), but there is a limitation, because the parameters passed with the babel mechanism are not allowed. The `\AddToHook` mechanism does *not* replace the current one in 'babel'. Its main advantage is you can reconfigure 'babel' even before loading it. See the example below.

`\AddBabelHook` `[⟨lang⟩]{⟨name⟩}{⟨event⟩}{⟨code⟩}`

The same name can be applied to several events. Hooks with a certain `{⟨name⟩}` may be enabled and disabled for all defined events with `\EnableBabelHook{⟨name⟩}`, `\DisableBabelHook{⟨name⟩}`. Names containing the string `babel` are reserved (they are used, for example, by `\useshortands*` to add a hook for the event `afterextras`).

**New 3.33** They may be also applied to a specific language with the optional argument; language-specific settings are executed after global ones.

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

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

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

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

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

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

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

**write** This event comes just after the switching commands are written to the aux file.

**beforeextras** Just before executing `\extras⟨language⟩`. This event and the next one should not contain language-dependent code (for that, add it to `\extras⟨language⟩`).

**afterextras** Just after executing `\extras⟨language⟩`. For example, the following deactivates shorthands in all languages:

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

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

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

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

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

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

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

**loadkernel** (file) By default just defines a few basic commands. It can be used to define different versions of them or to load a file.

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

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

**EXAMPLE** The generic unlocalized  $\TeX$  hooks are predefined, so that you can write:

```
\AddToHook{babel/*/afterextras}{\frenchspacing}
```

which is executed always after the extras for the language being selected (and just before the non-localized hooks defined with `\AddBabelHook`).

In addition, locale-specific hooks in the form `babel/⟨language-name⟩/⟨event-name⟩` are *recognized* (executed just before the localized babel hooks), but they are *not predefined*. You have to do it yourself. For example, to set `\frenchspacing` only in bengali:

```
\ActivateGenericHook{babel/bengali/afterextras}
\AddToHook{babel/bengali/afterextras}{\frenchspacing}
```

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

## 1.27 Languages supported by babel with ldf files

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

**Afrikaans** afrikaans  
**Azerbaijani** azerbaijani  
**Basque** basque  
**Breton** breton  
**Bulgarian** bulgarian  
**Catalan** catalan  
**Croatian** croatian  
**Czech** czech  
**Danish** danish  
**Dutch** dutch  
**English** english, USenglish, american, UKenglish, british, canadian, australian, newzealand  
**Esperanto** esperanto  
**Estonian** estonian  
**Finnish** finnish  
**French** french, francais, canadien, acadian  
**Galician** galician



**German** austrian, german, germanb, ngerman, naustrian  
**Greek** greek, polutonikogreek  
**Hebrew** hebrew  
**Icelandic** icelandic  
**Indonesian** indonesian (bahasa, indon, bahasai)  
**Interlingua** interlingua  
**Irish Gaelic** irish  
**Italian** italian  
**Latin** latin  
**Lower Sorbian** lowersorbian  
**Malay** malay, melayu (bahasam)  
**North Sami** samin  
**Norwegian** norsk, nynorsk  
**Polish** polish  
**Portuguese** portuguese, brazilian (portuges, brazil)<sup>20</sup>  
**Romanian** romanian  
**Russian** russian  
**Scottish Gaelic** scottish  
**Spanish** spanish  
**Slovakian** slovak  
**Slovenian** slovene  
**Swedish** swedish  
**Serbian** serbian  
**Turkish** turkish  
**Ukrainian** ukrainian  
**Upper Sorbian** upporsorbian  
**Welsh** welsh

There are more languages not listed above, including hindi, thai, thaicjk, latvian, turkmen, magyar, mongolian, romansh, lithuanian, spanglish, vietnamese, japanese, pinyin, arabic, farsi, ibygreek, bgreek, serbianc, frenchle, ethiop and friulan. Most of them work out of the box, but some may require extra fonts, encoding files, a preprocessor or even a complete framework (like CJK or luatexja). For example, if you have got the velthuis/devnag package, you can create a file with extension .dn:

```

\documentclass{article}
\usepackage[hindi]{babel}
\begin{document}
{\dn devaanaa.m priya.h}
\end{document}

```

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

## 1.28 Unicode character properties in luatex

**New 3.32** Part of the babel job is to apply Unicode rules to some script-specific features based on some properties. Currently, they are 3, namely, direction (ie, bidi class), mirroring glyphs, and line breaking for CJK scripts. These properties are stored in lua tables, which you can modify with the following macro (for example, to set them for glyphs in the PUA).

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

**New 3.32** Here,  $\{\langle char-code \rangle\}$  is a number (with  $\TeX$  syntax). With the optional argument, you can set a range of values. There are three properties (with a short name, taken from Unicode): direction (bc), mirror (bmg), linebreak (lb). The settings are global, and this command is allowed only in vertical mode (the preamble or between paragraphs).

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



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`),  $\LaTeX$  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\extrasrussian{\inputencoding{koi8-r}}
```

- For the hyphenation to work correctly, `lccodes` cannot change, because  $\TeX$  only takes into account the values when the paragraph is hyphenated, i.e., when it has been finished.<sup>21</sup> So, if you write a chunk of French text with `\foreignlanguage`, the apostrophes might not be taken into account. This is a limitation of  $\TeX$ , not of `babel`. Alternatively, you may use `\useshorthands` to activate `'` and `\defineshortand`, or redefine `\textquoteright` (the latter is called by the non-ASCII right quote).

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

- `\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 a similar issue with floats, too. 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  $\TeX$  enter in an infinite loop in some rare cases. (Another issue in the 'to do' list, although there is a partial solution.)

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

**csquotes** Logical markup for quotes.

**iflang** Tests correctly the current language.

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

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

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

**biblatex** Programmable bibliographies and citations.

**bicaption** Bilingual captions.

**babelbib** Multilingual bibliographies.

**microtype** Adjusts the typesetting according to some languages (kerning and spacing).  
Ligatures can be disabled.

**substitutefont** Combines fonts in several encodings.

**mkpattern** Generates hyphenation patterns.

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

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

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

### 1.31 Current and future work

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

Useful additions would be, for example, time, currency, addresses and personal names.<sup>22</sup>.

But that is the easy part, because they don't require modifying the  $\TeX$  internals.

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

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

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 and deprecated functions, see the babel site.

#### Options for locales loaded on the fly

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

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

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

### Labels

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

## 2 Loading languages with language.dat

T<sub>E</sub>X and most engines based on it (pdfT<sub>E</sub>X, xetex,  $\epsilon$ -T<sub>E</sub>X, the main exception being luatex) require hyphenation patterns to be preloaded when a format is created (eg, L<sup>A</sup>T<sub>E</sub>X, XeL<sup>A</sup>T<sub>E</sub>X, pdfL<sup>A</sup>T<sub>E</sub>X). babel provides a tool which has become standard in many distributions and based on a “configuration file” named language.dat. The exact way this file is used depends on the distribution, so please, read the documentation for the latter (note also some distributions generate the file with some tool).

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

### 2.1 Format

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

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

```
% File      : language.dat
% Purpose   : tell iniTeX what files with patterns to load.
english     english.hyphenations
=british

dutch       hyphen.dutch exceptions.dutch % Nederlands
german      hyphen.ger
```

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

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

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:

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

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

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

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

```
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 TeX users, so the files have to be coded so that they can be read by both  $\LaTeX$  and plain 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  $\LaTeX$  option that is to be used. These macros and their functions are discussed below. You must define all or none for a language (or a dialect); defining, say, `\date<lang>` but not `\captions<lang>` does not raise an error but can lead to unexpected results.
- When a language definition file is loaded, it can define `\l@<lang>` to be a dialect of `\language0` when `\l@<lang>` is undefined.
- Language names must be all lowercase. If an unknown language is selected, babel will attempt setting it after lowercasing its name.
- The semantics of modifiers is not defined (on purpose). In most cases, they will just be simple separated options (eg, `spanish`), but a language might require, say, a set of options organized as a tree with suboptions (in such a case, the recommended separator is `/`).

Some recommendations:

- The preferred shorthand is `"`, which is not used in  $\LaTeX$  (quotes are entered as `` `` and `' '`). Other good choices are characters which are not used in a certain context (eg, `=` in an ancient language). Note however `=`, `<`, `>`, `:` and the like can be dangerous, because they may be used as part of the syntax of some elements (numeric expressions, key/value pairs, etc.).
- Captions should not contain shorthands or encoding-dependent commands (the latter is not always possible, but should be clearly documented). They should be defined using the LICR. You may also use the new tools for encoded strings, described below.

- Avoid adding things to `\noextras⟨lang⟩` except for `umlauthigh` and friends, `\bbl@deactivate`, `\bbl@(non)frenchspacing`, and language-specific macros. Use always, if possible, `\bbl@save` and `\bbl@savevariable` (except if you still want to have access to the previous value). Do not reset a macro or a setting to a hardcoded value. Never. Instead save its value in `\extras⟨lang⟩`.
- Do not switch scripts. If you want to make sure a set of glyphs is used, switch either the font encoding (low-level) or the language (high-level, which in turn may switch the font encoding). Usage of things like `\latintext` is deprecated.<sup>27</sup>
- Please, for “private” internal macros do not use the `\bbl@` prefix. It is used by `babel` and it can lead to incompatibilities.

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

### 3.1 Guidelines for contributed languages

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

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

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

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

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

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

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. The macro `\langle lang \rangle 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.)

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

### 3.3 Skeleton

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

```

\ProvidesLanguage{<language>}
[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>

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

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

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

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

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

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

\EndBabelCommands

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

\ldf@finish{<language>}

```

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

```

\AtEndOfPackage{%
  \RequirePackage{dingbat}%      Delay package
  \savebox{\myeye}{\eye}}%      And direct usage
\newsavebox{\myeye}
\newcommand\myanchor{\anchor}%  But OK inside command

```

### 3.4 Support for active characters

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

`\initiate@active@char` The internal macro `\initiate@active@char` is used in language definition files to instruct



TeX to give a character the category code ‘active’. When a character has been made active it will remain that way until the end of the document. Its definition may vary.

`\bbl@activate` 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 TeXbook states: “Plain TeX includes a macro called `\dospecials` that is essentially a set macro, representing the set of all characters that have a special category code.” [4, p. 380]  
`\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`. TeX adds another macro called `\@sanitize` representing the same character set, but without the curly braces. The macros `\bbl@add@special⟨char⟩` and `\bbl@remove@special⟨char⟩` add and remove the character `⟨char⟩` to these two sets.

### 3.5 Support for saving macro definitions

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

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

`\babel@savevariable` A second macro is provided to save the current value of a variable. In this context, 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

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



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

`\bbl@nonfrenchspacing`

### 3.8 Encoding-dependent strings

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

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

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

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

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

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 a encoded way).

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

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

```

\StartBabelCommands{language}{captions}
  [unicode, fontenc=TU EU1 EU2, charset=utf8]
\SetString{chaptername}{utf8-string}

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

\EndBabelCommands

```

A real example is:

```

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

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

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

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

\StartBabelCommands{german,austrian}{date}
  \SetString\monthiiname{Februar}
  \SetString\monthiiname{M\"a\"rz}
  \SetString\monthivname{April}
  \SetString\monthvname{Mai}
  \SetString\monthviname{Juni}
  \SetString\monthviiname{Juli}
  \SetString\monthviiname{August}
  \SetString\monthixname{September}
  \SetString\monthxiname{Oktober}
  \SetString\monthxiiname{November}
  \SetString\monthxiiname{Dezenber}
  \SetString\today{\number\day.-%
    \csmname month\romannumeral\month name\endcsmname\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  $\langle date \rangle \langle language \rangle$  exists).

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

The starred version just forces strings to take a value – if not set as package option, then the default for the engine is used. This is not done by default to prevent backward incompatibilities, but if you are creating a new language this version is better. It's up to the maintainers of the current languages to decide if using it is appropriate.<sup>30</sup>

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

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

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

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

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

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

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

**\SetStringLoop**  $\{\langle macro-name \rangle\}\{\langle string-list \rangle\}$

A convenient way to define several ordered names at once. For example, to define `\abmoniname`, `\abmoniname`, etc. (and similarly with `abday`):

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

#1 is replaced by the roman numeral.

**\SetCase**  $[\langle map-list \rangle]\{\langle toupper-code \rangle\}\{\langle tolower-code \rangle\}$

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

**New 3.9g** Case mapping serves in  $\text{\TeX}$  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.

### 3.9 Executing code based on the selector

`\IfBabelSelectorTF` {⟨selectors⟩}{⟨true⟩}{⟨false⟩}

**New 3.67** Sometimes a different setup is desired depending on the selector used. Values allowed in `⟨selectors⟩` are `select`, `other`, `foreign`, `other*` (and also `foreign*` for the tentative starred version), and it can consist of a comma-separated list. For example:

```
\IfBabelSelectorTF{other, other*}{A}{B}
```

is true with these two environment selectors.

Its natural place of use is in hooks or in `\extras⟨language⟩`.

## 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>).

## 4 Identification and loading of required files

*Code documentation is still under revision.*

**The following description is no longer valid, because `switch` and `plain` have been merged into `babel.def`.**

The `babel` package after unpacking consists of the following files:

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

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

**babel.sty** is the  $\LaTeX$  package, which set options and load language styles.

**plain.def** defines some  $\text{\LaTeX}$  macros required by `babel.def` and provides a few tools for Plain.  
**hyphen.cfg** is the file to be used when generating the formats to load hyphenation patterns.

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

## 5 locale directory

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

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

This is a preliminary documentation.

ini files contain the actual data; tex files are currently just proxies to the corresponding ini files. Most keys are self-explanatory.

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

**version** of the ini file

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

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

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

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

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

Keys may be further qualified in a particular language with a suffix starting with a uppercase letter. It can be just a letter (eg, babel.name.A, babel.name.B) or a name (eg, date.long.Nominative, date.long.Formal, but no language is currently using the latter). *Multi-letter* qualifiers are forward compatible in the sense they won’t conflict with new “global” keys (which start always with a lowercase case). There is an exception, however: the section counters has been devised to have arbitrary keys, so you can add lowercased keys if you want.

## 6 Tools

```
1 <<version=3.75.2737>>
2 <<date=2022/05/14>>
```

**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@c1#1{\csname bbl@#1@languagenam\endcsname}
14 \def\bbl@loop#1#2#3{\bbl@loop#1{#3}#2,\@nnil,}
15 \def\bbl@loopx#1#2{\expandafter\bbl@loop\expandafter#1\expandafter{#2}}
```

```

16 \def\bbl@loop#1#2#3,{%
17   \ifx\@nnil#3\relax\else
18     \def#1{#3}#2\bbl@afterfi\bbl@loop#1{#2}%
19   \fi}
20 \def\bbl@for#1#2#3{\bbl@loopx#1{#2}{\ifx#1\@empty\else#3\fi}}

```

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

```

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

```

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

```

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

```

**\bbl@exp** Now, just syntactical sugar, but it makes partial expansion of some code a lot more simple and readable. Here `\` stands for `\noexpand`, `\<.>` for `\noexpand` applied to a built macro name (which does not define the macro if undefined to `\relax`, because it is created locally), and `\[...]` for one-level expansion (where `...` is the macro name without the backslash). The result may be followed by extra arguments, if necessary.

```

29 \def\bbl@exp#1{%
30   \begingroup
31     \let\<\noexpand
32     \let\<\bbl@exp@en
33     \let\[\bbl@exp@ue
34     \edef\bbl@exp@aux{\endgroup#1}%
35   \bbl@exp@aux}
36 \def\bbl@exp@en#1>{\expandafter\noexpand\csname#1\endcsname}%
37 \def\bbl@exp@ue#1]{%
38   \unexpanded\expandafter\expandafter\expandafter{\csname#1\endcsname}}%

```

**\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.

```

39 \def\bbl@tempa#1{%
40   \long\def\bbl@trim##1##2{%
41     \futurelet\bbl@trim@a\bbl@trim@c##2\@nil\@nil#1\@nil\relax{##1}}%
42   \def\bbl@trim@c{%
43     \ifx\bbl@trim@a\@sptoken
44       \expandafter\bbl@trim@b
45     \else
46       \expandafter\bbl@trim@b\expandafter#1%
47     \fi}%
48   \long\def\bbl@trim@b#1##1 \@nil{\bbl@trim@i#1}}
49 \bbl@tempa{ }
50 \long\def\bbl@trim@i#1\@nil#2\relax#3{#3{#1}}
51 \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 does not waste memory.

```

52 \begingroup
53   \gdef\bbl@ifunset#1{%

```

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

```

54 \expandafter\ifx\csname#1\endcsname\relax
55 \expandafter\@firstoftwo
56 \else
57 \expandafter\@secondoftwo
58 \fi}
59 \bbl@ifunset{ifcsname}% TODO. A better test?
60 {}%
61 {\gdef\bbl@ifunset#1{%
62 \ifcsname#1\endcsname
63 \expandafter\ifx\csname#1\endcsname\relax
64 \bbl@afterelse\expandafter\@firstoftwo
65 \else
66 \bbl@afterfi\expandafter\@secondoftwo
67 \fi
68 \else
69 \expandafter\@firstoftwo
70 \fi}}
71 \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,

```

72 \def\bbl@ifblank#1{%
73 \bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
74 \long\def\bbl@ifblank@i#1#2\@nil#3#4#5\@nil{#4}
75 \def\bbl@ifset#1#2#3{%
76 \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).

```

77 \def\bbl@forkv#1#2{%
78 \def\bbl@kvcmd##1##2##3{#2}%
79 \bbl@kvnext#1,\@nil,}
80 \def\bbl@kvnext#1,{%
81 \ifx\@nil#1\relax\else
82 \bbl@ifblank{#1}{\bbl@forkv@eq#1=\@empty=\@nil{#1}}%
83 \expandafter\bbl@kvnext
84 \fi}
85 \def\bbl@forkv@eq#1=#2=#3\@nil#4{%
86 \bbl@trim\def\bbl@forkv@a{#1}%
87 \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).

```

88 \def\bbl@vforeach#1#2{%
89 \def\bbl@forcmd##1{#2}%
90 \bbl@fornext#1,\@nil,}
91 \def\bbl@fornext#1,{%
92 \ifx\@nil#1\relax\else
93 \bbl@ifblank{#1}{\bbl@trim\bbl@forcmd{#1}}%
94 \expandafter\bbl@fornext
95 \fi}
96 \def\bbl@foreach#1{\expandafter\bbl@vforeach\expandafter{#1}}

```

**\bbl@replace** Returns implicitly \toks@ with the modified string.

```

97 \def\bbl@replace#1#2#3{% in #1 -> repl #2 by #3
98 \toks@{}}
99 \def\bbl@replace@aux##1#2##2#2{%
100 \ifx\bbl@nil##2%
101 \toks@\expandafter{\the\toks@##1}%
102 \else
103 \toks@\expandafter{\the\toks@##1#3}%
104 \bbl@afterfi

```

```

105 \bbl@replace@aux##2#2%
106 \fi}%
107 \expandafter\bbl@replace@aux#1#2\bbl@nil#2%
108 \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).

```

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

```

Two further tools. \bbl@ifsamestring 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 pdfTeX, 1 is luatex, and 2 is xetex. You may use the latter it in your language style if you want.

```

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

```



```
157 \fi
```

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

```
158 \def\bbl@bsphack{%
159   \ifhmode
160     \hskip\z@skip
161     \def\bbl@esphack{\loop\ifdim\lastskip>\z@\unskip\repeat\unskip}%
162   \else
163     \let\bbl@esphack\empty
164   \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`.

```
165 \def\bbl@cased{%
166   \ifx\oe\OE
167     \expandafter\in@\expandafter
168       {\expandafter\OE\expandafter}\expandafter{\oe}%
169   \ifin@
170     \bbl@afterelse\expandafter\MakeUppercase
171   \else
172     \bbl@afterfi\expandafter\MakeLowercase
173   \fi
174 \else
175   \expandafter\@firstofone
176 \fi}
```

An alternative to `\IfFormatAtLeastTF` for old versions. Temporary.

```
177 \ifx\IfFormatAtLeastTF\@undefined
178   \def\bbl@ifformatlater{\@ifl@t@r\fmtversion}
179 \else
180   \let\bbl@ifformatlater\IfFormatAtLeastTF
181 \fi
```

The following adds some code to `\extras...` both before and after, while avoiding doing it twice. It's somewhat convoluted, to deal with `#`'s. Used to deal with `alph`, `Alph` and `frenchspacing` when there are already changes (with `\babel@save`).

```
182 \def\bbl@extras@wrap#1#2#3{% 1:in-test, 2:before, 3:after
183   \toks@\expandafter\expandafter\expandafter{%
184     \csname extras\language\endcsname}%
185   \bbl@exp{\in@{#1}}{\the\toks@}%
186   \ifin@
187     \@temptokena{#2}%
188     \edef\bbl@tempc{\the\@temptokena\the\toks@}%
189     \toks@\expandafter{\bbl@tempc#3}%
190     \expandafter\edef\csname extras\language\endcsname{\the\toks@}%
191   \fi}
192 <</Basic macros>>
```

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

```
193 <<(*Make sure ProvidesFile is defined)>> \equiv
194 \ifx\ProvidesFile\@undefined
195   \def\ProvidesFile#1[#2 #3 #4]{%
196     \wlog{File: #1 #4 #3 <#2>}%
197     \let\ProvidesFile\@undefined}
198 \fi
199 <</Make sure ProvidesFile is defined>>
```

## 6.1 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.

```

200 <<*Define core switching macros>> ≡
201 \ifx\language\@undefined
202   \csname newcount\endcsname\language
203 \fi
204 <</Define core switching macros>>

```

`\last@language` Another counter is used to keep track of the allocated languages.  $\TeX$  and  $\LaTeX$  reserves for this purpose the count 19.

`\addlanguage` This macro was introduced for  $\TeX < 2$ . Preserved for compatibility.

```

205 <<*Define core switching macros>> ≡
206 \countdef\last@language=19
207 \def\addlanguage{\csname newlanguage\endcsname}
208 <</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. 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.

## 6.2 The Package File ( $\LaTeX$ , `babel.sty`)

```

209 <*package>
210 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
211 \ProvidesPackage{babel}[<<date>> <<version>> The Babel package]

```

Start with some “private” debugging tool, and then define macros for errors.

```

212 \@ifpackagewith{babel}{debug}
213   {\providecommand\bbl@trace[1]{\message{^^J[ #1 ]}}%
214    \let\bbl@debug\@firstofone
215    \ifx\directlua\@undefined\else
216      \directlua{ Babel = Babel or {}
217        Babel.debug = true }%
218      \input{babel-debug.tex}%
219    \fi}
220 {\providecommand\bbl@trace[1]{}%
221  \let\bbl@debug\@gobble
222  \ifx\directlua\@undefined\else
223    \directlua{ Babel = Babel or {}
224      Babel.debug = false }%
225  \fi}
226 \def\bbl@error#1#2{%
227   \begingroup
228     \def\{\MessageBreak}%
229     \PackageError{babel}{#1}{#2}%
230   \endgroup}
231 \def\bbl@warning#1{%
232   \begingroup
233     \def\{\MessageBreak}%
234     \PackageWarning{babel}{#1}%
235   \endgroup}
236 \def\bbl@infowarn#1{%
237   \begingroup
238     \def\{\MessageBreak}%
239     \GenericWarning
240       {(babel) \spaces\spaces\spaces}%
241       {Package babel Info: #1}%
242   \endgroup}
243 \def\bbl@info#1{%
244   \begingroup

```

```

245 \def\{MessageBreak}%
246 \PackageInfo{babel}{#1}%
247 \endgroup}

```

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.

But first, include here the *Basic macros* defined above.

```

248 <<Basic macros>>
249 \ifpackagewith{babel}{silent}
250 {\let\bbl@info@gobble
251 \let\bbl@infowarn@gobble
252 \let\bbl@warning@gobble}
253 {}
254 %
255 \def\AfterBabelLanguage#1{%
256 \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.

```

257 \ifx\bbl@languages\@undefined\else
258 \begingroup
259 \catcode\^^I=12
260 \ifpackagewith{babel}{showlanguages}{%
261 \begingroup
262 \def\bbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}%
263 \wlog{<*languages>}%
264 \bbl@languages
265 \wlog{</languages>}%
266 \endgroup}{%
267 \endgroup
268 \def\bbl@elt#1#2#3#4{%
269 \ifnum#2=\z@
270 \gdef\bbl@nulllanguage{#1}%
271 \def\bbl@elt##1##2##3##4{}}%
272 \fi}%
273 \bbl@languages
274 \fi%

```

## 6.3 base

The first 'real' option to be processed is base, which sets the hyphenation patterns then resets ver@babel.sty so that L<sup>A</sup>T<sub>E</sub>X 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.

```

275 \bbl@trace{Defining option 'base'}
276 \ifpackagewith{babel}{base}{%
277 \let\bbl@onlyswitch\@empty
278 \let\bbl@provide@locale\relax
279 \input babel.def
280 \let\bbl@onlyswitch\@undefined
281 \ifx\directlua\@undefined
282 \DeclareOption*{\bbl@patterns{\CurrentOption}}%
283 \else
284 \input luababel.def
285 \DeclareOption*{\bbl@patterns@lua{\CurrentOption}}%
286 \fi
287 \DeclareOption{base}{}%
288 \DeclareOption{showlanguages}{}%
289 \ProcessOptions
290 \global\expandafter\let\csname opt@babel.sty\endcsname\relax
291 \global\expandafter\let\csname ver@babel.sty\endcsname\relax

```

```

292 \global\let@ifl@ter@@\@ifl@ter
293 \def@ifl@ter#1#2#3#4#5{\global\let@ifl@ter\@ifl@ter@@}%
294 \endinput}{}}%

```

## 6.4 key=value options and other general option

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

```

295 \bbl@trace{key=value and another general options}
296 \bbl@csarg\let{tempa\expandafter}\csname opt@babel.sty\endcsname
297 \def\bbl@tempb#1.#2{% Remove trailing dot
298   #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%
299 \def\bbl@tempd#1.#2\@nnil{% TODO. Refactor lists?
300   \ifx\@empty#2%
301     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
302   \else
303     \in@{,provide=}{, #1}%
304     \ifin@
305       \edef\bbl@tempc{%
306         \ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.\bbl@tempb#2}%
307     \else
308       \in@{=}{ #1}%
309       \ifin@
310         \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
311       \else
312         \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
313         \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
314       \fi
315     \fi
316   \fi}
317 \let\bbl@tempc\@empty
318 \bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty\@nnil}
319 \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.

```

320 \DeclareOption{KeepShorthandsActive}{}
321 \DeclareOption{activeacute}{}
322 \DeclareOption{activegrave}{}
323 \DeclareOption{debug}{}
324 \DeclareOption{noconfigs}{}
325 \DeclareOption{showlanguages}{}
326 \DeclareOption{silent}{}
327 % \DeclareOption{mono}{}
328 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
329 \chardef\bbl@iniflag\z@
330 \DeclareOption{provide=*}{\chardef\bbl@iniflag\@ne} % main -> +1
331 \DeclareOption{provide+=*}{\chardef\bbl@iniflag\tw@} % add = 2
332 \DeclareOption{provide*=*}{\chardef\bbl@iniflag\thr@@} % add + main
333 % A separate option
334 \let\bbl@autoload@options\@empty
335 \DeclareOption{provide@=*}{\def\bbl@autoload@options{import}}
336 % Don't use. Experimental. TODO.
337 \newif\ifbbl@single
338 \DeclareOption{selectors=off}{\bbl@singletrue}
339 <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.

```
340 \let\bbl@opt@shorthands\@nnil
341 \let\bbl@opt@config\@nnil
342 \let\bbl@opt@main\@nnil
343 \let\bbl@opt@headfoot\@nnil
344 \let\bbl@opt@layout\@nnil
345 \let\bbl@opt@provide\@nnil
```

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

```
346 \def\bbl@tempa#1=#2\bbl@tempa{%
347   \bbl@csarg\ifx{opt@#1}\@nnil
348     \bbl@csarg\edef{opt@#1}{#2}%
349   \else
350     \bbl@error
351     {Bad option '#1=#2'. Either you have misspelled the\\%
352       key or there is a previous setting of '#1'. Valid\\%
353       keys are, among others, 'shorthands', 'main', 'bidi',\\%
354       'strings', 'config', 'headfoot', 'safe', 'math'.}%
355     {See the manual for further details.}
356   \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.

```
357 \let\bbl@language@opts\@empty
358 \DeclareOption*{%
359   \bbl@xin@{\string=}{\CurrentOption}%
360   \ifin@
361     \expandafter\bbl@tempa\CurrentOption\bbl@tempa
362   \else
363     \bbl@add@list\bbl@language@opts{\CurrentOption}%
364   \fi}
```

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

```
365 \ProcessOptions*
366 \ifx\bbl@opt@provide\@nnil
367   \let\bbl@opt@provide\@empty % %%% MOVE above
368 \else
369   \chardef\bbl@iniflag\@ne
370   \bbl@exp{\bbl@forkv{\@nameuse{@raw@opt@babel.sty}}}{%
371     \in@{,provide,}{, #1,}%
372     \ifin@
373       \def\bbl@opt@provide{#2}%
374       \bbl@replace\bbl@opt@provide{;}{,}%
375     \fi}
376 \fi
377 %
```

## 6.5 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=...

```
378 \bbl@trace{Conditional loading of shorthands}
379 \def\bbl@sh@string#1{%
380   \ifx#1\@empty\else
381     \ifx#1t\string~%
382     \else\ifx#1c\string,%
383     \else\string#1%
384   \fi\fi
```

```

385 \expandafter\bb1@sh@string
386 \fi}
387 \ifx\bb1@opt@shorthands\@nnil
388 \def\bb1@ifshorthand#1#2#3{#2}%
389 \else\ifx\bb1@opt@shorthands\@empty
390 \def\bb1@ifshorthand#1#2#3{#3}%
391 \else

```

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

```

392 \def\bb1@ifshorthand#1{%
393 \bb1@xin@{\string#1}{\bb1@opt@shorthands}%
394 \ifin@
395 \expandafter\@firstoftwo
396 \else
397 \expandafter\@secondoftwo
398 \fi}

```

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

```

399 \edef\bb1@opt@shorthands{%
400 \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.

```

401 \bb1@ifshorthand{'}%
402 {\PassOptionsToPackage{activeacute}{babel}}{}
403 \bb1@ifshorthand{`}%
404 {\PassOptionsToPackage{activegrave}{babel}}{}
405 \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.

```

406 \ifx\bb1@opt@headfoot\@nnil\else
407 \g@addto@macro\@resetactivechars{%
408 \set@typeset@protect
409 \expandafter\select@language@x\expandafter{\bb1@opt@headfoot}%
410 \let\protect\noexpand}
411 \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.

```

412 \ifx\bb1@opt@safe\@undefined
413 \def\bb1@opt@safe{BR}
414 \fi

```

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

```

415 \bb1@trace{Defining IfBabelLayout}
416 \ifx\bb1@opt@layout\@nnil
417 \newcommand\IfBabelLayout[3]{#3}%
418 \else
419 \newcommand\IfBabelLayout[1]{%
420 \@expandtwoargs\in@{.#1.}{.\bb1@opt@layout.}%
421 \ifin@
422 \expandafter\@firstoftwo
423 \else
424 \expandafter\@secondoftwo
425 \fi}
426 \fi
427 </package>
428 <*core>

```

## 6.6 Interlude for Plain

Because of the way docstrip works, we need to insert some code for Plain here. However, the tools provided by the babel installer for literate programming makes this section a short interlude, because the actual code is below, tagged as *Emulate LaTeX*.

```
429 \ifx\ldf@quit\undefined\else
430 \endinput\fi % Same line!
431 <<Make sure ProvidesFile is defined>>
432 \ProvidesFile{babel.def}[\<date>] \<version> Babel common definitions]
433 \ifx\AtBeginDocument\undefined % TODO. change test.
434 <<Emulate LaTeX>>
435 \fi
```

That is all for the moment. Now follows some common stuff, for both Plain and  $\text{\LaTeX}$ . After it, we will resume the  $\text{\LaTeX}$ -only stuff.

```
436 </core>
437 <*package | core>
```

## 7 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.

```
438 \def\bbl@version{\<version>}
439 \def\bbl@date{\<date>}
440 <<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.

```
441 \def\adddialect#1#2{%
442   \global\chardef#1#2\relax
443   \bbl@usehooks{adddialect}{\{#1\}\{#2\}}%
444   \begingroup
445     \count@#1\relax
446     \def\bbl@elt###1##2###3###4{%
447       \ifnum\count@=##2\relax
448         \edef\bbl@tempa{\expandafter\@gobbletwo\string#1}%
449         \bbl@info{Hyphen rules for '\expandafter\@gobble\bbl@tempa'
450               set to \expandafter\string\csname l@##1\endcsname\%
451               (\string\language\the\count@). Reported}%
452         \def\bbl@elt####1####2####3####4{%
453           \fi}%
454         \bbl@cs{languages}%
455       \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 an attempt 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.

```
456 \def\bbl@fixname#1{%
457   \begingroup
458     \def\bbl@tempe{l@}%
459     \edef\bbl@tempd{\noexpand\ifundefined{\noexpand\bbl@tempe#1}}%
460     \bbl@tempd
461     {\lowercase\expandafter{\bbl@tempd}%
462     {\uppercase\expandafter{\bbl@tempd}%
463     \empty
464     {\edef\bbl@tempd{\def\noexpand#1{#1}}%
465     \uppercase\expandafter{\bbl@tempd}}}%
466     {\edef\bbl@tempd{\def\noexpand#1{#1}}%
```

```

467 \lowercase\expandafter{\bbl@tempd}}}%
468 \empty
469 \edef\bbl@tempd{\endgroup\def\noexpand#1{#1}}%
470 \bbl@tempd
471 \bbl@exp{\bbl@usehooks{language}{\language}{#1}}}%
472 \def\bbl@iflanguage#1{%
473 \ifundefined{1@#1}{\@nolanerr{#1}\gobble}\@firstofone}

```

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

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

```

474 \def\bbl@bcpcase#1#2#3#4\@#5{%
475 \ifx\empty#3%
476 \uppercase{\def#5{#1#2}}%
477 \else
478 \uppercase{\def#5{#1}}%
479 \lowercase{\edef#5{#5#2#3#4}}%
480 \fi}
481 \def\bbl@bcpllookup#1-#2-#3-#4\@{%
482 \let\bbl@bcp\relax
483 \lowercase{\def\bbl@tempa{#1}}%
484 \ifx\empty#2%
485 \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
486 \else\ifx\empty#3%
487 \bbl@bcpcase#2\empty\empty\@{\bbl@tempb
488 \IfFileExists{babel-\bbl@tempa-\bbl@tempb.ini}%
489 {\edef\bbl@bcp{\bbl@tempa-\bbl@tempb}}%
490 }%
491 \ifx\bbl@bcp\relax
492 \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
493 \fi
494 \else
495 \bbl@bcpcase#2\empty\empty\@{\bbl@tempb
496 \bbl@bcpcase#3\empty\empty\@{\bbl@tempc
497 \IfFileExists{babel-\bbl@tempa-\bbl@tempb-\bbl@tempc.ini}%
498 {\edef\bbl@bcp{\bbl@tempa-\bbl@tempb-\bbl@tempc}}%
499 }%
500 \ifx\bbl@bcp\relax
501 \IfFileExists{babel-\bbl@tempa-\bbl@tempc.ini}%
502 {\edef\bbl@bcp{\bbl@tempa-\bbl@tempc}}%
503 }%
504 \fi
505 \ifx\bbl@bcp\relax
506 \IfFileExists{babel-\bbl@tempa-\bbl@tempc.ini}%
507 {\edef\bbl@bcp{\bbl@tempa-\bbl@tempc}}%
508 }%
509 \fi
510 \ifx\bbl@bcp\relax
511 \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
512 \fi
513 \fi\fi}
514 \let\bbl@initoload\relax
515 \def\bbl@provide@locale{%
516 \ifx\babelprovide\undefined
517 \bbl@error{For a language to be defined on the fly 'base'\\%
518 is not enough, and the whole package must be\\%
519 loaded. Either delete the 'base' option or\\%
520 request the languages explicitly}%
521 {See the manual for further details.}%
522 \fi
523 % TODO. Option to search if loaded, with \LocaleForEach

```



```

524 \let\bbl@auxname\language % Still necessary. TODO
525 \bbl@ifunset{\bbl@bcp@map@\language}\{}% Move uplevel??
526 \edef\language{\@nameuse{\bbl@bcp@map@\language}}}%
527 \ifbbl@bcpallowed
528 \expandafter\ifx\csname date\language\endcsname\relax
529 \expandafter
530 \bbl@bcplookup\language-\@empty-\@empty-\@empty\@
531 \ifx\bbl@bcp\relax\else % Returned by \bbl@bcplookup
532 \edef\language{\bbl@bcp@prefix\bbl@bcp}%
533 \edef\localename{\bbl@bcp@prefix\bbl@bcp}%
534 \expandafter\ifx\csname date\language\endcsname\relax
535 \let\bbl@initoload\bbl@bcp
536 \bbl@exp{\bbl@babelprovide[\bbl@autoload@bcptoptions]{\language}}%
537 \let\bbl@initoload\relax
538 \fi
539 \bbl@csarg\xdef{bcp@map@\bbl@bcp}{\localename}%
540 \fi
541 \fi
542 \fi
543 \expandafter\ifx\csname date\language\endcsname\relax
544 \IfFileExists{babel-\language.tex}%
545 {\bbl@exp{\bbl@babelprovide[\bbl@autoload@options]{\language}}}%
546 {}%
547 \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.

```

548 \def\iflanguage#1{%
549 \bbl@iflanguage{#1}{%
550 \ifnum\csname l@#1\endcsname=\language
551 \expandafter\@firstoftwo
552 \else
553 \expandafter\@secondoftwo
554 \fi}}

```

## 7.1 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.

```

555 \let\bbl@select@type\z@
556 \edef\selectlanguage{%
557 \noexpand\protect
558 \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`.

```

559 \ifx\undefined\protect\let\protect\relax\fi

```

The following definition is preserved for backwards compatibility (eg, arabi, koma). It is related to a trick for 2.09, now discarded.

```

560 \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.

```
561 \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`

```
562 \def\bbl@push@language{%
563   \ifx\language\@undefined\else
564     \ifx\currentgrouplevel\@undefined
565       \xdef\bbl@language@stack{\language+\bbl@language@stack}%
566     \else
567       \ifnum\currentgrouplevel=\z@
568         \xdef\bbl@language@stack{\language+}%
569       \else
570         \xdef\bbl@language@stack{\language+\bbl@language@stack}%
571       \fi
572     \fi
573 }
```

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`.

```
574 \def\bbl@pop@lang#1+#2\@{%
575   \edef\language{#1}%
576   \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).

```
577 \let\bbl@ifrestoring\@secondoftwo
578 \def\bbl@pop@language{%
579   \expandafter\bbl@pop@lang\bbl@language@stack\@
580   \let\bbl@ifrestoring\@firstoftwo
581   \expandafter\bbl@set@language\expandafter{\language}%
582   \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).

```
583 \chardef\localeid\z@
584 \def\bbl@id@last{0} % No real need for a new counter
585 \def\bbl@id@assign{%
586   \bbl@ifunset\bbl@id@\@language}%
587   {\count\bbl@id@last\relax
588   \advance\count\bbl@id@last\@ne
589   \bbl@csarg\chardef{id@\@language}\count\bbl@id@last}
590   \edef\bbl@id@last{\the\count\bbl@id@last}%
591   \ifcase\bbl@engine\or
592     \directlua{
593       Babel = Babel or {}
594       Babel.locale_props = Babel.locale_props or {}
595       Babel.locale_props[\bbl@id@last] = {}}
```

```

596         Babel.locale_props[\bbl@id@last].name = '\language'
597     }%
598     \fi}%
599     {}%
600     \chardef\localeid\bbl@cl{id@}}

```

The unprotected part of \selectlanguage.

```

601 \expandafter\def\csname selectlanguage \endcsname#1{%
602   \ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel\tw@\fi
603   \bbl@push@language
604   \aftergroup\bbl@pop@language
605   \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.

`\bbl@savelastskip` is used to deal with skips before the write whatsit (as suggested by U Fischer).

Adapted from hyperref, but it might fail, so I'll consider it a temporary hack, while I study other options (the ideal, but very likely unfeasible except perhaps in luatex, is to avoid the \write altogether when not needed).

```

606 \def\BabelContentsFiles{toc,lof,lot}
607 \def\bbl@set@language#1{% from selectlanguage, pop@
608   % The old buggy way. Preserved for compatibility.
609   \edef\language{%
610     \ifnum\escapechar=\expandafter`\string#1\@empty
611       \else\string#1\@empty\fi}%
612   \ifcat\relax\noexpand#1%
613     \expandafter\ifx\csname date\language\endcsname\relax
614       \edef\language{#1}%
615       \let\localename\language
616     \else
617       \bbl@info{Using '\string\language' instead of 'language' is\\%
618         deprecated. If what you want is to use a\\%
619         macro containing the actual locale, make\\%
620         sure it does not not match any language.\\%
621         Reported}%
622       \ifx\scantokens\@undefined
623         \def\localename{??}%
624       \else
625         \scantokens\expandafter{\expandafter
626           \def\expandafter\localename\expandafter{\language}}%
627       \fi
628     \fi
629   \else
630     \def\localename{#1}% This one has the correct catcodes
631   \fi
632   \select@language{\language}%
633   % write to auxs
634   \expandafter\ifx\csname date\language\endcsname\relax\else
635     \if@filesw
636       \ifx\babel@aux\@gobbletwo\else % Set if single in the first, redundant
637         \bbl@savelastskip
638         \protected@write\@auxout{}\string\babel@aux{\bbl@auxname}{}%
639         \bbl@restorelastskip
640       \fi
641       \bbl@usehooks{write}{}%
642     \fi
643   \fi}

```

```

644 %
645 \let\bbl@restorelastskip\relax
646 \let\bbl@savelastskip\relax
647 %
648 \newif\ifbbl@bcpallowed
649 \bbl@bcpallowedfalse
650 \def\select@language#1{% from set@, babel@aux
651   \ifx\bbl@selectorname\@empty
652     \def\bbl@selectorname{select}%
653   % set hymap
654   \fi
655   \ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi
656   % set name
657   \edef\language#1}%
658   \bbl@fixname\language
659   % TODO. name@map must be here?
660   \bbl@provide@locale
661   \bbl@iflanguage\language{%
662     \expandafter\ifx\csname date\language\endcsname\relax
663       \bbl@error
664       {Unknown language '\language'. Either you have\\%
665        misspelled its name, it has not been installed,\\%
666        or you requested it in a previous run. Fix its name,\\%
667        install it or just rerun the file, respectively. In\\%
668        some cases, you may need to remove the aux file}%
669       {You may proceed, but expect wrong results}%
670     \else
671       % set type
672       \let\bbl@select@type\z@
673       \expandafter\bbl@switch\expandafter{\language}%
674     \fi}}
675 \def\babel@aux#1#2{%
676   \select@language{#1}%
677   \bbl@foreach\BabelContentsFiles{% \relax -> don't assume vertical mode
678     \@writefile{##1}{\babel@toc{#1}{#2}\relax}}}% TODO - plain?
679 \def\babel@toc#1#2{%
680   \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.

```

681 \newif\ifbbl@usedategroup
682 \def\bbl@switch#1{% from select@, foreign@
683   % make sure there is info for the language if so requested
684   \bbl@ensureinfo{#1}%
685   % restore
686   \originalTeX
687   \expandafter\def\expandafter\originalTeX\expandafter{%
688     \csname noextras#1\endcsname
689     \let\originalTeX\@empty
690     \babel@beginsave}%
691   \bbl@usehooks{afterreset}}}%
692   \languageshorthands{none}%
693   % set the locale id

```

```

694 \bbl@id@assign
695 % switch captions, date
696 % No text is supposed to be added here, so we remove any
697 % spurious spaces.
698 \bbl@bsphack
699 \ifcase\bbl@select@type
700 \csname captions#1\endcsname\relax
701 \csname date#1\endcsname\relax
702 \else
703 \bbl@xin@{,captions,}{, \bbl@select@opts,}%
704 \ifin@
705 \csname captions#1\endcsname\relax
706 \fi
707 \bbl@xin@{,date,}{, \bbl@select@opts,}%
708 \ifin@ % if \foreign... within \<lang>date
709 \csname date#1\endcsname\relax
710 \fi
711 \fi
712 \bbl@esphack
713 % switch extras
714 \bbl@usehooks{beforeextras}{}%
715 \csname extras#1\endcsname\relax
716 \bbl@usehooks{afterextras}{}%
717 % > babel-ensure
718 % > babel-sh-<short>
719 % > babel-bidi
720 % > babel-fontspec
721 % hyphenation - case mapping
722 \ifcase\bbl@opt@hyphenmap\or
723 \def\BabelLower##1##2{\lccode##1=##2\relax}%
724 \ifnum\bbl@hymapsel>4\else
725 \csname\language\name @bbl@hyphenmap\endcsname
726 \fi
727 \chardef\bbl@opt@hyphenmap\z@
728 \else
729 \ifnum\bbl@hymapsel>\bbl@opt@hyphenmap\else
730 \csname\language\name @bbl@hyphenmap\endcsname
731 \fi
732 \fi
733 \let\bbl@hymapsel\@cclv
734 % hyphenation - select rules
735 \ifnum\csname l@\language\endcsname=\l@unhyphenated
736 \edef\bbl@tempa{u}%
737 \else
738 \edef\bbl@tempa{\bbl@c1{lnbrk}}%
739 \fi
740 % linebreaking - handle u, e, k (v in the future)
741 \bbl@xin@{/u}{/\bbl@tempa}%
742 \ifin@ \else \bbl@xin@{/e}{/\bbl@tempa} \fi % elongated forms
743 \ifin@ \else \bbl@xin@{/k}{/\bbl@tempa} \fi % only kashida
744 \ifin@ \else \bbl@xin@{/v}{/\bbl@tempa} \fi % variable font
745 \ifin@
746 % unhyphenated/kashida/elongated = allow stretching
747 \language\l@unhyphenated
748 \babel@savevariable\emergencystretch
749 \emergencystretch\maxdimen
750 \babel@savevariable\hbadness
751 \hbadness\@M
752 \else
753 % other = select patterns
754 \bbl@patterns{#1}%
755 \fi
756 % hyphenation - mins

```

```

757 \babel@savevariable\lefthyphenmin
758 \babel@savevariable\righthyphenmin
759 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
760   \set@hyphenmins\tw@thr@@\relax
761 \else
762   \expandafter\expandafter\expandafter\set@hyphenmins
763     \csname #1hyphenmins\endcsname\relax
764 \fi
765 \let\bbl@selectorname\@empty}

```

`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.

```

766 \long\def\otherlanguage#1{%
767   \def\bbl@selectorname{other}%
768   \ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel\thr@@\fi
769   \csname selectlanguage\endcsname{#1}%
770   \ignorespaces}

```

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

```

771 \long\def\endotherlanguage{%
772   \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`.

```

773 \expandafter\def\csname otherlanguage*\endcsname{%
774   \@ifnextchar[\bbl@otherlanguage@s{\bbl@otherlanguage@s[]}}
775 \def\bbl@otherlanguage@s[#1]#2{%
776   \def\bbl@selectorname{other*}%
777   \ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi
778   \def\bbl@select@opts{#1}%
779   \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”.

```

780 \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`.

```

781 \providecommand\bbl@beforeforeign{}
782 \edef\foreignlanguage{%
783   \noexpand\protect
784   \expandafter\noexpand\csname foreignlanguage \endcsname}
785 \expandafter\def\csname foreignlanguage \endcsname{%
786   \@ifstar\bbl@foreign@s\bbl@foreign@x}
787 \providecommand\bbl@foreign@x[3][]{%
788   \begingroup
789     \def\bbl@selectorname{foreign}%
790     \def\bbl@select@opts{#1}%
791     \let\BabelText\@firstofone
792     \bbl@beforeforeign
793     \foreign@language{#2}%
794     \bbl@usehooks{foreign}{}%
795     \BabelText{#3}% Now in horizontal mode!
796   \endgroup}
797 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \setpar, ?\@par
798   \begingroup
799     {\par}%
800     \def\bbl@selectorname{foreign*}%
801     \let\bbl@select@opts\empty
802     \let\BabelText\@firstofone
803     \foreign@language{#1}%
804     \bbl@usehooks{foreign*}{}%
805     \bbl@dirparastext
806     \BabelText{#2}% Still in vertical mode!
807     {\par}%
808   \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`.

```

809 \def\foreign@language#1{%
810   % set name
811   \edef\language#1%
812   \ifbbl@usedategroup
813     \bbl@add\bbl@select@opts{,date,}%
814     \bbl@usedategroupfalse
815   \fi
816   \bbl@fixname\language
817   % TODO. name@map here?
818   \bbl@provide@locale
819   \bbl@iflanguage\language%
820     \expandafter\ifx\csname date\language\endcsname\relax
821       \bbl@warning   % TODO - why a warning, not an error?
822       {Unknown language '#1'. Either you have\\%
823        misspelled its name, it has not been installed,\\%
824        or you requested it in a previous run. Fix its name,\\%
825        install it or just rerun the file, respectively. In\\%
826        some cases, you may need to remove the aux file.\\%
827        I'll proceed, but expect wrong results.\\%
828        Reported}%
829     \fi
830     % set type
831     \let\bbl@select@type\@ne
832     \expandafter\bbl@switch\expandafter{\language}}

```

The following macro executes conditionally some code based on the selector being used.

```

833 \def\IfBabelSelectorTF#1{%
834   \bbl@xin@{\bbl@selectorname,}{,\zap@space#1 \empty,}%
835   \ifin@
836     \expandafter\@firstoftwo
837   \else

```

```

838 \expandafter\@secondoftwo
839 \fi}

```

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

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

```

840 \let\bbl@hyphlist\@empty
841 \let\bbl@hyphenation@\relax
842 \let\bbl@pttnlist\@empty
843 \let\bbl@patterns@\relax
844 \let\bbl@hymapsel=\@cclv
845 \def\bbl@patterns#1{%
846   \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
847     \csname l@#1\endcsname
848     \edef\bbl@tempa{#1}%
849   \else
850     \csname l@#1:\f@encoding\endcsname
851     \edef\bbl@tempa{#1:\f@encoding}%
852   \fi
853   \@expandtwoargs\bbl@usehooks{patterns}{#1}{\bbl@tempa}}%
854 % > luatex
855 \@ifundefined{bbl@hyphenation@}{% Can be \relax!
856   \begingroup
857     \bbl@xin@{\number\language,}{,\bbl@hyphlist}%
858     \ifin@ \else
859       \@expandtwoargs\bbl@usehooks{hyphenation}{#1}{\bbl@tempa}}%
860     \hyphenation{%
861       \bbl@hyphenation@
862       \@ifundefined{bbl@hyphenation@#1}%
863       \@empty
864       {\space\csname bbl@hyphenation@#1\endcsname}}%
865     \xdef\bbl@hyphlist{\bbl@hyphlist\number\language,}%
866   \fi
867   \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, `\lcode's` and font encodings are not set at all, so in most cases you should use `otherlanguage*`.

```

868 \def\hyphenrules#1{%
869   \edef\bbl@tempf{#1}%
870   \bbl@fixname\bbl@tempf
871   \bbl@iflanguage\bbl@tempf{%
872     \expandafter\bbl@patterns\expandafter{\bbl@tempf}%
873     \ifx\languageshorthands\undefined\else
874       \languageshorthands{none}%
875     \fi
876     \expandafter\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
877       \set@hyphenmins\tw@\thr@@\relax
878     \else
879       \expandafter\expandafter\expandafter\set@hyphenmins
880       \csname\bbl@tempf hyphenmins\endcsname\relax
881     \fi}}
882 \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.



```

883 \def\providehyphenmins#1#2{%
884   \expandafter\ifx\csname #1hyphenmins\endcsname\relax
885     \@namedef{#1hyphenmins}{#2}%
886   \fi}

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

887 \def\set@hyphenmins#1#2{%
888   \lefthyphenmin#1\relax
889   \righthyphenmin#2\relax}

\ProvidesLanguage The identification code for each file is something that was introduced in  $\text{\LaTeX}$  2 $\epsilon$ . When the
command \ProvidesFile does not exist, a dummy definition is provided temporarily. For use in the
language definition file the command \ProvidesLanguage is defined by babel.
Depending on the format, ie, on if the former is defined, we use a similar definition or not.

890 \ifx\ProvidesFile\undefined
891   \def\ProvidesLanguage#1[#2 #3 #4]{%
892     \wlog{Language: #1 #4 #3 <#2>}%
893   }
894 \else
895   \def\ProvidesLanguage#1{%
896     \begingroup
897     \catcode\ 10 %
898     \@makeother\/%
899     \@ifnextchar[%]
900       {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}
901   \def\@provideslanguage#1[#2]{%
902     \wlog{Language: #1 #2}%
903     \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
904     \endgroup}
905 \fi

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

906 \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.

907 \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’:

908 \providecommand\setlocale{%
909   \bbl@error
910   {Not yet available}%
911   {Find an armchair, sit down and wait}}
912 \let\uselocale\setlocale
913 \let\locale\setlocale
914 \let\selectlocale\setlocale
915 \let\textlocale\setlocale
916 \let\textlanguage\setlocale
917 \let\languagetext\setlocale

```

## 7.2 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  $\text{\LaTeX}$  2 $\epsilon$ , so we can safely use its error handling interface. Otherwise we’ll have to ‘keep it simple’.

Infos are not written to the console, but on the other hand many people think warnings are errors, so a further message type is defined: an important info which is sent to the console.

```

918 \edef\bbl@nulllanguage{\string\language=0}
919 \def\bbl@nocaption{\protect\bbl@nocaption@i}
920 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
921   \global\@namedef{#2}{\textbf{?#1?}}}%
922   \@nameuse{#2}%
923 \edef\bbl@tempa{#1}%
924 \bbl@sreplace\bbl@tempa{name}{}}%
925 \bbl@warning{% TODO.
926   \@backslashchar#1 not set for '\language'. Please,\\%
927   define it after the language has been loaded\\%
928   (typically in the preamble) with:\\%
929   \string\setlocalecaption{\language}{\bbl@tempa}{..}\\%
930   Reported}}
931 \def\bbl@tentative{\protect\bbl@tentative@i}
932 \def\bbl@tentative@i#1{%
933   \bbl@warning{%
934     Some functions for '#1' are tentative.\\%
935     They might not work as expected and their behavior\\%
936     could change in the future.\\%
937     Reported}}
938 \def\@nolanerr#1{%
939   \bbl@error
940   {You haven't defined the language '#1' yet.\\%
941     Perhaps you misspelled it or your installation\\%
942     is not complete}%
943   {Your command will be ignored, type <return> to proceed}}
944 \def\@nopatterns#1{%
945   \bbl@warning
946   {No hyphenation patterns were preloaded for\\%
947     the language '#1' into the format.\\%
948     Please, configure your TeX system to add them and\\%
949     rebuild the format. Now I will use the patterns\\%
950     preloaded for \bbl@nulllanguage\space instead}}
951 \let\bbl@usehooks\@gobbles
952 \ifx\bbl@onlyswitch\@empty\endinput\fi
953 % Here ended switch.def

```

Here ended the now discarded switch.def. Here also (currently) ends the base option.

```

954 \ifx\directlua\@undefined\else
955   \ifx\bbl@luapatterns\@undefined
956     \input luababel.def
957   \fi
958 \fi
959 <Basic macros>
960 \bbl@trace{Compatibility with language.def}
961 \ifx\bbl@languages\@undefined
962   \ifx\directlua\@undefined
963     \openin1 = language.def % TODO. Remove hardcoded number
964     \ifeof1
965       \closein1
966       \message{I couldn't find the file language.def}
967     \else
968       \closein1
969       \begin{group}
970         \def\addlanguage#1#2#3#4#5{%
971           \expandafter\ifx\csname lang@#1\endcsname\relax\else
972             \global\expandafter\let\csname l@#1\endcsname
973             \csname lang@#1\endcsname
974           \fi}%
975         \def\uselanguage#1{%
976           \input language.def

```

```

977     \endgroup
978   \fi
979 \fi
980 \chardef\l@english\z@
981 \fi

```

`\addto` It takes two arguments, a *<control sequence>* and  $\TeX$ -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.

```

982 \def\addto#1#2{%
983   \ifx#1\undefined
984     \def#1{#2}%
985   \else
986     \ifx#1\relax
987       \def#1{#2}%
988     \else
989       {\toks@\expandafter{#1#2}%
990        \xdef#1{\the\toks@}}%
991   \fi
992 \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?

```

993 \def\bbl@withactive#1#2{%
994   \begingroup
995   \lccode`~=`#2\relax
996   \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`.

```

997 \def\bbl@redefine#1{%
998   \edef\bbl@tempa{\bbl@stripslash#1}%
999   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
1000   \expandafter\def\csname\bbl@tempa\endcsname}
1001 \@onlypreamble\bbl@redefine

```

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

```

1002 \def\bbl@redefine@long#1{%
1003   \edef\bbl@tempa{\bbl@stripslash#1}%
1004   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
1005   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
1006 \@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_`.

```

1007 \def\bbl@redefineroobust#1{%
1008   \edef\bbl@tempa{\bbl@stripslash#1}%
1009   \bbl@ifunset{\bbl@tempa\space}%
1010   {\expandafter\let\csname org@\bbl@tempa\endcsname#1%
1011    \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
1012   {\bbl@exp{\let\<org@\bbl@tempa>\<\bbl@tempa\space>}}%
1013   \namedef{\bbl@tempa\space}}
1014 \@onlypreamble\bbl@redefineroobust

```

### 7.3 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.

```

1015 \bbl@trace{Hooks}
1016 \newcommand\AddBabelHook[3][\%
1017   \bbl@ifunset{bbl@hk@#2}{\EnableBabelHook{#2}}{\%
1018   \def\bbl@tempa##1,##2,##3\@empty{\def\bbl@tempb{##2}}\%
1019   \expandafter\bbl@tempa\bbl@evargs,##3,\@empty
1020   \bbl@ifunset{bbl@ev@#2@#3@#1}\%
1021     {\bbl@csarg\bbl@add{ev@#3@#1}{\bbl@elth{#2}}}\%
1022     {\bbl@csarg\let{ev@#2@#3@#1}\relax}\%
1023   \bbl@csarg\newcommand{ev@#2@#3@#1}{\bbl@tempb}}
1024 \newcommand\EnableBabelHook[1]{\bbl@csarg\let{hk@#1}\@firstofone}
1025 \newcommand\DisableBabelHook[1]{\bbl@csarg\let{hk@#1}\@gobble}
1026 \def\bbl@usehooks#1#2{\%
1027   \ifx\UseHook\@undefined\else\UseHook{babel/*/#1}\fi
1028   \def\bbl@elth##1{\%
1029     \bbl@cs{hk@##1}{\bbl@cs{ev@##1@#1@#2}}\%
1030     \bbl@cs{ev@#1@}\%
1031     \ifx\language\@undefined\else % Test required for Plain (?)
1032       \ifx\UseHook\@undefined\else\UseHook{babel/\language/#1}\fi
1033       \def\bbl@elth##1{\%
1034         \bbl@cs{hk@##1}{\bbl@cl{ev@##1@#1@#2}}\%
1035         \bbl@cl{ev@#1@}\%
1036       \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).

```

1037 \def\bbl@evargs{\,% <- don't delete this comma
1038   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
1039   adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
1040   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
1041   hyphenation=2,initiateactive=3,afterreset=0,foreign=0,foreign*=0,%
1042   beforestart=0,language=2}
1043 \ifx\NewHook\@undefined\else
1044   \def\bbl@tempa#1=#2\@{\NewHook{babel/#1}}
1045   \bbl@foreach\bbl@evargs{\bbl@tempa#1\@}
1046 \fi

```

`\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.

```

1047 \bbl@trace{Defining babelensure}
1048 \newcommand\babelensure[2][\% TODO - revise test files
1049   \AddBabelHook{babel-ensure}{afterextras}{\%
1050     \ifcase\bbl@select@type
1051       \bbl@cl{e}\%
1052     \fi}\%
1053   \beginngroup
1054     \let\bbl@ens@include\@empty
1055     \let\bbl@ens@exclude\@empty
1056     \def\bbl@ens@fontenc{\relax}\%
1057     \def\bbl@tempb##1{\%
1058       \ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}\%

```

```

1059 \edef\bbl@tempa{\bbl@tempb#1\@empty}%
1060 \def\bbl@tempb##1=##2\@{\@namedef{\bbl@ens@##1}{##2}}%
1061 \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
1062 \def\bbl@tempc{\bbl@ensure}%
1063 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
1064 \expandafter{\bbl@ens@include}}%
1065 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
1066 \expandafter{\bbl@ens@exclude}}%
1067 \toks@\expandafter{\bbl@tempc}%
1068 \bbl@exp{%
1069 \endgroup
1070 \def\<bbl@e@#2>{\the\toks@{\bbl@ens@fontenc}}}%
1071 \def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
1072 \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
1073 \ifx##1\@undefined % 3.32 - Don't assume the macro exists
1074 \edef##1{\noexpand\bbl@nocaption
1075 {\bbl@stripslash##1}{\language\language\bbl@stripslash##1}}%
1076 \fi
1077 \ifx##1\@empty\else
1078 \in@{##1}{#2}%
1079 \ifin@else
1080 \bbl@ifunset{\bbl@ensure@\language}%
1081 {\bbl@exp{%
1082 \\\DeclareRobustCommand\<bbl@ensure@\language>[1]{%
1083 \\\foreignlanguage{\language}%
1084 {\ifx\relax#3\else
1085 \\\fontencoding{#3}\selectfont
1086 \fi
1087 #####1}}}%
1088 {}%
1089 \toks@\expandafter{##1}%
1090 \edef##1{%
1091 \bbl@csarg\noexpand{ensure@\language}%
1092 {\the\toks@}}%
1093 \fi
1094 \expandafter\bbl@tempb
1095 \fi}%
1096 \expandafter\bbl@tempb\bbl@captionslist\today\@empty
1097 \def\bbl@tempa##1{% elt for include list
1098 \ifx##1\@empty\else
1099 \bbl@csarg\in@{ensure@\language\expandafter}\expandafter{##1}%
1100 \ifin@else
1101 \bbl@tempb##1\@empty
1102 \fi
1103 \expandafter\bbl@tempa
1104 \fi}%
1105 \bbl@tempa#1\@empty}
1106 \def\bbl@captionslist{%
1107 \prefacename\refname\abstractname\bibname\chaptername\appendixname
1108 \contentsname\listfigurename\listtablename\indexname\figurename
1109 \tablename\partname\enclname\ccname\headtoname\pagename\seename
1110 \alsoname\proofname\glossaryname}

```

## 7.4 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.

```

1111 \bbl@trace{Macros for setting language files up}
1112 \def\bbl@ldfinit{%
1113   \let\bbl@screset\@empty
1114   \let\BabelStrings\bbl@opt@string
1115   \let\BabelOptions\@empty
1116   \let\BabelLanguages\relax
1117   \ifx\originalTeX\@undefined
1118     \let\originalTeX\@empty
1119   \else
1120     \originalTeX
1121   \fi}
1122 \def\LdfInit#1#2{%
1123   \chardef\atcatcode=\catcode`\@
1124   \catcode`\@=11\relax
1125   \chardef\eqcatcode=\catcode`\=
1126   \catcode`\==12\relax
1127   \expandafter\if\expandafter\@backslashchar
1128     \expandafter\@car\string#2\@nil
1129   \ifx#2\@undefined\else
1130     \ldf@quit{#1}%
1131   \fi
1132 \else
1133   \expandafter\ifx\csname#2\endcsname\relax\else
1134     \ldf@quit{#1}%
1135   \fi
1136 \fi
1137 \bbl@ldfinit}

```

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

```

1138 \def\ldf@quit#1{%
1139   \expandafter\main@language\expandafter{#1}%
1140   \catcode`\@=\atcatcode \let\atcatcode\relax
1141   \catcode`\==\eqcatcode \let\eqcatcode\relax
1142   \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.

```

1143 \def\bbl@afterldf#1{% TODO. Merge into the next macro? Unused elsewhere
1144   \bbl@afterlang
1145   \let\bbl@afterlang\relax
1146   \let\BabelModifiers\relax
1147   \let\bbl@screset\relax}%
1148 \def\ldf@finish#1{%
1149   \loadlocalcfg{#1}%
1150   \bbl@afterldf{#1}%
1151   \expandafter\main@language\expandafter{#1}%
1152   \catcode`\@=\atcatcode \let\atcatcode\relax
1153   \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  $\LaTeX$ .

```
1154 \@onlypreamble\LdfInit
1155 \@onlypreamble\ldf@quit
1156 \@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.

```
1157 \def\main@language#1{%
1158   \def\bbl@main@language{#1}%
1159   \let\language\main@language % TODO. Set localename
1160   \bbl@id@assign
1161   \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`.

```
1162 \def\bbl@beforestart{%
1163   \def\@nolanerr##1{%
1164     \bbl@warning{Undefined language '##1' in aux.\\Reported}}%
1165   \bbl@usehooks{beforestart}}%
1166   \global\let\bbl@beforestart\relax}
1167 \AtBeginDocument{%
1168   {\@nameuse{bbl@beforestart}}% Group!
1169   \if@filesw
1170     \providecommand\babel@aux[2]{}%
1171     \immediate\write\@mainaux{%
1172       \string\providecommand\string\babel@aux[2]}%
1173     \immediate\write\@mainaux{\string\@nameuse{bbl@beforestart}}%
1174   \fi
1175   \expandafter\selectlanguage\expandafter{\bbl@main@language}%
1176   \ifbbl@single % must go after the line above.
1177     \renewcommand\selectlanguage[1]{}%
1178     \renewcommand\foreignlanguage[2]{#2}%
1179     \global\let\babel@aux\@gobbletwo % Also as flag
1180   \fi
1181   \ifcase\bbl@engine\or\pagedir\bodydir\fi % TODO - a better place
```

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

```
1182 \def\select@language@x#1{%
1183   \ifcase\bbl@select@type
1184     \bbl@ifsamestring\language{#1}{\select@language{#1}}%
1185   \else
1186     \select@language{#1}%
1187   \fi}
```

## 7.5 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  $\LaTeX$  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.

```
1188 \bbl@trace{Shorhands}
1189 \def\bbl@add@special#1{% 1:a macro like \", \?, etc.
1190   \bbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
1191   \bbl@ifunset{@sanitize}{\bbl@add\@sanitize{\@makeother#1}}%
1192   \ifx\nfss@catcodes\undefined\else % TODO - same for above
1193     \begingroup
1194       \catcode`#1\active
1195     \nfss@catcodes
```

```

1196     \ifnum\catcode`#1=\active
1197     \endgroup
1198     \bbl@add\nfss@catcodes{\@makeother#1}%
1199     \else
1200     \endgroup
1201     \fi
1202 \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.

```

1203 \def\bbl@remove@special#1{%
1204   \begingroup
1205   \def\x##1##2{\ifnum`#1=##2\noexpand\@empty
1206     \else\noexpand##1\noexpand##2\fi}%
1207   \def\do{\x\do}%
1208   \def\@makeother{\x\@makeother}%
1209   \edef\x{\endgroup
1210     \def\noexpand\dospecials{\dospecials}%
1211     \expandafter\ifx\csname @sanitize\endcsname\relax\else
1212     \def\noexpand\@sanitize{\@sanitize}%
1213     \fi}%
1214   \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).

```

1215 \def\bbl@active@def#1#2#3#4{%
1216   \@namedef{#3#1}{%
1217     \expandafter\ifx\csname#2@sh@#1\endcsname\relax
1218     \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
1219     \else
1220     \bbl@afterfi\csname#2@sh@#1\endcsname
1221     \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.

```

1222   \long\@namedef{#3@arg#1}##1{%
1223     \expandafter\ifx\csname#2@sh@#1\string##1\endcsname\relax
1224     \bbl@afterelse\csname#4#1\endcsname##1%
1225     \else
1226     \bbl@afterfi\csname#2@sh@#1\string##1\endcsname
1227     \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.

```

1228 \def\initiate@active@char#1{%
1229   \bbl@ifunset{active@char\string#1}%

```



```

1230    {\bbl@withactive
1231      {\expandafter\@initiate@active@char\expandafter}\#1\string#1#1}%
1232    {}

```

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` and preserving some degree of protection).

```

1233 \def\@initiate@active@char#1#2#3{%
1234   \bbl@csarg\edef\oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%
1235   \ifx#1\@undefined
1236     \bbl@csarg\def\oridef@#2}{\def#1{\active@prefix#1\@undefined}}%
1237   \else
1238     \bbl@csarg\let\oridef@@#2}#1%
1239     \bbl@csarg\edef\oridef@#2}{%
1240       \let\noexpand#1%
1241       \expandafter\noexpand\csname bbl@oridef@@#2\endcsname}%
1242   \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`).

```

1243   \ifx#1#3\relax
1244     \expandafter\let\csname normal@char#2\endcsname#3%
1245   \else
1246     \bbl@info{Making #2 an active character}%
1247     \ifnum\mathcode`#2=\ifodd\bbl@engine"1000000 \else"8000 \fi
1248     \@namedef{normal@char#2}{%
1249       \textormath{#3}{\csname bbl@oridef@@#2\endcsname}}%
1250   \else
1251     \@namedef{normal@char#2}{#3}%
1252   \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).

```

1253   \bbl@restoreactive{#2}%
1254   \AtBeginDocument{%
1255     \catcode`#2\active
1256     \if@filesw
1257       \immediate\write\@mainaux{\catcode`\string#2\active}%
1258     \fi}%
1259   \expandafter\bbl@add@special\csname#2\endcsname
1260   \catcode`#2\active
1261 \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⟩`).

```

1262   \let\bbl@tempa\@firstoftwo
1263   \if\string^#2%
1264     \def\bbl@tempa{\noexpand\textormath}%
1265   \else
1266     \ifx\bbl@mathnormal\@undefined\else
1267       \let\bbl@tempa\bbl@mathnormal
1268     \fi
1269   \fi
1270   \expandafter\edef\csname active@char#2\endcsname{%

```

```

1271 \bbl@tempa
1272 {\noexpand\if@safe@actives
1273 \noexpand\expandafter
1274 \expandafter\noexpand\csname normal@char#2\endcsname
1275 \noexpand\else
1276 \noexpand\expandafter
1277 \expandafter\noexpand\csname bbl@doactive#2\endcsname
1278 \noexpand\fi}%
1279 {\expandafter\noexpand\csname normal@char#2\endcsname}}%
1280 \bbl@csarg\edef{doactive#2}{%
1281 \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

$$\backslash active@prefix \langle char \rangle \backslash normal@char \langle char \rangle$$

(where  $\backslash active@char \langle char \rangle$  is *one* control sequence!).

```

1282 \bbl@csarg\edef{active@#2}{%
1283 \noexpand\active@prefix\noexpand#1%
1284 \expandafter\noexpand\csname active@char#2\endcsname}%
1285 \bbl@csarg\edef{normal@#2}{%
1286 \noexpand\active@prefix\noexpand#1%
1287 \expandafter\noexpand\csname normal@char#2\endcsname}%
1288 \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.

```

1289 \bbl@active@def#2\user@group{user@active}{language@active}%
1290 \bbl@active@def#2\language@group{language@active}{system@active}%
1291 \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  $\backslash protect' \backslash protect'$ . To prevent this from happening a couple of shorthand needs to be defined at user level.

```

1292 \expandafter\edef\csname\user@group @sh@#2@@\endcsname
1293 {\expandafter\noexpand\csname normal@char#2\endcsname}%
1294 \expandafter\edef\csname\user@group @sh@#2@\string\protect\endcsname
1295 {\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  $\backslash prim@s$  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.

```

1296 \if\string'#2%
1297 \let\prim@s\bbl@prim@s
1298 \let\active@math@prime#1%
1299 \fi
1300 \bbl@usehooks{initiateactive}{{#1}{#2}{#3}}

```

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

```

1301 \langle *More package options \rangle \equiv
1302 \DeclareOption{math=active}{}
1303 \DeclareOption{math=normal}{{\def\bbl@mathnormal{\noexpand\textormath}}}
1304 \rangle /More package options \rangle

```

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.

```

1305 \ifpackagewith{babel}{KeepShorthandsActive}%
1306 {\let\bbl@restoreactive\@gobble}%

```

```

1307 {\def\bbl@restoreactive#1{%
1308   \bbl@exp{%
1309     \\\AfterBabelLanguage\\CurrentOption
1310     {\catcode`#1=\the\catcode`#1\relax}%
1311     \\\AtEndOfPackage
1312     {\catcode`#1=\the\catcode`#1\relax}}}%
1313   \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.

```

1314 \def\bbl@sh@select#1#2{%
1315   \expandafter\ifx\csname#1@sh#2@sel\endcsname\relax
1316     \bbl@afterelse\bbl@scndcs
1317   \else
1318     \bbl@afterfi\csname#1@sh#2@sel\endcsname
1319   \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.

```

1320 \begingroup
1321 \bbl@ifunset{ifincsname}% TODO. Ugly. Correct? Only Plain?
1322 {\gdef\active@prefix#1{%
1323   \ifx\protect\@typeset@protect
1324   \else
1325     \ifx\protect\@unexpandable@protect
1326       \noexpand#1%
1327     \else
1328       \protect#1%
1329     \fi
1330     \expandafter\@gobble
1331   \fi}}
1332 {\gdef\active@prefix#1{%
1333   \ifincsname
1334     \string#1%
1335     \expandafter\@gobble
1336   \else
1337     \ifx\protect\@typeset@protect
1338     \else
1339       \ifx\protect\@unexpandable@protect
1340         \noexpand#1%
1341       \else
1342         \protect#1%
1343       \fi
1344       \expandafter\expandafter\expandafter\@gobble
1345     \fi
1346   \fi}}
1347 \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*).

```

1348 \newif\if@safe@actives
1349 \@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.

```

1350 \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.

1351 \chardef\bbl@activated\z@
1352 \def\bbl@activate#1{%
1353   \chardef\bbl@activated\@ne
1354   \bbl@withactive{\expandafter\let\expandafter}\#1%
1355   \csname bbl@active@\string#1\endcsname}
1356 \def\bbl@deactivate#1{%
1357   \chardef\bbl@activated\tw@
1358   \bbl@withactive{\expandafter\let\expandafter}\#1%
1359   \csname bbl@normal@\string#1\endcsname}

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

1360 \def\bbl@firstcs#1#2{\csname#1\endcsname}
1361 \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.

1362 \def\babel@texpdf#1#2#3#4{%
1363   \ifx\texorpdfstring\undefined
1364     \textormath{#1}{#3}%
1365   \else
1366     \texorpdfstring{\textormath{#1}{#3}}{#2}%
1367     % \texorpdfstring{\textormath{#1}{#3}}{\textormath{#2}{#4}}%
1368   \fi}
1369 %
1370 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil}
1371 \def\@decl@short#1#2#3\@nil#4{%
1372   \def\bbl@tempa{#3}%
1373   \ifx\bbl@tempa\@empty
1374     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs
1375     \bbl@ifunset{#1@sh@\string#2@}{}%
1376     {\def\bbl@tempa{#4}%
1377      \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa
1378      \else
1379        \bbl@info
1380        {Redefining #1 shorthand \string#2\}%
1381        in language \CurrentOption}%
1382     \fi}%
1383   \@namedef{#1@sh@\string#2@}{#4}%
1384 \else
1385   \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs
1386   \bbl@ifunset{#1@sh@\string#2@\string#3@}{}%
1387   {\def\bbl@tempa{#4}%
1388    \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
1389    \else
1390      \bbl@info
1391      {Redefining #1 shorthand \string#2\string#3\}%
1392      in language \CurrentOption}%
1393   \fi}%

```

```

1394 \namedef{#1@sh@string#2@string#3@}{#4}%
1395 \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.

1396 \def\textormath{%
1397 \ifmode
1398 \expandafter\@secondoftwo
1399 \else
1400 \expandafter\@firstoftwo
1401 \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’.

1402 \def\user@group{user}
1403 \def\language@group{english} % TODO. I don't like defaults
1404 \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.

1405 \def\useshorthands{%
1406 \ifstar\bbl@usesh@s{\bbl@usesh@x{}}%
1407 \def\bbl@usesh@s#1{%
1408 \bbl@usesh@x
1409 {\AddBabelHook{babel-sh-}\string#1}{afterextras}{\bbl@activate{#1}}}%
1410 {#1}}
1411 \def\bbl@usesh@x#1#2{%
1412 \bbl@ifshorthand{#2}%
1413 {\def\user@group{user}%
1414 \initiate@active@char{#2}%
1415 #1%
1416 \bbl@activate{#2}}%
1417 {\bbl@error
1418 {I can't declare a shorthand turned off (\string#2)}
1419 {Sorry, but you can't use shorthands which have been\\
1420 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.

1421 \def\user@language@group{user@\language@group}
1422 \def\bbl@set@user@generic#1#2{%
1423 \bbl@ifunset{user@generic@active#1}%
1424 {\bbl@active@def#1\user@language@group{user@active}{user@generic@active}%
1425 \bbl@active@def#1\user@group{user@generic@active}{language@active}%
1426 \expandafter\edef\csname#2@sh@#1@\endcsname{%
1427 \expandafter\noexpand\csname normal@char#1\endcsname}%
1428 \expandafter\edef\csname#2@sh@#1@\string\protect@\endcsname{%
1429 \expandafter\noexpand\csname user@active#1\endcsname}}%
1430 \@empty}
1431 \newcommand\defineshorthand[3][user]{%
1432 \edef\bbl@tempa{\zap@space#1 \@empty}%
1433 \bbl@for\bbl@tempb\bbl@tempa{%
1434 \if*\expandafter\car\bbl@tempb\nil
1435 \edef\bbl@tempb{user@\expandafter@gobble\bbl@tempb}%
1436 \@expandtwoargs
1437 \bbl@set@user@generic{\expandafter\string\car#2\nil}\bbl@tempb
1438 \fi
1439 \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].

```
1440 \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`.

```
1441 \def \aliasshorthand#1#2{%
1442   \bbl@ifshorthand{#2}%
1443   {\expandafter \ifx \csname active@char \string#2 \endcsname \relax
1444     \ifx \document \notprerr
1445       \notshorthand{#2}%
1446     \else
1447       \initiate@active@char{#2}%
1448       \expandafter \let \csname active@char \string#2 \expandafter \endcsname
1449         \csname active@char \string#1 \endcsname
1450       \expandafter \let \csname normal@char \string#2 \expandafter \endcsname
1451         \csname normal@char \string#1 \endcsname
1452       \bbl@activate{#2}%
1453     \fi
1454   \fi}%
1455   {\bbl@error
1456     {Cannot declare a shorthand turned off (\string#2)}
1457     {Sorry, but you cannot use shorthands which have been\\
1458       turned off in the package options}}}
```

`\notshorthand`

```
1459 \def \notshorthand#1{%
1460   \bbl@error{%
1461     The character '\string #1' should be made a shorthand character;\\%
1462     add the command \string\usesshorthands\string{#1\string} to
1463     the preamble.\\%
1464     I will ignore your instruction}%
1465   {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.

```
1466 \newcommand* \shorthandon[1]{\bbl@switch@sh \@ne#1 \@nnil}
1467 \DeclareRobustCommand* \shorthandoff{%
1468   \@ifstar{\bbl@shorthandoff\tw}{\bbl@shorthandoff\z@}}
1469 \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.

```
1470 \def \bbl@switch@sh#1#2{%
1471   \ifx#2 \@nnil \else
1472     \bbl@ifunset{\bbl@active@ \string#2}%
1473     {\bbl@error
1474       {I can't switch '\string#2' on or off--not a shorthand}%
1475       {This character is not a shorthand. Maybe you made\\
1476         a typing mistake? I will ignore your instruction.}}%
1477     {\ifcase#1%   off, on, off*
1478       \catcode`#2\relax
1479     \or
1480       \catcode`#2\active
1481     \bbl@ifunset{\bbl@shdef@ \string#2}%
1482     {\bbl@shdef@ \string#2 \active}}
```

```

1482         {}%
1483         {\bbl@withactive{\expandafter\let\expandafter}\#2%
1484             \csname bbl@shdef@\string#2\endcsname
1485             \bbl@csarg\let{shdef@\string#2}\relax}%
1486         \ifcase\bbl@activated\or
1487             \bbl@activate{#2}%
1488         \else
1489             \bbl@deactivate{#2}%
1490         \fi
1491     \or
1492         \bbl@ifunset{bbl@shdef@\string#2}%
1493         {\bbl@withactive{\bbl@csarg\let{shdef@\string#2}}{#2}%
1494         {}%
1495         \csname bbl@oricat@\string#2\endcsname
1496         \csname bbl@oridef@\string#2\endcsname
1497     \fi}%
1498     \bbl@afterfi\bbl@switch@sh#1%
1499 \fi}

```

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

```

1500 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
1501 \def\bbl@putsh#1{%
1502     \bbl@ifunset{bbl@active@\string#1}%
1503     {\bbl@putsh@i#1\@empty\@nnil}%
1504     {\csname bbl@active@\string#1\endcsname}}
1505 \def\bbl@putsh@i#1#2\@nnil{%
1506     \csname\language@group @sh@\string#1@%
1507     \ifx\@empty#2\else\string#2\fi\endcsname}
1508 \ifx\bbl@opt@shorthands\@nnil\else
1509     \let\bbl@s@initiate@active@char\initiate@active@char
1510     \def\initiate@active@char#1{%
1511         \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
1512     \let\bbl@s@switch@sh\bbl@switch@sh
1513     \def\bbl@switch@sh#1#2{%
1514         \ifx#2\@nnil\else
1515             \bbl@afterfi
1516             \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
1517         \fi}
1518     \let\bbl@s@activate\bbl@activate
1519     \def\bbl@activate#1{%
1520         \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
1521     \let\bbl@s@deactivate\bbl@deactivate
1522     \def\bbl@deactivate#1{%
1523         \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
1524 \fi

```

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

```

1525 \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.

```

1526 \def\bbl@prim@s{%
1527     \prime\futurelet\@let@token\bbl@pr@m@s}
1528 \def\bbl@if@primes#1#2{%
1529     \ifx#1\@let@token
1530         \expandafter\@firstoftwo
1531     \else\ifx#2\@let@token
1532         \bbl@afterelse\expandafter\@firstoftwo
1533     \else
1534         \bbl@afterfi\expandafter\@secondoftwo

```

```

1535 \fi\fi}
1536 \begingroup
1537 \catcode`\^=7 \catcode`\*=\active \lccode`\*=\^
1538 \catcode`\'=12 \catcode`\\"=\active \lccode`\\"=\'
1539 \lowercase{%
1540 \gdef\bbl@pr@m@s{%
1541 \bbl@if@primes''%
1542 \pr@@@s
1543 {\bbl@if@primes*^\pr@@@t\egroup}}
1544 \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).

```

1545 \initiate@active@char{~}
1546 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
1547 \bbl@activate{~}

```

`\OT1dpos` The position of the double quote character is different for the OT1 and T1 encodings. It will later be  
`\T1dpos` selected using the `\f@encoding` macro. Therefore we define two macros here to store the position of the character in these encodings.

```

1548 \expandafter\def\csname OT1dpos\endcsname{127}
1549 \expandafter\def\csname T1dpos\endcsname{4}

```

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

```

1550 \ifx\f@encoding\undefined
1551 \def\f@encoding{OT1}
1552 \fi

```

## 7.6 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.

```

1553 \bbl@trace{Language attributes}
1554 \newcommand\languageattribute[2]{%
1555 \def\bbl@tempc{#1}%
1556 \bbl@fixname\bbl@tempc
1557 \bbl@iflanguage\bbl@tempc{%
1558 \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.

```

1559 \ifx\bbl@known@attribs\undefined
1560 \in@false
1561 \else
1562 \bbl@xin@{\bbl@tempc-##1,}{\bbl@known@attribs,%}
1563 \fi
1564 \ifin@
1565 \bbl@warning{%
1566 You have more than once selected the attribute '##1'\%
1567 for language #1. Reported}%
1568 \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.

```

1569      \bbl@exp{%
1570        \\bbl@add@list\\bbl@known@attribs{\bbl@tempc-##1}}%
1571      \edef\bbl@tempa{\bbl@tempc-##1}%
1572      \expandafter\bbl@ifknown@trib\expandafter{\bbl@tempa}\bbl@attributes%
1573      {\csname\bbl@tempc @attr@##1\endcsname}%
1574      {\@attrerr{\bbl@tempc}{##1}}%
1575      \fi}}
1576 \onlypreamble\languageattribute

```

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

```

1577 \newcommand*{\@attrerr}[2]{%
1578   \bbl@error
1579   {The attribute #2 is unknown for language #1.}%
1580   {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}`.

```

1581 \def\bbl@declare@ttribute#1#2#3{%
1582   \bbl@xin@{,#2,}{,\BabelModifiers,}%
1583   \ifin@
1584     \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
1585   \fi
1586   \bbl@add@list\bbl@attributes{#1-#2}%
1587   \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.

```

1588 \def\bbl@ifattributeset#1#2#3#4{%
1589   \ifx\bbl@known@attribs\@undefined
1590     \in@false
1591   \else
1592     \bbl@xin@{,#1-#2,}{,\bbl@known@attribs,}%
1593   \fi
1594   \ifin@
1595     \bbl@afterelse#3%
1596   \else
1597     \bbl@afterfi#4%
1598   \fi}

```

**\bbl@ifknown@trib** 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.

```

1599 \def\bbl@ifknown@trib#1#2{%
1600   \let\bbl@tempa\@secondoftwo
1601   \bbl@loopx\bbl@tempb{#2}{%
1602     \expandafter\in\expandafter{\expandafter,\bbl@tempb,}{,#1,}%
1603     \ifin@
1604       \let\bbl@tempa\@firstoftwo
1605     \else
1606       \fi}%
1607   \bbl@tempa}

```

`\bbl@clear@ttribs` This macro removes all the attribute code from L<sup>A</sup>T<sub>E</sub>X's memory at `\begin{document}` time (if any is present).

```

1608 \def\bbl@clear@ttribs{%
1609   \ifx\bbl@attributes\undefined\else
1610     \bbl@loopx\bbl@tempa{\bbl@attributes}{%
1611       \expandafter\bbl@clear@ttrib\bbl@tempa.
1612     }%
1613   \let\bbl@attributes\undefined
1614   \fi}
1615 \def\bbl@clear@ttrib#1-#2.{%
1616   \expandafter\let\csname#1@attr@#2\endcsname\undefined}
1617 \AtBeginDocument{\bbl@clear@ttribs}

```

## 7.7 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`

```

1618 \bbl@trace{Macros for saving definitions}
1619 \def\babel@beginsave{\babel@savecnt\z@}

```

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

```

1620 \newcount\babel@savecnt
1621 \babel@beginsave

```

`\babel@save` The macro `\babel@save⟨csname⟩` saves the current meaning of the control sequence `⟨csname⟩` to `\originalTeX`<sup>32</sup>. To do this, we let the current meaning to a temporary control 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.

```

1622 \def\babel@save#1{%
1623   \expandafter\let\csname babel@\number\babel@savecnt\endcsname#1\relax
1624   \toks@\expandafter{\originalTeX\let#1=}
1625   \bbl@exp{%
1626     \def\originalTeX{\the\toks@<\babel@\number\babel@savecnt>\relax}}%
1627   \advance\babel@savecnt\@ne}
1628 \def\babel@savevariable#1{%
1629   \toks@\expandafter{\originalTeX #1=}
1630   \bbl@exp{\def\originalTeX{\the\toks@the#1\relax}}}

```

`\bbl@frenchspacing` Some languages need to have `\frenchspacing` in effect. Others don't want that. The command `\bbl@nonfrenchspacing` switches it on when it isn't already in effect and `\bbl@nonfrenchspacing` switches it off if necessary. A more refined way to switch the catcodes is done with ini files. Here an auxiliary macro is defined, but the main part is in `\babelprovide`. This new method should be ideally the default one.

```

1631 \def\bbl@frenchspacing{%
1632   \ifnum\the\sfcode`\.=\@m
1633     \let\bbl@nonfrenchspacing\relax
1634   \else
1635     \frenchspacing
1636     \let\bbl@nonfrenchspacing\nonfrenchspacing
1637   \fi}
1638 \let\bbl@nonfrenchspacing\nonfrenchspacing
1639 \let\bbl@elt\relax

```

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

```

1640 \edef\bbl@fs@chars{%
1641   \bbl@elt{\string.}\@m{3000}\bbl@elt{\string?}\@m{3000}%
1642   \bbl@elt{\string!}\@m{3000}\bbl@elt{\string:}\@m{2000}%
1643   \bbl@elt{\string;}\@m{1500}\bbl@elt{\string,}\@m{1250}}%
1644 \def\bbl@pre@fs{%
1645   \def\bbl@elt##1##2##3{\sfcode`##1=\the\sfcode`##1\relax}%
1646   \edef\bbl@save@sfcodes{\bbl@fs@chars}}%
1647 \def\bbl@post@fs{%
1648   \bbl@save@sfcodes
1649   \edef\bbl@tempa{\bbl@cl{frspc}}}%
1650   \edef\bbl@tempa{\expandafter\@car\bbl@tempa\@nil}%
1651   \if u\bbl@tempa      % do nothing
1652   \else\if n\bbl@tempa % non french
1653     \def\bbl@elt##1##2##3{%
1654       \ifnum\sfcode`##1=##2\relax
1655         \babel@savevariable{\sfcode`##1}%
1656         \sfcode`##1=##3\relax
1657       \fi}%
1658     \bbl@fs@chars
1659   \else\if y\bbl@tempa % french
1660     \def\bbl@elt##1##2##3{%
1661       \ifnum\sfcode`##1=##3\relax
1662         \babel@savevariable{\sfcode`##1}%
1663         \sfcode`##1=##2\relax
1664       \fi}%
1665     \bbl@fs@chars
1666   \fi\fi\fi}

```

## 7.8 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.

```

1667 \bbl@trace{Short tags}
1668 \def\babeltags#1{%
1669   \edef\bbl@tempa{\zap@space#1 \@empty}%
1670   \def\bbl@tempb##1=##2\@{%
1671     \edef\bbl@tempc{%
1672       \noexpand\newcommand
1673       \expandafter\noexpand\csname ##1\endcsname{%
1674         \noexpand\protect
1675         \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
1676       \noexpand\newcommand
1677       \expandafter\noexpand\csname text##1\endcsname{%
1678         \noexpand\foreignlanguage{##2}}
1679     \bbl@tempc}%
1680   \bbl@for\bbl@tempa\bbl@tempa{%
1681     \expandafter\bbl@tempb\bbl@tempa\@{}}

```

## 7.9 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.

```

1682 \bbl@trace{Hyphens}
1683 \@onlypreamble\babelhyphenation
1684 \AtEndOfPackage{%
1685   \newcommand\babelhyphenation[2][\@empty]{%
1686     \ifx\bbl@hyphenation@\relax
1687     \let\bbl@hyphenation@\@empty
1688     \fi
1689     \ifx\bbl@hyphlist\@empty\else

```

```

1690     \bbl@warning{%
1691         You must not intermingle \string\selectlanguage\space and\\%
1692         \string\babelhyphenation\space or some exceptions will not\\%
1693         be taken into account. Reported}%
1694     \fi
1695     \ifx\@empty#1%
1696         \protected@edef\bbl@hyphenation@\bbl@hyphenation@\space#2}%
1697     \else
1698         \bbl@vforeach{#1}{%
1699             \def\bbl@tempa{##1}%
1700             \bbl@fixname\bbl@tempa
1701             \bbl@iflanguage\bbl@tempa{%
1702                 \bbl@csarg\protected@edef{hyphenation@\bbl@tempa}{%
1703                     \bbl@ifunset{bbl@hyphenation@\bbl@tempa}%
1704                     {}%
1705                     {\csname bbl@hyphenation@\bbl@tempa\endcsname\space}%
1706                     #2}}}%
1707     \fi}}

```

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

```

1708 \def\bbl@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}
1709 \def\bbl@t@one{T1}
1710 \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`.

```

1711 \newcommand\babellnullhyphen{\char\hyphenchar\font}
1712 \def\babelhyphen{\active@prefix\babelhyphen\bbl@hyphen}
1713 \def\bbl@hyphen{%
1714     \@ifstar{\bbl@hyphen@i @}{\bbl@hyphen@i \@empty}}
1715 \def\bbl@hyphen@i#1#2{%
1716     \bbl@ifunset{bbl@hy@#1#2\@empty}%
1717     {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%
1718     {\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.

```

1719 \def\bbl@usehyphen#1{%
1720     \leavevmode
1721     \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
1722     \nobreak\hskip\z@skip}
1723 \def\bbl@@usehyphen#1{%
1724     \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

The following macro inserts the hyphen char.

```

1725 \def\bbl@hyphenchar{%
1726     \ifnum\hyphenchar\font=\m@ne
1727         \babellnullhyphen
1728     \else
1729         \char\hyphenchar\font
1730     \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.

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

```

1731 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1732 \def\bbl@hy@@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1733 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
1734 \def\bbl@hy@@hard{\bbl@usehyphen\bbl@hyphenchar}
1735 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}{}}
1736 \def\bbl@hy@@nobreak{\mbox{\bbl@hyphenchar}}
1737 \def\bbl@hy@repeat{%
1738   \bbl@usehyphen{%
1739     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
1740 \def\bbl@hy@@repeat{%
1741   \bbl@usehyphen{%
1742     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
1743 \def\bbl@hy@empty{\hskip\zskip}
1744 \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.

```

1745 \def\bbl@disc#1#2{\nobreak\discretionary{#2-}{#1}\bbl@allowhyphens}

```

## 7.10 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.

```

1746 \bbl@trace{Multiencoding strings}
1747 \def\bbl@tglobal#1{\global\let#1#1}
1748 \def\bbl@recatcode#1{% TODO. Used only once?
1749   \@tempcnta="7F
1750   \def\bbl@tempa{%
1751     \ifnum\@tempcnta>"FF\else
1752       \catcode\@tempcnta=#1\relax
1753       \advance\@tempcnta\@ne
1754       \expandafter\bbl@tempa
1755     \fi}%
1756   \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).

```

1757 \ifpackagewith{babel}{nocase}%
1758   {\let\bbl@patchuclc\relax}%
1759   {\def\bbl@patchuclc{%
1760     \global\let\bbl@patchuclc\relax
1761     \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}%
1762     \gdef\bbl@uclc##1{%
1763       \let\bbl@encoded\bbl@encoded@uclc
1764       \bbl@ifunset{\language @bbl@uclc}% and resumes it
1765       {##1}%
1766       {\let\bbl@tempa##1\relax % Used by LANG@bbl@uclc
1767         \csname\language @bbl@uclc\endcsname}%
1768       {\bbl@tolower\@empty}{\bbl@toupper\@empty}}}%

```

```

1769 \gdef\bbl@tolower{\csname\languagename @bbl@lc\endcsname}%
1770 \gdef\bbl@toupper{\csname\languagename @bbl@uc\endcsname}}

1771 <{*More package options}> ≡
1772 \DeclareOption{nocase}{}
1773 <{/More package options}>

```

The following package options control the behavior of \SetString.

```

1774 <{*More package options}> ≡
1775 \let\bbl@opt@strings\@nnil % accept strings=value
1776 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1777 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1778 \def\BabelStringsDefault{generic}
1779 <{/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.

```

1780 \@onlypreamble\StartBabelCommands
1781 \def\StartBabelCommands{%
1782   \begingroup
1783   \bbl@recatcode{11}%
1784   <<Macros local to BabelCommands>>
1785   \def\bbl@provstring##1##2{%
1786     \providecommand##1{##2}%
1787     \bbl@toglobal##1}%
1788   \global\let\bbl@scafter\@empty
1789   \let\StartBabelCommands\bbl@startcmds
1790   \ifx\BabelLanguages\relax
1791     \let\BabelLanguages\CurrentOption
1792   \fi
1793   \begingroup
1794   \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
1795   \StartBabelCommands}
1796 \def\bbl@startcmds{%
1797   \ifx\bbl@screset\@nnil\else
1798     \bbl@usehooks{stopcommands}{}%
1799   \fi
1800   \endgroup
1801   \begingroup
1802   \@ifstar
1803   {\ifx\bbl@opt@strings\@nnil
1804     \let\bbl@opt@strings\BabelStringsDefault
1805   \fi
1806   \bbl@startcmds@i}%
1807   \bbl@startcmds@i}
1808 \def\bbl@startcmds@i#1#2{%
1809   \edef\bbl@L{\zap@space#1 \@empty}%
1810   \edef\bbl@G{\zap@space#2 \@empty}%
1811   \bbl@startcmds@ii}
1812 \let\bbl@startcmds\StartBabelCommands

```

Parse the encoding info to get the label, input, and font parts.

Select the behavior of \SetString. There are two main cases, depending of if there is an optional argument: without it and strings=encoded, strings are defined always; otherwise, they are set only if they are still undefined (ie, fallback values). With labelled blocks and strings=encoded, define the strings, but with another value, define strings only if the current label or font encoding is the value of strings; otherwise (ie, no strings or a block whose label is not in strings=) do nothing.

We presume the current block is not loaded, and therefore set (above) a couple of default values to gobble the arguments. Then, these macros are redefined if necessary according to several parameters.

```

1813 \newcommand\bbl@startcmds@ii[1][\@empty]{%
1814   \let\SetString@gobbletwo

```

```

1815 \let\bbl@stringdef\@gobbletwo
1816 \let\AfterBabelCommands\@gobble
1817 \ifx\@empty#1%
1818   \def\bbl@sc@label{generic}%
1819   \def\bbl@encstring##1##2{%
1820     \ProvideTextCommandDefault##1{##2}%
1821     \bbl@tglobal##1%
1822     \expandafter\bbl@tglobal\csname\string?\string##1\endcsname}%
1823     \let\bbl@sctest\in@true
1824   \else
1825     \let\bbl@sc@charset\space % <- zapped below
1826     \let\bbl@sc@fontenc\space % <- " "
1827     \def\bbl@tempa##1=##2\@nil{%
1828       \bbl@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
1829     \bbl@vforeach{label=#1}{\bbl@tempa##1\@nil}%
1830     \def\bbl@tempa##1 ##2{% space -> comma
1831       ##1%
1832       \ifx\@empty##2\else\ifx,##1,\else,\fi\bbl@afterfi\bbl@tempa##2\fi}%
1833     \edef\bbl@sc@fontenc{\expandafter\bbl@tempa\bbl@sc@fontenc\@empty}%
1834     \edef\bbl@sc@label{\expandafter\zap@space\bbl@sc@label\@empty}%
1835     \edef\bbl@sc@charset{\expandafter\zap@space\bbl@sc@charset\@empty}%
1836     \def\bbl@encstring##1##2{%
1837       \bbl@foreach\bbl@sc@fontenc{%
1838         \bbl@ifunset{T####1}%
1839         }%
1840         {\ProvideTextCommand##1{####1}{##2}%
1841         \bbl@tglobal##1%
1842         \expandafter
1843         \bbl@tglobal\csname####1\string##1\endcsname}}}%
1844     \def\bbl@sctest{%
1845       \bbl@xin@{\bbl@opt@strings,}{\bbl@sc@label,\bbl@sc@fontenc,}}%
1846   \fi
1847   \ifx\bbl@opt@strings\@nnil % ie, no strings key -> defaults
1848   \else\ifx\bbl@opt@strings\relax % ie, strings=encoded
1849     \let\AfterBabelCommands\bbl@aftercmds
1850     \let\SetString\bbl@setstring
1851     \let\bbl@stringdef\bbl@encstring
1852   \else % ie, strings=value
1853     \bbl@sctest
1854   \ifin@
1855     \let\AfterBabelCommands\bbl@aftercmds
1856     \let\SetString\bbl@setstring
1857     \let\bbl@stringdef\bbl@provstring
1858   \fi\fi\fi
1859   \bbl@scswitch
1860   \ifx\bbl@G\@empty
1861     \def\SetString##1##2{%
1862       \bbl@error{Missing group for string \string##1}%
1863       {You must assign strings to some category, typically\\%
1864       captions or extras, but you set none}}%
1865   \fi
1866   \ifx\@empty#1%
1867     \bbl@usehooks{defaultcommands}{}%
1868   \else
1869     \@expandtwoargs
1870     \bbl@usehooks{encodedcommands}{\bbl@sc@charset}{\bbl@sc@fontenc}%
1871   \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) .

```

1872 \def\bbl@forlang#1#2{%
1873   \bbl@for#1\bbl@L{%
1874     \bbl@xin@{,#1,}{,\BabelLanguages,}%
1875     \ifin@#2\relax\fi}}
1876 \def\bbl@scswitch{%
1877   \bbl@forlang\bbl@tempa{%
1878     \ifx\bbl@G\@empty\else
1879       \ifx\SetString\@gobbletwo\else
1880         \edef\bbl@GL{\bbl@G\bbl@tempa}%
1881         \bbl@xin@{\bbl@GL,}{,\bbl@screset,}%
1882         \ifin@\else
1883           \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
1884           \xdef\bbl@screset{\bbl@screset,\bbl@GL}%
1885           \fi
1886         \fi
1887       \fi}}
1888 \AtEndOfPackage{%
1889   \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{\#2}}}%
1890   \let\bbl@scswitch\relax
1891   \@onlypreamble\EndBabelCommands
1892   \def\EndBabelCommands{%
1893     \bbl@usehooks{stopcommands}{}%
1894     \endgroup
1895     \endgroup
1896     \bbl@scafter}
1897   \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.

```

1898 \def\bbl@setstring#1#2{% eg, \prefacename{<string>}
1899   \bbl@forlang\bbl@tempa{%
1900     \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
1901     \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
1902     {\bbl@exp{%
1903       \global\let\bbl@add\<\bbl@G\bbl@tempa>\bbl@scset\#1\<\bbl@LC>}}}%
1904     }%
1905     \def\BabelString{#2}%
1906     \bbl@usehooks{stringprocess}{}%
1907     \expandafter\bbl@stringdef
1908     \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.

```

1909 \ifx\bbl@opt@strings\relax
1910   \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
1911   \bbl@patchuclc
1912   \let\bbl@encoded\relax
1913   \def\bbl@encoded@uclc#1{%
1914     \@inmathwarn#1%
1915     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
1916       \expandafter\ifx\csname ?\string#1\endcsname\relax
1917         \TextSymbolUnavailable#1%
1918       \else
1919         \csname ?\string#1\endcsname
1920       \fi

```



```

1921 \else
1922 \csname\cf@encoding\string#1\endcsname
1923 \fi}
1924 \else
1925 \def\bbl@scset#1#2{\def#1{#2}}
1926 \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.

```

1927 <<(*Macros local to BabelCommands)>> ≡
1928 \def\SetStringLoop##1##2{%
1929 \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
1930 \count@z@
1931 \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
1932 \advance\count@\@ne
1933 \toks@\expandafter{\bbl@tempa}%
1934 \bbl@exp{%
1935 \\\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
1936 \count@=\the\count@\relax}}}%
1937 <</Macros local to BabelCommands>>

```

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

```

1938 \def\bbl@aftercmds#1{%
1939 \toks@\expandafter{\bbl@scafter#1}%
1940 \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.

```

1941 <<(*Macros local to BabelCommands)>> ≡
1942 \newcommand\SetCase[3][]{%
1943 \bbl@patchuclc
1944 \bbl@forlang\bbl@tempa{%
1945 \expandafter\bbl@encstring
1946 \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
1947 \expandafter\bbl@encstring
1948 \csname\bbl@tempa @bbl@uc\endcsname{##2}%
1949 \expandafter\bbl@encstring
1950 \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
1951 <</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.

```

1952 <<(*Macros local to BabelCommands)>> ≡
1953 \newcommand\SetHyphenMap[1]{%
1954 \bbl@forlang\bbl@tempa{%
1955 \expandafter\bbl@stringdef
1956 \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
1957 <</Macros local to BabelCommands>>

```

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

```

1958 \newcommand\BabelLower[2]{% one to one.
1959 \ifnum\lccode#1=#2\else
1960 \babel@savevariable{\lccode#1}%
1961 \lccode#1=#2\relax
1962 \fi}
1963 \newcommand\BabelLowerMM[4]{% many-to-many
1964 \@tempcnta=#1\relax
1965 \@tempcntb=#4\relax
1966 \def\bbl@tempa{%

```

```

1967 \ifnum\@tempcnta>#2\else
1968 \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
1969 \advance\@tempcnta#3\relax
1970 \advance\@tempcntb#3\relax
1971 \expandafter\bb1@tempa
1972 \fi}%
1973 \bb1@tempa}
1974 \newcommand\BabelLowerM0[4]{% many-to-one
1975 \@tempcnta=#1\relax
1976 \def\bb1@tempa{%
1977 \ifnum\@tempcnta>#2\else
1978 \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
1979 \advance\@tempcnta#3
1980 \expandafter\bb1@tempa
1981 \fi}%
1982 \bb1@tempa}

```

The following package options control the behavior of hyphenation mapping.

```

1983 <{*More package options}> ≡
1984 \DeclareOption{hyphenmap=off}{\chardef\bb1@opt@hyphenmap\z@}
1985 \DeclareOption{hyphenmap=first}{\chardef\bb1@opt@hyphenmap\@ne}
1986 \DeclareOption{hyphenmap=select}{\chardef\bb1@opt@hyphenmap\tw@}
1987 \DeclareOption{hyphenmap=other}{\chardef\bb1@opt@hyphenmap\thr@@}
1988 \DeclareOption{hyphenmap=other*}{\chardef\bb1@opt@hyphenmap4\relax}
1989 <{/More package options}>

```

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

```

1990 \AtEndOfPackage{%
1991 \ifx\bb1@opt@hyphenmap\undefined
1992 \bb1@xin@{,}{\bb1@language@opts}%
1993 \chardef\bb1@opt@hyphenmap\ifin@4\else\@ne\fi
1994 \fi}

```

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

```

1995 \newcommand\setlocalecaption{% TODO. Catch typos. What about ensure?
1996 \@ifstar\bb1@setcaption@x\bb1@setcaption@x}
1997 \def\bb1@setcaption@x#1#2#3{% language caption-name string
1998 \bb1@trim@def\bb1@tempa{#2}%
1999 \bb1@xin@{.template}{\bb1@tempa}%
2000 \ifin@
2001 \bb1@ini@captions@template{#3}{#1}%
2002 \else
2003 \edef\bb1@tempd{%
2004 \expandafter\expandafter\expandafter
2005 \strip@prefix\expandafter\meaning\csname captions#1\endcsname}%
2006 \bb1@xin@
2007 {\expandafter\string\csname #2name\endcsname}%
2008 {\bb1@tempd}%
2009 \ifin@ % Renew caption
2010 \bb1@xin@{\string\bb1@scset}{\bb1@tempd}%
2011 \ifin@
2012 \bb1@exp{%
2013 \\\bb1@ifsamestring{\bb1@tempa}{\language@name}%
2014 {\\\bb1@scset\<#2name>\<#1#2name>}%
2015 {}}%
2016 \else % Old way converts to new way
2017 \bb1@ifunset{#1#2name}%
2018 {\bb1@exp{%
2019 \\\bb1@add\<captions#1>{\def\<#2name>{\<#1#2name>}}%
2020 \\\bb1@ifsamestring{\bb1@tempa}{\language@name}%
2021 {\def\<#2name>{\<#1#2name>}}%
2022 {}}}%

```

```

2023     {}%
2024     \fi
2025     \else
2026         \bbl@xin@{\string\bbl@scset}{\bbl@tempd}% New
2027         \ifin@ % New way
2028             \bbl@exp{%
2029                 \\bbl@add\<captions#1>{\bbl@scset\<#2name>\<#1#2name>}%
2030                 \\bbl@ifsamestring{\bbl@tempa}{\languagename}%
2031                 {\bbl@scset\<#2name>\<#1#2name>}%
2032                 {}}%
2033         \else % Old way, but defined in the new way
2034             \bbl@exp{%
2035                 \\bbl@add\<captions#1>{\def\<#2name>\<#1#2name>}}%
2036                 \\bbl@ifsamestring{\bbl@tempa}{\languagename}%
2037                 {\def\<#2name>\<#1#2name>}}%
2038                 {}}%
2039         \fi%
2040     \fi
2041     \@namedef{#1#2name}{#3}%
2042     \toks@\expandafter{\bbl@captionslist}%
2043     \bbl@exp{\in@{\<#2name>}{\the\toks@}}%
2044     \ifin@\else
2045         \bbl@exp{\bbl@add\bbl@captionslist{\<#2name>}}%
2046         \bbl@tglobal\bbl@captionslist
2047     \fi
2048     \fi}
2049 % \def\bbl@setcaption@s#1#2#3{} % TODO. Not yet implemented

```

## 7.11 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.

```

2050 \bbl@trace{Macros related to glyphs}
2051 \def\set@low@box#1{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}%
2052     \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@%
2053     \setbox\z@\hbox{\lower\dimen\z@ \box\z@\ht\z@\ht\tw@ \dp\z@\dp\tw@}

```

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

```

2054 \def\save@s@q#1{\leavevmode
2055     \begingroup
2056     \edef\@SF{\spacefactor\the\spacefactor}#1\@SF
2057     \endgroup}

```

## 7.12 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.

### 7.12.1 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.

```

2058 \ProvideTextCommand{\quotedblbase}{OT1}{%
2059     \save@s@q{\set@low@box{\textquotedblright\}%
2060     \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.

```

2061 \ProvideTextCommandDefault{\quotedblbase}{%
2062     \UseTextSymbol{OT1}{\quotedblbase}}

```

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

```
2063 \ProvideTextCommand{\quotesinglbase}{OT1}{%
2064   \save@sf@q{\set@low@box{\textquoteright\}%
2065     \box\z@\kern-.04em\bb1@allowhyphens}}
```

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

```
2066 \ProvideTextCommandDefault{\quotesinglbase}{%
2067   \UseTextSymbol{OT1}{\quotesinglbase}}
```

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

```
2068 \ProvideTextCommand{\guillemetleft}{OT1}{%
2069   \ifmmode
2070     \ll
2071   \else
2072     \save@sf@q{\nobreak
2073       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bb1@allowhyphens}%
2074   \fi}
2075 \ProvideTextCommand{\guillemetright}{OT1}{%
2076   \ifmmode
2077     \gg
2078   \else
2079     \save@sf@q{\nobreak
2080       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bb1@allowhyphens}%
2081   \fi}
2082 \ProvideTextCommand{\guillemotleft}{OT1}{%
2083   \ifmmode
2084     \ll
2085   \else
2086     \save@sf@q{\nobreak
2087       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bb1@allowhyphens}%
2088   \fi}
2089 \ProvideTextCommand{\guillemotright}{OT1}{%
2090   \ifmmode
2091     \gg
2092   \else
2093     \save@sf@q{\nobreak
2094       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bb1@allowhyphens}%
2095   \fi}
```

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

```
2096 \ProvideTextCommandDefault{\guillemetleft}{%
2097   \UseTextSymbol{OT1}{\guillemetleft}}
2098 \ProvideTextCommandDefault{\guillemetright}{%
2099   \UseTextSymbol{OT1}{\guillemetright}}
2100 \ProvideTextCommandDefault{\guillemotleft}{%
2101   \UseTextSymbol{OT1}{\guillemotleft}}
2102 \ProvideTextCommandDefault{\guillemotright}{%
2103   \UseTextSymbol{OT1}{\guillemotright}}
```

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

```
2104 \ProvideTextCommand{\guilsinglleft}{OT1}{%
2105   \ifmmode
2106     <%
2107   \else
2108     \save@sf@q{\nobreak
2109       \raise.2ex\hbox{$\scriptscriptstyle<$}\bb1@allowhyphens}%
2110   \fi}
2111 \ProvideTextCommand{\guilsinglright}{OT1}{%
2112   \ifmmode
2113     >%
2114   \else
2115     \save@sf@q{\nobreak
```

```

2116      \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
2117 \fi}

```

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

```

2118 \ProvideTextCommandDefault{\guilsinglleft}{%
2119   \UseTextSymbol{OT1}{\guilsinglleft}}
2120 \ProvideTextCommandDefault{\guilsinglright}{%
2121   \UseTextSymbol{OT1}{\guilsinglright}}

```

### 7.12.2 Letters

`\ij` The dutch language uses the letter ‘ij’. It is available in T1 encoded fonts, but not in the OT1 encoded fonts. Therefore we fake it for the OT1 encoding.

```

2122 \DeclareTextCommand{\ij}{OT1}{%
2123   i\kern-0.02em\bbl@allowhyphens j}
2124 \DeclareTextCommand{\IJ}{OT1}{%
2125   I\kern-0.02em\bbl@allowhyphens J}
2126 \DeclareTextCommand{\ij}{T1}{\char188}
2127 \DeclareTextCommand{\IJ}{T1}{\char156}

```

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

```

2128 \ProvideTextCommandDefault{\ij}{%
2129   \UseTextSymbol{OT1}{\ij}}
2130 \ProvideTextCommandDefault{\IJ}{%
2131   \UseTextSymbol{OT1}{\IJ}}

```

`\dj` The croatian language needs the letters `\dj` and `\DJ`; they are available in the T1 encoding, but not in the OT1 encoding by default.

Some code to construct these glyphs for the OT1 encoding was made available to me by Stipčević Mario, (stipcevic@olimp.irb.hr).

```

2132 \def\crrtic@{\hrule height0.1ex width0.3em}
2133 \def\crttic@{\hrule height0.1ex width0.33em}
2134 \def\ddj@{%
2135   \setbox0\hbox{d}\dimen@=\ht0
2136   \advance\dimen@1ex
2137   \dimen@.45\dimen@
2138   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
2139   \advance\dimen@ii.5ex
2140   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
2141 \def\DDJ@{%
2142   \setbox0\hbox{D}\dimen@=.55\ht0
2143   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
2144   \advance\dimen@ii.15ex % correction for the dash position
2145   \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
2146   \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
2147   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
2148 %
2149 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
2150 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}

```

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

```

2151 \ProvideTextCommandDefault{\dj}{%
2152   \UseTextSymbol{OT1}{\dj}}
2153 \ProvideTextCommandDefault{\DJ}{%
2154   \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.

```

2155 \DeclareTextCommand{\SS}{OT1}{\SS}
2156 \ProvideTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}

```

### 7.12.3 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
2157 \ProvideTextCommandDefault{\glq}{%
2158   \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}

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

2159 \ProvideTextCommand{\grq}{T1}{%
2160   \textormath{\kern\z@\textquoteleft}{\mbox{\textquoteleft}}}
2161 \ProvideTextCommand{\grq}{TU}{%
2162   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
2163 \ProvideTextCommand{\grq}{OT1}{%
2164   \save@sf@q{\kern-.0125em
2165     \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
2166     \kern.07em\relax}}
2167 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}
```

`\glqq` The ‘german’ double quotes.

```
\grqq
2168 \ProvideTextCommandDefault{\glqq}{%
2169   \textormath{\quotedblbase}{\mbox{\quotedblbase}}}

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

2170 \ProvideTextCommand{\grqq}{T1}{%
2171   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
2172 \ProvideTextCommand{\grqq}{TU}{%
2173   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
2174 \ProvideTextCommand{\grqq}{OT1}{%
2175   \save@sf@q{\kern-.07em
2176     \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
2177     \kern.07em\relax}}
2178 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}
```

`\flq` The ‘french’ single guillemets.

```
\frq
2179 \ProvideTextCommandDefault{\flq}{%
2180   \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
2181 \ProvideTextCommandDefault{\frq}{%
2182   \textormath{\guilsinglright}{\mbox{\guilsinglright}}}
```

`\flqq` The ‘french’ double guillemets.

```
\frqq
2183 \ProvideTextCommandDefault{\flqq}{%
2184   \textormath{\guillemetleft}{\mbox{\guillemetleft}}}
2185 \ProvideTextCommandDefault{\frqq}{%
2186   \textormath{\guillemetright}{\mbox{\guillemetright}}}
```

### 7.12.4 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).

```
2187 \def\umlauthigh{%
2188   \def\bbl@umlauta##1{\leavevmode\bgroup%
2189     \expandafter\accent\csname\f@encoding dqpos\endcsname
2190     ##1\bbl@allowhyphens\egroup}%
2191   \let\bbl@umlaute\bbl@umlauta}
2192 \def\umlautlow{%
```

```

2193 \def\bbl@umlauta{\protect\lower@umlaut}}
2194 \def\umlautelow{%
2195 \def\bbl@umlaute{\protect\lower@umlaut}}
2196 \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.

```

2197 \expandafter\ifx\csname U@D\endcsname\relax
2198 \csname newdimen\endcsname\U@D
2199 \fi

```

The following code fools  $\TeX$ 's `make_accent` procedure about the current x-height of the font to force another placement of the umlaut character. First we have to save the current x-height of the font, because we'll change this font dimension and this is always done globally.

Then we compute the new x-height in such a way that the umlaut character is lowered to the base character. The value of `.45ex` depends on the METAFONT parameters with which the fonts were built. (Just try out, which value will look best.) If the new x-height is too low, it is not changed. Finally we call the `\accent` primitive, reset the old x-height and insert the base character in the argument.

```

2200 \def\lower@umlaut#1{%
2201 \leavevmode\bgroup
2202 \U@D 1ex%
2203 {\setbox\z@\hbox{%
2204 \expandafter\char\csname f@encoding dqpos\endcsname}%
2205 \dimen@ -.45ex\advance\dimen@\ht\z@
2206 \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
2207 \expandafter\accent\csname f@encoding dqpos\endcsname
2208 \fontdimen5\font\U@D #1%
2209 \egroup}

```

For all vowels we declare `\` to be a composite command which uses `\bbl@umlauta` or `\bbl@umlaute` to position the umlaut character. We need to be sure that these definitions override the ones that are provided when the package `fontenc` with option `OT1` is used. Therefore these declarations are postponed until the beginning of the document. Note these definitions only apply to some languages, but `babel` sets them for *all* languages – you may want to redefine `\bbl@umlauta` and/or `\bbl@umlaute` for a language in the corresponding `ldf` (using the `babel` switching mechanism, of course).

```

2210 \AtBeginDocument{%
2211 \DeclareTextCompositeCommand{\}{OT1}{a}{\bbl@umlauta{a}}%
2212 \DeclareTextCompositeCommand{\}{OT1}{e}{\bbl@umlaute{e}}%
2213 \DeclareTextCompositeCommand{\}{OT1}{i}{\bbl@umlaute{i}}%
2214 \DeclareTextCompositeCommand{\}{OT1}{\i}{\bbl@umlaute{i}}%
2215 \DeclareTextCompositeCommand{\}{OT1}{o}{\bbl@umlauta{o}}%
2216 \DeclareTextCompositeCommand{\}{OT1}{u}{\bbl@umlauta{u}}%
2217 \DeclareTextCompositeCommand{\}{OT1}{A}{\bbl@umlauta{A}}%
2218 \DeclareTextCompositeCommand{\}{OT1}{E}{\bbl@umlaute{E}}%
2219 \DeclareTextCompositeCommand{\}{OT1}{I}{\bbl@umlaute{I}}%
2220 \DeclareTextCompositeCommand{\}{OT1}{O}{\bbl@umlauta{O}}%
2221 \DeclareTextCompositeCommand{\}{OT1}{U}{\bbl@umlauta{U}}%

```

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

```

2222 \ifx\l@english\undefined
2223 \chardef\l@english\z@
2224 \fi
2225 % The following is used to cancel rules in ini files (see Amharic).
2226 \ifx\l@unhyphenated\undefined
2227 \newlanguage\l@unhyphenated
2228 \fi

```

## 7.13 Layout

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

```

2229 \bbl@trace{Bidi layout}
2230 \providecommand\IfBabelLayout[3]{#3}%
2231 \newcommand\BabelPatchSection[1]{%
2232   \@ifundefined{#1}{}{%
2233     \bbl@exp{\let\<bbl@ss@#1>\<#1>}%
2234     \@namedef{#1}{%
2235       \ifstar{\bbl@presec@s{#1}}%
2236       {\@dblarg{\bbl@presec@x{#1}}}}}%
2237 \def\bbl@presec@x#1[#2]#3{%
2238   \bbl@exp{%
2239     \\\select@language@x{\bbl@main@language}%
2240     \\\bbl@cs{sspre@#1}%
2241     \\\bbl@cs{ss@#1}%
2242     [\\foreignlanguage{\language}{\unexpanded{#2}}}%
2243     {\\foreignlanguage{\language}{\unexpanded{#3}}}%
2244     \\\select@language@x{\language}}}%
2245 \def\bbl@presec@s#1#2{%
2246   \bbl@exp{%
2247     \\\select@language@x{\bbl@main@language}%
2248     \\\bbl@cs{sspre@#1}%
2249     \\\bbl@cs{ss@#1}*%
2250     {\\foreignlanguage{\language}{\unexpanded{#2}}}%
2251     \\\select@language@x{\language}}}%
2252 \IfBabelLayout{sectioning}%
2253   {\BabelPatchSection{part}%
2254    \BabelPatchSection{chapter}%
2255    \BabelPatchSection{section}%
2256    \BabelPatchSection{subsection}%
2257    \BabelPatchSection{subsubsection}%
2258    \BabelPatchSection{paragraph}%
2259    \BabelPatchSection{subparagraph}%
2260    \def\babel@toc#1{%
2261      \select@language@x{\bbl@main@language}}}%
2262 \IfBabelLayout{captions}%
2263   {\BabelPatchSection{caption}}}%

```

## 7.14 Load engine specific macros

```

2264 \bbl@trace{Input engine specific macros}
2265 \ifcase\bbl@engine
2266   \input txtbabel.def
2267 \or
2268   \input luababel.def
2269 \or
2270   \input xebabel.def
2271 \fi

```

## 7.15 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.

```

2272 \bbl@trace{Creating languages and reading ini files}
2273 \let\bbl@extend@ini@gobble
2274 \newcommand\babelprovide[2][{}]{%
2275   \let\bbl@savelangname\language
2276   \edef\bbl@savelocaleid{\the\localeid}%
2277   % Set name and locale id
2278   \edef\language{#2}%
2279   \bbl@id@assign
2280   % Initialize keys
2281   \let\bbl@KVP@captions\@nil
2282   \let\bbl@KVP@date\@nil

```



```

2283 \let\bbl@KVP@import\@nil
2284 \let\bbl@KVP@main\@nil
2285 \let\bbl@KVP@script\@nil
2286 \let\bbl@KVP@language\@nil
2287 \let\bbl@KVP@hyphenrules\@nil
2288 \let\bbl@KVP@linebreaking\@nil
2289 \let\bbl@KVP@justification\@nil
2290 \let\bbl@KVP@mapfont\@nil
2291 \let\bbl@KVP@maparabic\@nil
2292 \let\bbl@KVP@mapdigits\@nil
2293 \let\bbl@KVP@intraspace\@nil
2294 \let\bbl@KVP@intrapenalty\@nil
2295 \let\bbl@KVP@onchar\@nil
2296 \let\bbl@KVP@transforms\@nil
2297 \global\let\bbl@release@transforms\@empty
2298 \let\bbl@KVP@alph\@nil
2299 \let\bbl@KVP@Alph\@nil
2300 \let\bbl@KVP@labels\@nil
2301 \bbl@csarg\let{KVP@labels*}\@nil
2302 \global\let\bbl@inidata\@empty
2303 \global\let\bbl@extend@ini\@gobble
2304 \gdef\bbl@key@list{;}%
2305 \bbl@forkv{#1}{% TODO - error handling
2306   \in{/}{##1}%
2307   \ifin@
2308     \global\let\bbl@extend@ini\bbl@extend@ini@aux
2309     \bbl@renewinikey##1\@{##2}%
2310   \else
2311     \bbl@csarg\def{KVP@##1}{##2}%
2312   \fi}%
2313 \chardef\bbl@howloaded=% 0:none; 1:ldf without ini; 2:ini
2314 \bbl@ifunset{date#2}\z{\bbl@ifunset{\bbl@llevel@#2}\@ne\tw@}%
2315 % == init ==
2316 \ifx\bbl@screset\@undefined
2317   \bbl@ldfinit
2318 \fi
2319 % ==
2320 \let\bbl@lbkflag\relax % \@empty = do setup linebreak
2321 \ifcase\bbl@howloaded
2322   \let\bbl@lbkflag\@empty % new
2323 \else
2324   \ifx\bbl@KVP@hyphenrules\@nil\else
2325     \let\bbl@lbkflag\@empty
2326   \fi
2327   \ifx\bbl@KVP@import\@nil\else
2328     \let\bbl@lbkflag\@empty
2329   \fi
2330 \fi
2331 % == import, captions ==
2332 \ifx\bbl@KVP@import\@nil\else
2333   \bbl@exp{\bbl@ifblank{\bbl@KVP@import}}%
2334   {\ifx\bbl@initload\relax
2335     \begingroup
2336       \def\BabelBeforeIni##1##2{\gdef\bbl@KVP@import{##1}\endinput}%
2337       \bbl@input@texini{#2}%
2338     \endgroup
2339   \else
2340     \xdef\bbl@KVP@import{\bbl@initload}%
2341   \fi}%
2342 {}%
2343 \fi
2344 \ifx\bbl@KVP@captions\@nil
2345   \let\bbl@KVP@captions\bbl@KVP@import

```

```

2346 \fi
2347 % ==
2348 \ifx\bbbl@KVP@transforms\@nil\else
2349 \bbbl@replace\bbbl@KVP@transforms{ },}%
2350 \fi
2351 % == Load ini ==
2352 \ifcase\bbbl@howloaded
2353 \bbbl@provide@new{#2}%
2354 \else
2355 \bbbl@ifblank{#1}%
2356 { }% With \bbbl@load@basic below
2357 { \bbbl@provide@renew{#2} }%
2358 \fi
2359 % Post tasks
2360 % -----
2361 % == subsequent calls after the first provide for a locale ==
2362 \ifx\bbbl@inidata\@empty\else
2363 \bbbl@extend@ini{#2}%
2364 \fi
2365 % == ensure captions ==
2366 \ifx\bbbl@KVP@captions\@nil\else
2367 \bbbl@ifunset{bbbl@extracaps@#2}%
2368 { \bbbl@exp{\\babelensure[exclude=\\today]{#2}}}%
2369 { \bbbl@exp{\\babelensure[exclude=\\today,
2370 include=\[bbbl@extracaps@#2]]{#2}}}%
2371 \bbbl@ifunset{bbbl@ensure@\language}%
2372 { \bbbl@exp{%
2373 \\DeclareRobustCommand\<bbbl@ensure@\language>[1]{%
2374 \\foreignlanguage{\language}%
2375 {###1}}}%
2376 } }%
2377 \bbbl@exp{%
2378 \\bbbl@toglobal\<bbbl@ensure@\language>%
2379 \\bbbl@toglobal\<bbbl@ensure@\language\space>}%
2380 \fi
2381 % ==
2382 % At this point all parameters are defined if 'import'. Now we
2383 % execute some code depending on them. But what about if nothing was
2384 % imported? We just set the basic parameters, but still loading the
2385 % whole ini file.
2386 \bbbl@load@basic{#2}%
2387 % == script, language ==
2388 % Override the values from ini or defines them
2389 \ifx\bbbl@KVP@script\@nil\else
2390 \bbbl@csarg\edef{sname@#2}{\bbbl@KVP@script}%
2391 \fi
2392 \ifx\bbbl@KVP@language\@nil\else
2393 \bbbl@csarg\edef{lname@#2}{\bbbl@KVP@language}%
2394 \fi
2395 \ifcase\bbbl@engine\or
2396 \bbbl@ifunset{bbbl@chrng@\language}{ }%
2397 { \directlua{
2398 Babel.set_chranges_b('\bbbl@cl{sbc}', '\bbbl@cl{chrng}') }}%
2399 \fi
2400 % == onchar ==
2401 \ifx\bbbl@KVP@onchar\@nil\else
2402 \bbbl@luahyphenate
2403 \bbbl@exp{%
2404 \\AddToHook{env/document/before}{\\select@language{#2}}}%
2405 \directlua{
2406 if Babel.locale_mapped == nil then
2407 Babel.locale_mapped = true
2408 Babel.linebreaking.add_before(Babel.locale_map)

```

```

2409     Babel.loc_to_scr = {}
2410     Babel.chr_to_loc = Babel.chr_to_loc or {}
2411 end}%
2412 \bbl@xin@{ ids }{ \bbl@KVP@onchar\space}%
2413 \ifin@
2414     \ifx\bbl@starthyphens\undefined % Needed if no explicit selection
2415         \AddBabelHook{babel-onchar}{beforestart}{\bbl@starthyphens}%
2416     \fi
2417     \bbl@exp{\bbl@add\bbl@starthyphens
2418         {\bbl@patterns@lua{\language}\space}}%
2419 % TODO - error/warning if no script
2420 \directlua{
2421     if Babel.script_blocks['\bbl@cl{sbc}'] then
2422         Babel.loc_to_scr[\the\localeid] =
2423             Babel.script_blocks['\bbl@cl{sbc}']
2424         Babel.locale_props[\the\localeid].lc = \the\localeid\space
2425         Babel.locale_props[\the\localeid].lg = \the\nameuse{1@\language}\space
2426     end
2427 }%
2428 \fi
2429 \bbl@xin@{ fonts }{ \bbl@KVP@onchar\space}%
2430 \ifin@
2431     \bbl@ifunset{bbl@lsys@\language}{\bbl@provide@lsys{\language}}{}%
2432     \bbl@ifunset{bbl@wdir@\language}{\bbl@provide@dirs{\language}}{}%
2433 \directlua{
2434     if Babel.script_blocks['\bbl@cl{sbc}'] then
2435         Babel.loc_to_scr[\the\localeid] =
2436             Babel.script_blocks['\bbl@cl{sbc}']
2437     end}%
2438 \ifx\bbl@mapselect\undefined % TODO. almost the same as mapfont
2439     \AtBeginDocument{%
2440         \bbl@patchfont{\bbl@mapselect}%
2441         {\selectfont}}%
2442     \def\bbl@mapselect{%
2443         \let\bbl@mapselect\relax
2444         \edef\bbl@prefontid{\fontid\font}}%
2445     \def\bbl@mapdir##1{%
2446         {\def\language{##1}%
2447         \let\bbl@ifrestoring\@firstoftwo % To avoid font warning
2448         \bbl@switchfont
2449         \ifnum\fontid\font>\z@ % A hack, for the pgf nullfont hack
2450             \directlua{
2451                 Babel.locale_props[\the\csname bbl@id@##1\endcsname]%
2452                     [\bbl@prefontid] = \fontid\font\space}%
2453             \fi}}%
2454     \fi
2455     \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\language}}}%
2456 \fi
2457 % TODO - catch non-valid values
2458 \fi
2459 % == mapfont ==
2460 % For bidi texts, to switch the font based on direction
2461 \ifx\bbl@KVP@mapfont\@nil\else
2462     \bbl@ifsamestring{\bbl@KVP@mapfont}{direction}{}%
2463     {\bbl@error{Option '\bbl@KVP@mapfont' unknown for\%
2464         mapfont. Use 'direction'.%
2465         {See the manual for details.}}}%
2466     \bbl@ifunset{bbl@lsys@\language}{\bbl@provide@lsys{\language}}{}%
2467     \bbl@ifunset{bbl@wdir@\language}{\bbl@provide@dirs{\language}}{}%
2468 \ifx\bbl@mapselect\undefined % TODO. See onchar.
2469     \AtBeginDocument{%
2470         \bbl@patchfont{\bbl@mapselect}%
2471         {\selectfont}}%

```

```

2472 \def\bbl@mapselect{%
2473 \let\bbl@mapselect\relax
2474 \edef\bbl@prefontid{\fontid\font}}%
2475 \def\bbl@mapdir##1{%
2476 {\def\language{##1}%
2477 \let\bbl@ifrestoring\@firstoftwo % avoid font warning
2478 \bbl@switchfont
2479 \directlua{Babel.fontmap
2480 [\the\csname bbl@wdir@##1\endcsname]%
2481 [\bbl@prefontid]=\fontid\font}}}%
2482 \fi
2483 \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir\language}}}%
2484 \fi
2485 % == Line breaking: intraspace, intrapenalty ==
2486 % For CJK, East Asian, Southeast Asian, if interspace in ini
2487 \ifx\bbl@KVP@intraspace\@nil\else % We can override the ini or set
2488 \bbl@csarg\edef{intsp@#2}{\bbl@KVP@intraspace}%
2489 \fi
2490 \bbl@provide@intraspace
2491 % == Line breaking: CJK quotes ==
2492 \ifcase\bbl@engine\or
2493 \bbl@xin@{/c}{\bbl@cl{lnbrk}}}%
2494 \ifin@
2495 \bbl@ifunset{bbl@quote@language}{}%
2496 {\directlua{
2497 Babel.locale_props[\the\localeid].cjk_quotes = {}
2498 local cs = 'op'
2499 for c in string.utfvalues(
2500 [[\csname bbl@quote@language\endcsname]]) do
2501 if Babel.cjk_characters[c].c == 'qu' then
2502 Babel.locale_props[\the\localeid].cjk_quotes[c] = cs
2503 end
2504 cs = (cs == 'op') and 'cl' or 'op'
2505 end
2506 }}%
2507 \fi
2508 \fi
2509 % == Line breaking: justification ==
2510 \ifx\bbl@KVP@justification\@nil\else
2511 \let\bbl@KVP@linebreaking\bbl@KVP@justification
2512 \fi
2513 \ifx\bbl@KVP@linebreaking\@nil\else
2514 \bbl@xin@{,\bbl@KVP@linebreaking,}{,elongated,kashida,cjk,unhyphenated,}%
2515 \ifin@
2516 \bbl@csarg\xdef
2517 {lnbrk@language}{\expandafter\@car\bbl@KVP@linebreaking\@nil}%
2518 \fi
2519 \fi
2520 \bbl@xin@{/e}{\bbl@cl{lnbrk}}}%
2521 \ifin@else\bbl@xin@{/k}{\bbl@cl{lnbrk}}\fi
2522 \ifin@\bbl@arabicjust\fi
2523 % == Line breaking: hyphenate.other.(locale|script) ==
2524 \ifx\bbl@lbkflag\@empty
2525 \bbl@ifunset{bbl@hyotl@language}{}%
2526 {\bbl@csarg\bbl@replace{hyotl@language}{,},}%
2527 \bbl@startcommands*\language}%
2528 \bbl@csarg\bbl@foreach{hyotl@language}{%
2529 \ifcase\bbl@engine
2530 \ifnum##1<257
2531 \SetHyphenMap{\BabelLower{##1}{##1}}%
2532 \fi
2533 \else
2534 \SetHyphenMap{\BabelLower{##1}{##1}}%

```

```

2535         \fi}%
2536     \bbl@endcommands}%
2537 \bbl@ifunset{\bbl@hyots@\language}\{ }%
2538 {\bbl@csarg\bbl@replace{\hyots@\language}{ }{ },}%
2539 \bbl@csarg\bbl@foreach{\hyots@\language}{%
2540     \ifcase\bbl@engine
2541     \ifnum##1<257
2542         \global\lccode##1=##1\relax
2543     \fi
2544     \else
2545         \global\lccode##1=##1\relax
2546     \fi}}%
2547 \fi
2548 % == Counters: maparabic ==
2549 % Native digits, if provided in ini (TeX level, xe and lua)
2550 \ifcase\bbl@engine\else
2551     \bbl@ifunset{\bbl@dgnat@\language}\{ }%
2552     {\expandafter\ifx\csname bbl@dgnat@\language\endcsname\@empty\else
2553         \expandafter\expandafter\expandafter
2554         \bbl@setdigits\csname bbl@dgnat@\language\endcsname
2555         \ifx\bbl@KVP@maparabic\@nil\else
2556             \ifx\bbl@latinarabic\@undefined
2557                 \expandafter\let\expandafter\@arabic
2558                 \csname bbl@counter@\language\endcsname
2559             \else % ie, if layout=counters, which redefines \@arabic
2560                 \expandafter\let\expandafter\bbl@latinarabic
2561                 \csname bbl@counter@\language\endcsname
2562             \fi
2563         \fi
2564     \fi}%
2565 \fi
2566 % == Counters: mapdigits ==
2567 % Native digits (lua level).
2568 \ifodd\bbl@engine
2569     \ifx\bbl@KVP@mapdigits\@nil\else
2570         \bbl@ifunset{\bbl@dgnat@\language}\{ }%
2571         {\RequirePackage{luatexbase}%
2572         \bbl@activate@preotf
2573         \directlua{
2574             Babel = Babel or {} %%% -> presets in luababel
2575             Babel.digits_mapped = true
2576             Babel.digits = Babel.digits or {}
2577             Babel.digits[\the\localeid] =
2578                 table.pack(string.utfvalue('\bbl@cl{dgnat}'))
2579             if not Babel.numbers then
2580                 function Babel.numbers(head)
2581                     local LOCALE = Babel.attr_locale
2582                     local GLYPH = node.id'glyph'
2583                     local inmath = false
2584                     for item in node.traverse(head) do
2585                         if not inmath and item.id == GLYPH then
2586                             local temp = node.get_attribute(item, LOCALE)
2587                             if Babel.digits[temp] then
2588                                 local chr = item.char
2589                                 if chr > 47 and chr < 58 then
2590                                     item.char = Babel.digits[temp][chr-47]
2591                                 end
2592                             end
2593                         elseif item.id == node.id'math' then
2594                             inmath = (item.subtype == 0)
2595                         end
2596                     end
2597                     return head

```

```

2598         end
2599     end
2600 }}%
2601 \fi
2602 \fi
2603 % == Counters: alph, Alph ==
2604 % What if extras<lang> contains a \babel@save\@alph? It won't be
2605 % restored correctly when exiting the language, so we ignore
2606 % this change with the \bbl@alph@saved trick.
2607 \ifx\bbl@KVP@alph\@nil\else
2608     \bbl@extras@wrap{\bbl@alph@saved}%
2609     {\let\bbl@alph@saved\@alph}%
2610     {\let\@alph\bbl@alph@saved
2611      \babel@save\@alph}%
2612     \bbl@exp{%
2613         \bbl@add\<extras\language>%
2614         \let\@alph\<bbl@cntr\bbl@KVP@alph @\language>}}%
2615 \fi
2616 \ifx\bbl@KVP@Alph\@nil\else
2617     \bbl@extras@wrap{\bbl@Alph@saved}%
2618     {\let\bbl@Alph@saved\@Alph}%
2619     {\let\@Alph\bbl@Alph@saved
2620      \babel@save\@Alph}%
2621     \bbl@exp{%
2622         \bbl@add\<extras\language>%
2623         \let\@Alph\<bbl@cntr\bbl@KVP@Alph @\language>}}%
2624 \fi
2625 % == require.babel in ini ==
2626 % To load or reload the babel-*.tex, if require.babel in ini
2627 \ifx\bbl@beforestart\relax\else % But not in doc aux or body
2628     \bbl@ifunset\bbl@rqtex@\language{}{%
2629         {\expandafter\ifx\csname bbl@rqtex@\language\endcsname\@empty\else
2630             \let\BabelBeforeIni\@gobbletwo
2631             \chardef\atcatcode=\catcode`\@
2632             \catcode`\@=11\relax
2633             \bbl@input@texini{\bbl@cs{rqtex@\language}}%
2634             \catcode`\@=\atcatcode
2635             \let\atcatcode\relax
2636             \global\bbl@csarg\let{rqtex@\language}\relax
2637         \fi}%
2638 \fi
2639 % == frenchspacing ==
2640 \ifcase\bbl@howloaded\in@true\else\in@false\fi
2641 \ifin@else\bbl@xin@{typography/frenchspacing}{\bbl@key@list}\fi
2642 \ifin@
2643     \bbl@extras@wrap{\bbl@pre@fs}%
2644     {\bbl@pre@fs}%
2645     {\bbl@post@fs}%
2646 \fi
2647 % == Release saved transforms ==
2648 \bbl@release@transforms\relax % \relax closes the last item.
2649 % == main ==
2650 \ifx\bbl@KVP@main\@nil % Restore only if not 'main'
2651     \let\language\bbl@savelangname
2652     \chardef\localeid\bbl@savelocaleid\relax
2653 \fi}

```

Depending on whether or not the language exists (based on \date<language>), we define two macros. Remember \bbl@startcommands opens a group.

```

2654 \def\bbl@provide@new#1{%
2655     \@namedef{date#1}{}% marks lang exists - required by \StartBabelCommands
2656     \@namedef{extras#1}{}%
2657     \@namedef{noextras#1}{}%

```

```

2658 \bbl@startcommands*{#1}{captions}%
2659 \ifx\bbl@KVP@captions\@nil % and also if import, implicit
2660 \def\bbl@tempb##1{% elt for \bbl@captionslist
2661 \ifx##1\@empty\else
2662 \bbl@exp{%
2663 \\\SetString\\##1{%
2664 \\\bbl@nocaption{\bbl@stripslash##1}{#1\bbl@stripslash##1}}}%
2665 \expandafter\bbl@tempb
2666 \fi}%
2667 \expandafter\bbl@tempb\bbl@captionslist\@empty
2668 \else
2669 \ifx\bbl@initoload\relax
2670 \bbl@read@ini{\bbl@KVP@captions}2% % Here letters cat = 11
2671 \else
2672 \bbl@read@ini{\bbl@initoload}2% % Same
2673 \fi
2674 \fi
2675 \StartBabelCommands*{#1}{date}%
2676 \ifx\bbl@KVP@import\@nil
2677 \bbl@exp{%
2678 \\\SetString\\today{\bbl@nocaption{today}{#1today}}}%
2679 \else
2680 \bbl@savetoday
2681 \bbl@savestate
2682 \fi
2683 \bbl@endcommands
2684 \bbl@load@basic{#1}%
2685 % == hyphenmins == (only if new)
2686 \bbl@exp{%
2687 \gdef\<#1hyphenmins>{%
2688 {\bbl@ifunset{\bbl@lfthm@#1}{2}{\bbl@cs{lfthm@#1}}}%
2689 {\bbl@ifunset{\bbl@rgthm@#1}{3}{\bbl@cs{rgthm@#1}}}}}%
2690 % == hyphenrules (also in renew) ==
2691 \bbl@provide@hyphens{#1}%
2692 \ifx\bbl@KVP@main\@nil\else
2693 \expandafter\main@language\expandafter{#1}%
2694 \fi}
2695 %
2696 \def\bbl@provide@renew#1{%
2697 \ifx\bbl@KVP@captions\@nil\else
2698 \StartBabelCommands*{#1}{captions}%
2699 \bbl@read@ini{\bbl@KVP@captions}2% % Here all letters cat = 11
2700 \EndBabelCommands
2701 \fi
2702 \ifx\bbl@KVP@import\@nil\else
2703 \StartBabelCommands*{#1}{date}%
2704 \bbl@savetoday
2705 \bbl@savestate
2706 \EndBabelCommands
2707 \fi
2708 % == hyphenrules (also in new) ==
2709 \ifx\bbl@lbfkflag\@empty
2710 \bbl@provide@hyphens{#1}%
2711 \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.)

```

2712 \def\bbl@load@basic#1{%
2713 \ifcase\bbl@howloaded\or\or
2714 \ifcase\csname bbl@llevel@\language\endcsname
2715 \bbl@csarg\let{lname@\language}\relax
2716 \fi

```

```

2717 \fi
2718 \bbl@ifunset{bbl@lname@#1}%
2719 {\def\BabelBeforeIni##1##2{%
2720   \begingroup
2721     \let\bbl@ini@captions@aux\@gobbletwo
2722     \def\bbl@inidate ####1.####2.####3.####4\relax ####5####6{%
2723       \bbl@read@ini{##1}1%
2724       \ifx\bbl@initoload\relax\endinput\fi
2725     \endgroup}%
2726   \begingroup % boxed, to avoid extra spaces:
2727     \ifx\bbl@initoload\relax
2728       \bbl@input@texini{##1}%
2729     \else
2730       \setbox\z@\hbox{\BabelBeforeIni{\bbl@initoload}{}}%
2731     \fi
2732   \endgroup}%
2733 }}}
```

The hyphenrules option is handled with an auxiliary macro.

```

2734 \def\bbl@provide@hyphens#1{%
2735   \let\bbl@tempa\relax
2736   \ifx\bbl@KVP@hyphenrules\@nil\else
2737     \bbl@replace\bbl@KVP@hyphenrules{ }{,}%
2738     \bbl@foreach\bbl@KVP@hyphenrules{%
2739       \ifx\bbl@tempa\relax % if not yet found
2740         \bbl@ifsamestring{##1}{+}%
2741         {\bbl@exp{\addlanguage\<l@##1>}}}%
2742       {}%
2743       \bbl@ifunset{l@##1}%
2744       {}%
2745       {\bbl@exp{\let\bbl@tempa\<l@##1>}}}%
2746     \fi}%
2747 \fi
2748 \ifx\bbl@tempa\relax % if no opt or no language in opt found
2749   \ifx\bbl@KVP@import\@nil
2750     \ifx\bbl@initoload\relax\else
2751       \bbl@exp{% and hyphenrules is not empty
2752         \\bbl@ifblank{\bbl@cs{hyphr@#1}}%
2753         {}%
2754         {\let\\bbl@tempa\<l@bbl@cl{hyphr}>}}%
2755     \fi
2756   \else % if importing
2757     \bbl@exp{% and hyphenrules is not empty
2758       \\bbl@ifblank{\bbl@cs{hyphr@#1}}%
2759       {}%
2760       {\let\\bbl@tempa\<l@bbl@cl{hyphr}>}}%
2761   \fi
2762 \fi
2763 \bbl@ifunset{bbl@tempa}% ie, relax or undefined
2764 {\bbl@ifunset{l@#1}% no hyphenrules found - fallback
2765   {\bbl@exp{\adddialect\<l@#1>\language}}%
2766   {}}% so, l@<lang> is ok - nothing to do
2767 {\bbl@exp{\adddialect\<l@#1>\bbl@tempa}}% found in opt list or ini
```

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

```

2768 \def\bbl@input@texini#1{%
2769   \bbl@bspack
2770   \bbl@exp{%
2771     \catcode`\\%=14 \catcode`\\=\0
2772     \catcode`\\={1 \catcode`\\}=2
2773     \lowercase{\InputIfFileExists{babel-#1.tex}{}}%
2774     \catcode`\\%=\the\catcode`\%\relax
2775     \catcode`\\=\the\catcode`\%\relax
2776     \catcode`\\={\the\catcode`\%\relax
```



```

2777 \catcode`\=\the\catcode`\}\relax}%
2778 \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.

```

2779 \def\bbl@iniline#1\bbl@iniline{%
2780 \ifnextchar[\bbl@inisect{\ifnextchar\bbl@iniskip\bbl@inistore}#1\@@}% ]
2781 \def\bbl@inisect[#1]#2\@@{\def\bbl@section{#1}}
2782 \def\bbl@iniskip#1\@@{% if starts with ;
2783 \def\bbl@inistore#1=#2\@@{% full (default)
2784 \bbl@trim@def\bbl@tempa{#1}%
2785 \bbl@trim\toks@{#2}%
2786 \bbl@xin@;\bbl@section/\bbl@tempa;}{\bbl@key@list}%
2787 \ifin@
2788 \bbl@exp{%
2789 \g@addto@macro\bbl@inidata{%
2790 \bbl@elt{\bbl@section}{\bbl@tempa}{\the\toks@}}}%
2791 \fi}
2792 \def\bbl@inistore@min#1=#2\@@{% minimal (maybe set in \bbl@read@ini)
2793 \bbl@trim@def\bbl@tempa{#1}%
2794 \bbl@trim\toks@{#2}%
2795 \bbl@xin@{.identification.}{.\bbl@section.}%
2796 \ifin@
2797 \bbl@exp{\g@addto@macro\bbl@inidata{%
2798 \bbl@elt{identification}{\bbl@tempa}{\the\toks@}}}%
2799 \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.

```

2800 \ifx\bbl@readstream\undefined
2801 \csname newread\endcsname\bbl@readstream
2802 \fi
2803 \def\bbl@read@ini#1#2{%
2804 \global\let\bbl@extend@ini\gobble
2805 \openin\bbl@readstream=babel-#1.ini
2806 \ifeof\bbl@readstream
2807 \bbl@error
2808 {There is no ini file for the requested language\%
2809 (#1: \language). Perhaps you misspelled it or your\%
2810 installation is not complete.}%
2811 {Fix the name or reinstall babel.}%
2812 \else
2813 % == Store ini data in \bbl@inidata ==
2814 \catcode`\[=12 \catcode`\]=12 \catcode`\==12 \catcode`\&=12
2815 \catcode`\;=12 \catcode`\|=12 \catcode`\%=14 \catcode`\-=12
2816 \bbl@info{Importing
2817 \ifcase#2font and identification \or basic \fi
2818 data for \language\%
2819 from babel-#1.ini. Reported}%
2820 \ifnum#2=\z@
2821 \global\let\bbl@inidata\empty
2822 \let\bbl@inistore\bbl@inistore@min % Remember it's local
2823 \fi
2824 \def\bbl@section{identification}%
2825 \bbl@exp{\bbl@inistore tag.ini=#1\@@}%
2826 \bbl@inistore load.level=#2\@@
2827 \loop
2828 \if T\ifeof\bbl@readstream F\fi T\relax % Trick, because inside \loop
2829 \endlinechar\m@ne

```

```

2830 \read\bbbl@readstream to \bbbl@line
2831 \endlinechar`\^^M
2832 \ifx\bbbl@line\@empty\else
2833 \expandafter\bbbl@iniline\bbbl@line\bbbl@iniline
2834 \fi
2835 \repeat
2836 % == Process stored data ==
2837 \bbbl@csarg\xdef{lini@\languagename}{#1}%
2838 \bbbl@read@ini@aux
2839 % == 'Export' data ==
2840 \bbbl@ini@exports{#2}%
2841 \global\bbbl@csarg\let{inidata@\languagename}\bbbl@inidata
2842 \global\let\bbbl@inidata\@empty
2843 \bbbl@exp{\bbbl@add@list\bbbl@ini@loaded{\languagename}}%
2844 \bbbl@to\global\bbbl@ini@loaded
2845 \fi}
2846 \def\bbbl@read@ini@aux{%
2847 \let\bbbl@savestrings\@empty
2848 \let\bbbl@savetoday\@empty
2849 \let\bbbl@savestate\@empty
2850 \def\bbbl@elt##1##2##3{%
2851 \def\bbbl@section{##1}%
2852 \in@{=date.}{=##1}% Find a better place
2853 \ifin@
2854 \bbbl@ini@calendar{##1}%
2855 \fi
2856 \bbbl@ifunset{bbbl@inikv@##1}{}%
2857 {\csname bbl@inikv@##1\endcsname{##2}{##3}}%
2858 \bbbl@inidata}

```

A variant to be used when the ini file has been already loaded, because it's not the first  
\babelprovide for this language.

```

2859 \def\bbbl@extend@ini@aux#1{%
2860 \bbbl@startcommands*{#1}{captions}%
2861 % Activate captions/... and modify exports
2862 \bbbl@csarg\def{inikv@captions.ligr}##1##2{%
2863 \setlocalecaption{#1}{##1}{##2}%
2864 \def\bbbl@inikv@captions##1##2{%
2865 \bbbl@ini@captions@aux{##1}{##2}%
2866 \def\bbbl@stringdef##1##2{\gdef##1{##2}}%
2867 \def\bbbl@exportkey##1##2##3{%
2868 \bbbl@ifunset{bbbl@kv@##2}{}%
2869 {\expandafter\ifx\csname bbl@kv@##2\endcsname\@empty\else
2870 \bbbl@exp{\global\let\bbbl@##1@\languagename}>\bbbl@kv@##2}>%
2871 \fi}}%
2872 % As with \bbbl@read@ini, but with some changes
2873 \bbbl@read@ini@aux
2874 \bbbl@ini@exports\tw@
2875 % Update inidata@lang by pretending the ini is read.
2876 \def\bbbl@elt##1##2##3{%
2877 \def\bbbl@section{##1}%
2878 \bbbl@iniline##2=##3\bbbl@iniline}%
2879 \csname bbl@inidata@#1\endcsname
2880 \global\bbbl@csarg\let{inidata@#1}\bbbl@inidata
2881 \StartBabelCommands*{#1}{date}% And from the import stuff
2882 \def\bbbl@stringdef##1##2{\gdef##1{##2}}%
2883 \bbbl@savetoday
2884 \bbbl@savestate
2885 \bbbl@endcommands}

```

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

```

2886 \def\bbbl@ini@calendar#1{%
2887 \lowercase{\def\bbbl@tempa{=##1=}}%
2888 \bbbl@replace\bbbl@tempa{=date.gregorian}{}%

```

```

2889 \bbl@replace\bbl@tempa{=date.}{}%
2890 \in@{.licr=}{#1=}%
2891 \ifin@
2892 \ifcase\bbl@engine
2893 \bbl@replace\bbl@tempa{.licr=}{}%
2894 \else
2895 \let\bbl@tempa\relax
2896 \fi
2897 \fi
2898 \ifx\bbl@tempa\relax\else
2899 \bbl@replace\bbl@tempa{=}{}%
2900 \bbl@exp{%
2901 \def\<bbl@inikv@#1>###1####2{%
2902 \\\bbl@inidate###1...\relax{###2}{\bbl@tempa}}}%
2903 \fi}

```

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

```

2904 \def\bbl@renewinikey#1/#2\@#3{%
2905 \edef\bbl@tempa{\zap@space #1 \@empty}% section
2906 \edef\bbl@tempb{\zap@space #2 \@empty}% key
2907 \bbl@trim\toks@{#3}% value
2908 \bbl@exp{%
2909 \edef\\bbl@key@list{\bbl@key@list \bbl@tempa/\bbl@tempb;}%
2910 \\\g@addto@macro\\bbl@inidata{%
2911 \\\bbl@elt{\bbl@tempa}{\bbl@tempb}{\the\toks@}}}%

```

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

```

2912 \def\bbl@exportkey#1#2#3{%
2913 \bbl@ifunset{bbl@kv@#2}%
2914 {\bbl@csarg\gdef{#1@\language name}{#3}}%
2915 {\expandafter\ifx\csname bbl@kv@#2\endcsname\@empty
2916 \bbl@csarg\gdef{#1@\language name}{#3}}%
2917 \else
2918 \bbl@exp{\global\let\<bbl@#1@\language name>\<bbl@kv@#2>}%
2919 \fi}}

```

Key-value pairs are treated differently depending on the section in the ini file. The following macros are the readers for identification and typography. Note \bbl@ini@exports is called always (via \bbl@inisec), while \bbl@after@ini must be called explicitly after \bbl@read@ini if necessary.

```

2920 \def\bbl@iniwarning#1{%
2921 \bbl@ifunset{bbl@kv@identification.warning#1}{}%
2922 {\bbl@warning{%
2923 From babel-\bbl@cs{lini@\language name}.ini:\%
2924 \bbl@cs{kv@identification.warning#1}\%
2925 Reported }}}
2926 %
2927 \let\bbl@release@transforms\@empty
2928 %
2929 \def\bbl@ini@exports#1{%
2930 % Identification always exported
2931 \bbl@iniwarning}%
2932 \ifcase\bbl@engine
2933 \bbl@iniwarning{.pdf latex}%
2934 \or
2935 \bbl@iniwarning{.lua latex}%
2936 \or
2937 \bbl@iniwarning{.x latex}%
2938 \fi%
2939 \bbl@exportkey{llevel}{identification.load.level}{}%
2940 \bbl@exportkey{elname}{identification.name.english}{}%

```

```

2941 \bbl@exp{\bbl@exportkey{lname}{identification.name.opentype}%
2942   {\csname bbl@elname@languagename\endcsname}}%
2943 \bbl@exportkey{tbc}{identification.tag.bcp47}}}%
2944 \bbl@exportkey{lbc}{identification.language.tag.bcp47}}}%
2945 \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
2946 \bbl@exportkey{esname}{identification.script.name}}}%
2947 \bbl@exp{\bbl@exportkey{sname}{identification.script.name.opentype}%
2948   {\csname bbl@esname@languagename\endcsname}}%
2949 \bbl@exportkey{sbc}{identification.script.tag.bcp47}}}%
2950 \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}%
2951 % Also maps bcp47 -> languagename
2952 \ifbbl@bcptoname
2953   \bbl@csarg\xdef{bcp@map@bbl@cl{tbc}}{\languagename}%
2954 \fi
2955 % Conditional
2956 \ifnum#1>\z@ % 0 = only info, 1, 2 = basic, (re)new
2957   \bbl@exportkey{lnbrk}{typography.linebreaking}{h}%
2958   \bbl@exportkey{hyphr}{typography.hyphenrules}}}%
2959   \bbl@exportkey{lfthm}{typography.lefthyphenmin}{2}%
2960   \bbl@exportkey{rgthm}{typography.righthyphenmin}{3}%
2961   \bbl@exportkey{prehc}{typography.prehyphenchar}}}%
2962   \bbl@exportkey{hyotl}{typography.hyphenate.other.locale}}}%
2963   \bbl@exportkey{hyots}{typography.hyphenate.other.script}}}%
2964   \bbl@exportkey{intsp}{typography.intraspace}}}%
2965   \bbl@exportkey{frspc}{typography.frenchspacing}{u}%
2966   \bbl@exportkey{chrng}{characters.ranges}}}%
2967   \bbl@exportkey{quote}{characters.delimiters.quotes}}}%
2968   \bbl@exportkey{dgnat}{numbers.digits.native}}}%
2969   \ifnum#1=\tw@ % only (re)new
2970     \bbl@exportkey{rqtex}{identification.require.babel}}}%
2971     \bbl@toglobal\bbl@savetoday
2972     \bbl@toglobal\bbl@savestate
2973     \bbl@savestrings
2974   \fi
2975 \fi}

```

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

```

2976 \def\bbl@inikv#1#2{%      key=value
2977   \toks@{#2}%             This hides #'s from ini values
2978   \bbl@csarg\xdef{@kv@bbl@section.#1}{\the\toks@}}

```

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

```

2979 \let\bbl@inikv@identification\bbl@inikv
2980 \let\bbl@inikv@typography\bbl@inikv
2981 \let\bbl@inikv@characters\bbl@inikv
2982 \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’.

```

2983 \def\bbl@inikv@counters#1#2{%
2984   \bbl@ifsamestring{#1}{digits}%
2985   {\bbl@error{The counter name 'digits' is reserved for mapping\\%
2986     decimal digits}%
2987     {Use another name.}}}%
2988   {}%
2989   \def\bbl@tempc{#1}%
2990   \bbl@trim@def{\bbl@tempb*}{#2}%
2991   \in@{.1$}{#1$}%
2992   \ifin@
2993     \bbl@replace\bbl@tempc{.1}{}%
2994     \bbl@csarg\protected@xdef{cntr@bbl@tempc @languagename}{%
2995       \noexpand\bbl@alphanumeric{\bbl@tempc}}}%
2996   \fi

```

```

2997 \in@{.F.}{#1}%
2998 \ifin@else\in@{.S.}{#1}\fi
2999 \ifin@
3000 \bbl@csarg\protected@xdef{cntr@#1@\language}\bbl@tempb*}%
3001 \else
3002 \toks@{}% Required by \bbl@buildifcase, which returns \bbl@tempa
3003 \expandafter\bbl@buildifcase\bbl@tempb* \ \ % Space after \
3004 \bbl@csarg{\global\expandafter\let}{cntr@#1@\language}\bbl@tempa
3005 \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.

```

3006 \ifcase\bbl@engine
3007 \bbl@csarg\def{inikv@captions.licr}#1#2{%
3008 \bbl@ini@captions@aux{#1}{#2}}
3009 \else
3010 \def\bbl@inikv@captions#1#2{%
3011 \bbl@ini@captions@aux{#1}{#2}}
3012 \fi

```

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

```

3013 \def\bbl@ini@captions@template#1#2{% string language tempa=capt-name
3014 \bbl@replace\bbl@tempa{.template}{}}%
3015 \def\bbl@toreplace{#1}{}%
3016 \bbl@replace\bbl@toreplace{[ ]}{\nobreakspace}}%
3017 \bbl@replace\bbl@toreplace{[ ]}{\csname}%
3018 \bbl@replace\bbl@toreplace{[ ]}{\csname the}%
3019 \bbl@replace\bbl@toreplace{[ ]}{name\endcsname}}%
3020 \bbl@replace\bbl@toreplace{[ ]}{\endcsname}}%
3021 \bbl@xin@{\, \bbl@tempa,}{, chapter, appendix, part,}%
3022 \ifin@
3023 \@nameuse{bbl@patch\bbl@tempa}%
3024 \global\bbl@csarg\let{\bbl@tempa fmt@#2}\bbl@toreplace
3025 \fi
3026 \bbl@xin@{\, \bbl@tempa,}{, figure, table,}%
3027 \ifin@
3028 \toks@\expandafter{\bbl@toreplace}%
3029 \bbl@exp{\gdef\<fnum@\bbl@tempa>{\the\toks@}}%
3030 \fi}
3031 \def\bbl@ini@captions@aux#1#2{%
3032 \bbl@trim@def\bbl@tempa{#1}%
3033 \bbl@xin@{.template}{\bbl@tempa}%
3034 \ifin@
3035 \bbl@ini@captions@template{#2}\language
3036 \else
3037 \bbl@ifblank{#2}%
3038 {\bbl@exp{%
3039 \toks@{\bbl@nocaption{\bbl@tempa}{\language\bbl@tempa name}}}%
3040 {\bbl@trim\toks@{#2}}}%
3041 \bbl@exp{%
3042 \bbl@add{\bbl@savestrings{%
3043 \SetString\<\bbl@tempa name>{\the\toks@}}}%
3044 \toks@\expandafter{\bbl@captionslist}%
3045 \bbl@exp{\in@{\<\bbl@tempa name>}{\the\toks@}}}%
3046 \ifin@else
3047 \bbl@exp{%
3048 \bbl@add{\bbl@extracaps@\language>{\<\bbl@tempa name>}}%
3049 \bbl@tglobal\bbl@extracaps@\language>}%
3050 \fi
3051 \fi}

```

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

```

3052 \def\bbl@list@the{%

```

```

3053 part,chapter,section,subsection,subsubsection,paragraph,%
3054 subparagraph,enumi,enumii,enumiii,enumiv,equation,figure,%
3055 table,page,footnote,mpfootnote,mpfn}
3056 \def\bbl@map@cnt#1{% #1:roman,etc, // #2:enumi,etc
3057 \bbl@ifunset\bbl@map@#1@\languagename}%
3058 {\@nameuse{#1}}%
3059 {\@nameuse\bbl@map@#1@\languagename}}
3060 \def\bbl@inikv@labels#1#2{%
3061 \in@{.map}{#1}%
3062 \ifin@
3063 \ifx\bbl@KVP@labels\@nil\else
3064 \bbl@xin@{ map }{ \bbl@KVP@labels\space}%
3065 \ifin@
3066 \def\bbl@tempc{#1}%
3067 \bbl@replace\bbl@tempc{.map}{}%
3068 \in@{,#2,}{,arabic,roman,Roman,alph,Alph,fnsymbol,}%
3069 \bbl@exp{%
3070 \gdef\<bbl@map@\bbl@tempc @\languagename>%
3071 {\ifin@\<#2>\else\\localecounter{#2}\fi}}%
3072 \bbl@foreach\bbl@list@the{%
3073 \bbl@ifunset{the##1}{}%
3074 {\bbl@exp{\let\\bbl@tempd\<the##1>}%
3075 \bbl@exp{%
3076 \\bbl@sreplace\<the##1>%
3077 {\<\bbl@tempc>{##1}}{\\bbl@map@cnt{\bbl@tempc}{##1}}%
3078 \\bbl@sreplace\<the##1>%
3079 {\<\@empty @\bbl@tempc>\<c##1>}{\\bbl@map@cnt{\bbl@tempc}{##1}}}%
3080 \expandafter\ifx\csname the##1\endcsname\bbl@tempd\else
3081 \toks@\expandafter\expandafter\expandafter{%
3082 \csname the##1\endcsname}%
3083 \expandafter\xdef\csname the##1\endcsname{\the\toks@}}%
3084 \fi}}%
3085 \fi
3086 \fi
3087 %
3088 \else
3089 %
3090 % The following code is still under study. You can test it and make
3091 % suggestions. Eg, enumerate.2 = ([enumi]).([enumii]). It's
3092 % language dependent.
3093 \in@{enumerate.}{#1}%
3094 \ifin@
3095 \def\bbl@tempa{#1}%
3096 \bbl@replace\bbl@tempa{enumerate.}{}%
3097 \def\bbl@toreplace{#2}%
3098 \bbl@replace\bbl@toreplace{[ ]}{\nobreakspace{}}%
3099 \bbl@replace\bbl@toreplace{[ ]}{\csname the}%
3100 \bbl@replace\bbl@toreplace{ ]}{\endcsname{}}%
3101 \toks@\expandafter{\bbl@toreplace}%
3102 % TODO. Execute only once:
3103 \bbl@exp{%
3104 \\bbl@add\<extras\languagename>{%
3105 \\babel@save\<labelenum\romannumeral\bbl@tempa>%
3106 \def\<labelenum\romannumeral\bbl@tempa>\the\toks@}%
3107 \\bbl@tglobal\<extras\languagename>}%
3108 \fi
3109 \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.

```

3110 \def\bbl@chapttype{chapter}

```

```

3111 \ifx\@makechapterhead\@undefined
3112   \let\bbl@patchchapter\relax
3113 \else\ifx\thechapter\@undefined
3114   \let\bbl@patchchapter\relax
3115 \else\ifx\ps@headings\@undefined
3116   \let\bbl@patchchapter\relax
3117 \else
3118   \def\bbl@patchchapter{%
3119     \global\let\bbl@patchchapter\relax
3120     \gdef\bbl@chfmt{%
3121       \bbl@ifunset{\bbl\bbl@chapttype fmt@\languagename}%
3122       {\@chapapp\space\thechapter}
3123       {\@nameuse{\bbl\bbl@chapttype fmt@\languagename}}}%
3124     \bbl@add\appendix{\def\bbl@chapttype{appendix}}% Not harmful, I hope
3125     \bbl@sreplace\ps@headings{\@chapapp\ \thechapter}{\bbl@chfmt}%
3126     \bbl@sreplace\chaptermark{\@chapapp\ \thechapter}{\bbl@chfmt}%
3127     \bbl@sreplace\@makechapterhead{\@chapapp\space\thechapter}{\bbl@chfmt}%
3128     \bbl@tglobal\appendix
3129     \bbl@tglobal\ps@headings
3130     \bbl@tglobal\chaptermark
3131     \bbl@tglobal\@makechapterhead}
3132   \let\bbl@patchappendix\bbl@patchchapter
3133 \fi\fi\fi
3134 \ifx\@part\@undefined
3135   \let\bbl@patchpart\relax
3136 \else
3137   \def\bbl@patchpart{%
3138     \global\let\bbl@patchpart\relax
3139     \gdef\bbl@partformat{%
3140       \bbl@ifunset{\bbl@partfmt@\languagename}%
3141       {\partname\nobreakspace\thepart}
3142       {\@nameuse{\bbl@partfmt@\languagename}}}%
3143     \bbl@sreplace\@part{\partname\nobreakspace\thepart}{\bbl@partformat}%
3144     \bbl@tglobal\@part}
3145 \fi

```

#### **Date.** TODO. Document

```

3146 % Arguments are _not_ protected.
3147 \let\bbl@calendar\@empty
3148 \DeclareRobustCommand\localedate[1][\bbl@localedate{#1}]
3149 \def\bbl@localedate#1#2#3#4{%
3150   \begingroup
3151     \ifx\@empty#1\@empty\else
3152       \let\bbl@ld@calendar\@empty
3153       \let\bbl@ld@variant\@empty
3154       \edef\bbl@tempa{\zap@space#1 \@empty}%
3155       \def\bbl@tempb##1=##2\@{\@namedef{\bbl@ld@##1}{##2}}%
3156       \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
3157       \edef\bbl@calendar{%
3158         \bbl@ld@calendar
3159         \ifx\bbl@ld@variant\@empty\else
3160           .\bbl@ld@variant
3161         \fi}%
3162       \bbl@replace\bbl@calendar{gregorian}{}%
3163     \fi
3164     \bbl@cased
3165     {\@nameuse{\bbl@date@\languagename @\bbl@calendar}{#2}{#3}{#4}}%
3166   \endgroup}
3167 % eg: 1=months, 2=wide, 3=1, 4=dummy, 5=value, 6=calendar
3168 \def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
3169   \bbl@trim@def\bbl@tempa{#1.#2}%
3170   \bbl@ifsamestring{\bbl@tempa}{months.wide}%           to savedate
3171   {\bbl@trim@def\bbl@tempa{#3}%

```

```

3172 \bbl@trim\toks@{#5}%
3173 \@temptokena\expandafter{\bbl@savestate}%
3174 \bbl@exp{% Reverse order - in ini last wins
3175 \def\\bbl@savestate{%
3176 \\SetString\<month\romannumeral\bbl@tempa#6name>\the\toks@}%
3177 \the\@temptokena}}}%
3178 {\bbl@ifsamestring{\bbl@tempa}{date.long}% defined now
3179 {\lowercase{\def\bbl@tempb{#6}}}%
3180 \bbl@trim@def\bbl@toreplace{#5}%
3181 \bbl@TG@@date
3182 \bbl@ifunset{\bbl@date@\language @}%
3183 {\bbl@exp{% TODO. Move to a better place.
3184 \gdef\<\language date>\protect\<\language date >%
3185 \gdef\<\language date >####1####2####3{%
3186 \\bbl@usedategrouptue
3187 \<bbl@ensure@\language >%
3188 \\localedate{####1}{####2}{####3}}}%
3189 \\bbl@add\\bbl@savetoday{%
3190 \\SetString\\today{%
3191 \<\language date>%
3192 {\the\year}{\the\month}{\the\day}}}%
3193 {}%
3194 \global\bbl@csarg\let{date@\language @}\bbl@toreplace
3195 \ifx\bbl@tempb\@empty\else
3196 \global\bbl@csarg\let{date@\language @}\bbl@tempb\bbl@toreplace
3197 \fi}%
3198 {}}}

```

**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. Note after \bbl@replace\toks@ contains the resulting string, which is used by \bbl@replace@finish@iii (this implicit behavior doesn’t seem a good idea, but it’s efficient).

```

3199 \let\bbl@calendar\@empty
3200 \newcommand\BabelDateSpace{\nobreakspace}
3201 \newcommand\BabelDateDot{.\@} % TODO. \let instead of repeating
3202 \newcommand\BabelDated[1]{\number#1}
3203 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
3204 \newcommand\BabelDateM[1]{\number#1}
3205 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}
3206 \newcommand\BabelDateMMM[1]{\%
3207 \csname month\romannumeral#1\bbl@calendar name\endcsname}}%
3208 \newcommand\BabelDatey[1]{\number#1}%
3209 \newcommand\BabelDateyy[1]{\%
3210 \ifnum#1<10 0\number#1 %
3211 \else\ifnum#1<100 \number#1 %
3212 \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
3213 \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
3214 \else
3215 \bbl@error
3216 {Currently two-digit years are restricted to the\
3217 range 0-9999.}%
3218 {There is little you can do. Sorry.}%
3219 \fi\fi\fi\fi}}
3220 \newcommand\BabelDateyyyy[1]{\number#1} % TODO - add leading 0
3221 \def\bbl@replace@finish@iii#1{%
3222 \bbl@exp{\def\\#1####1####2####3{\the\toks@}}
3223 \def\bbl@TG@@date{%
3224 \bbl@replace\bbl@toreplace{[ ]}{\BabelDateSpace}}%
3225 \bbl@replace\bbl@toreplace{[. ]}{\BabelDateDot}}%
3226 \bbl@replace\bbl@toreplace{[d]}{\BabelDated{####3}}%
3227 \bbl@replace\bbl@toreplace{[dd]}{\BabelDatedd{####3}}%
3228 \bbl@replace\bbl@toreplace{[M]}{\BabelDateM{####2}}%

```



```

3229 \bbl@replace\bbl@toreplace{[MM]}\BabelDateMM{####2}%
3230 \bbl@replace\bbl@toreplace{[MMMM]}\BabelDateMMMM{####2}%
3231 \bbl@replace\bbl@toreplace{[y]}\BabelDatey{####1}%
3232 \bbl@replace\bbl@toreplace{[yy]}\BabelDateyy{####1}%
3233 \bbl@replace\bbl@toreplace{[yyyy]}\BabelDateyyyy{####1}%
3234 \bbl@replace\bbl@toreplace{[y]}\bbl@datecctr{####1}%
3235 \bbl@replace\bbl@toreplace{[m]}\bbl@datecctr{####2}%
3236 \bbl@replace\bbl@toreplace{[d]}\bbl@datecctr{####3}%
3237 \bbl@replace@finish@iii\bbl@toreplace}
3238 \def\bbl@datecctr\expandafter\bbl@xdatecctr\expandafter}
3239 \def\bbl@xdatecctr[#1|#2]\localenumeral{#2}{#1}}

```

### Transforms.

```

3240 \let\bbl@release@transforms\@empty
3241 \@namedef{bbl@inikv@transforms.prehyphenation}{%
3242 \bbl@transforms\babelprehyphenation}
3243 \@namedef{bbl@inikv@transforms.posthyphenation}{%
3244 \bbl@transforms\babelposthyphenation}
3245 \def\bbl@transforms@aux#1#2#3#4,#5\relax{%
3246 #1[#2]{#3}{#4}{#5}}
3247 \begingroup % A hack. TODO. Don't require an specific order
3248 \catcode`\%=12
3249 \catcode`\&=14
3250 \gdef\bbl@transforms#1#2#3{&
3251 \ifx\bbl@KVP@transforms\@nil\else
3252 \directlua{
3253 local str = [=[#2]=]
3254 str = str:gsub('%.%d+%.%d+$', '')
3255 tex.print([[def\string\babeltempa]] .. str .. [[]])
3256 }&
3257 \bbl@xin@{,\babeltempa,}{,\bbl@KVP@transforms,&}
3258 \ifin@
3259 \in@{.0$}{#2$}&
3260 \ifin@
3261 \directlua{
3262 local str = string.match([[bbl@KVP@transforms]],
3263 '%(([^%(-)]%)[^%)]-\babeltempa')
3264 if str == nil then
3265 tex.print([[def\string\babeltempb]])
3266 else
3267 tex.print([[def\string\babeltempb{,attribute=}] .. str .. [[]])
3268 end
3269 }
3270 \toks@{#3}&
3271 \bbl@exp{&
3272 \\\g@addto@macro\\bbl@release@transforms{&
3273 \relax & Closes previous \bbl@transforms@aux
3274 \\\bbl@transforms@aux
3275 \\\#1{label=\babeltempa\babeltempb}{\language}\the\toks@}}&
3276 \else
3277 \g@addto@macro\bbl@release@transforms{, {#3}}&
3278 \fi
3279 \fi
3280 \fi}
3281 \endgroup

```

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

```

3282 \def\bbl@provide@lsys#1{%
3283 \bbl@ifunset{bbl@lname@#1}%
3284 {\bbl@load@info{#1}}%
3285 {}%
3286 \bbl@csarg\let{lsys@#1}\@empty
3287 \bbl@ifunset{bbl@sname@#1}{\bbl@csarg\gdef{sname@#1}{Default}}}%

```

```

3288 \bbl@ifunset{\bbl@sotf@#1}{\bbl@csarg\gdef{sotf@#1}{DFLT}}{}%
3289 \bbl@csarg\bbl@add@list{lsys@#1}{Script=\bbl@cs{sname@#1}}%
3290 \bbl@ifunset{\bbl@lname@#1}{}%
3291 {\bbl@csarg\bbl@add@list{lsys@#1}{Language=\bbl@cs{lname@#1}}}%
3292 \ifcase\bbl@engine\or\or
3293 \bbl@ifunset{\bbl@prehc@#1}{}%
3294 {\bbl@exp{\bbl@ifblank{\bbl@cs{prehc@#1}}}%
3295 }%
3296 {\ifx\bbl@xenoxyph\undefined
3297 \let\bbl@xenoxyph\bbl@xenoxyph@d
3298 \ifx\AtBeginDocument\@notprerr
3299 \expandafter\@secondoftwo % to execute right now
3300 \fi
3301 \AtBeginDocument{%
3302 \bbl@patchfont{\bbl@xenoxyph}%
3303 \expandafter\selectlanguage\expandafter{\language}%
3304 \fi}}%
3305 \fi
3306 \bbl@csarg\bbl@tglobal{lsys@#1}}
3307 \def\bbl@xenoxyph@d{%
3308 \bbl@ifset{\bbl@prehc@language}%
3309 {\ifnum\hyphenchar\font=\defaultthyphenchar
3310 \iffontchar\font\bbl@c1{prehc}\relax
3311 \hyphenchar\font\bbl@c1{prehc}\relax
3312 \else\iffontchar\font"200B
3313 \hyphenchar\font"200B
3314 \else
3315 \bbl@warning
3316 {Neither 0 nor ZERO WIDTH SPACE are available\\%
3317 in the current font, and therefore the hyphen\\%
3318 will be printed. Try changing the fontspec's\\%
3319 'HyphenChar' to another value, but be aware\\%
3320 this setting is not safe (see the manual)}%
3321 \hyphenchar\font\defaultthyphenchar
3322 \fi\fi
3323 \fi}%
3324 {\hyphenchar\font\defaultthyphenchar}}
3325 % \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).

```

3326 \def\bbl@load@info#1{%
3327 \def\BabelBeforeIni##1##2{%
3328 \begingroup
3329 \bbl@read@ini{##1}0%
3330 \endinput % babel- .tex may contain only preamble's
3331 \endgroup}% boxed, to avoid extra spaces:
3332 {\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 TeX. Non-digits characters are kept. The first macro is the generic “localized” command.

```

3333 \def\bbl@setdigits#1#2#3#4#5{%
3334 \bbl@exp{%
3335 \def<\language name digits>####1{% ie, \langdigits
3336 \<\bbl@digits@language name>####1\\\nil}%
3337 \let<\bbl@cntr@digits@language name>\<\language name digits>%
3338 \def<\language name counter>####1{% ie, \langcounter
3339 \expandafter\<\bbl@counter@language name>%
3340 \csname c#####1\endcsname}%
3341 \def<\bbl@counter@language name>####1{% ie, \bbl@counter@lang
3342 \expandafter\<\bbl@digits@language name>%

```

[illegible]

```

3364 \def\bbl@buildifcase#1 {% Returns \bbl@tempa, requires \toks@={%
3365   \ifx\\#1%                % \\ before, in case #1 is multiletter
3366     \bbl@exp{%
3367       \def\\bbl@tempa####1{%
3368         \<ifcase>####1\space\the\toks@\<else>\\<ctrerr>\<fi>}}%
3369   \else
3370     \toks@\expandafter{\the\toks@\or #1}%
3371     \expandafter\bbl@buildifcase
3372   \fi}

```

```

3373 \newcommand\localecntr[2]{\bbl@cs{cntr@#1\language}\{#2}}
3374 \def\bbl@localecntr#1#2{\localenumerical{#2}{#1}}
3375 \newcommand\localecounter[2]{%
3376   \expandafter\bbl@localecntr
3377   \expandafter{\number\csname c@#2\endcsname}{#1}}
3378 \def\bbl@alphanumeric#1#2{%
3379   \expandafter\bbl@alphanumeric@i\number#2 76543210\@@{#1}}
3380 \def\bbl@alphanumeric@i#1#2#3#4#5#6#7#8\@#9{%
3381   \ifcase\@car#8\@nil\or    % Currenty <10000, but prepared for bigger
3382     \bbl@alphanumeric@ii{#9}000000#1\or
3383     \bbl@alphanumeric@ii{#9}00000#1#2\or
3384     \bbl@alphanumeric@ii{#9}0000#1#2#3\or
3385     \bbl@alphanumeric@ii{#9}000#1#2#3#4\else
3386     \bbl@alphanumeric@invalid{>9999}%
3387   \fi}
3388 \def\bbl@alphanumeric@ii#1#2#3#4#5#6#7#8{%
3389   \bbl@ifunset{\bbl@cntr@#1.F.\number#5#6#7#8\language}%
3390     {\bbl@cs{cntr@#1.4\language}#5%
3391     \bbl@cs{cntr@#1.3\language}#6%
3392     \bbl@cs{cntr@#1.2\language}#7%
3393     \bbl@cs{cntr@#1.1\language}#8%
3394     \ifnum#6#7#8>\z@ % TODO. An ad hoc rule for Greek. Ugly.
3395       \bbl@ifunset{\bbl@cntr@#1.S.321\language}{}%
3396       {\bbl@cs{cntr@#1.S.321\language}}%
3397     \fi}%
3398 {\bbl@cs{cntr@#1.F.\number#5#6#7#8\language}}

```

```

3399 \def\bbl@alphnum@invalid#1{%
3400   \bbl@error{Alphabetic numeral too large (#1)}%
3401   {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.

```

3402 \newcommand\localeinfo[1]{%
3403   \bbl@ifunset\bbl@csname\bbl@info@#1\endcsname @\language}%
3404   {\bbl@error{I've found no info for the current locale.\%
3405     The corresponding ini file has not been loaded\%
3406     Perhaps it doesn't exist}%
3407     {See the manual for details.}}%
3408   {\bbl@cs{\csname\bbl@info@#1\endcsname @\language}}}%
3409 % \@namedef\bbl@info@name.locale\{lcname}
3410 \@namedef\bbl@info@tag.ini\{lini}
3411 \@namedef\bbl@info@name.english\{elname}
3412 \@namedef\bbl@info@name.opentype\{lname}
3413 \@namedef\bbl@info@tag.bcp47\{tbc}
3414 \@namedef\bbl@info@language.tag.bcp47\{lbc}
3415 \@namedef\bbl@info@tag.opentype\{lotf}
3416 \@namedef\bbl@info@script.name\{esname}
3417 \@namedef\bbl@info@script.name.opentype\{sname}
3418 \@namedef\bbl@info@script.tag.bcp47\{sbcp}
3419 \@namedef\bbl@info@script.tag.opentype\{sotf}
3420 \let\bbl@ensureinfo\@gobble
3421 \newcommand\BabelEnsureInfo{%
3422   \ifx\InputIfFileExists\undefined\else
3423     \def\bbl@ensureinfo##1{%e
3424       \bbl@ifunset\bbl@lname@##1\{\bbl@load@info{##1}\}}%
3425     \fi
3426     \bbl@foreach\bbl@loaded{%
3427       \def\language{##1}%
3428       \bbl@ensureinfo{##1}}}%
3429 \AtEndOfPackage{% Test for plain.
3430   \ifx\undefined\bbl@loaded\else\BabelEnsureInfo\fi}

```

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`.

```

3431 \newcommand\getlocaleproperty{%
3432   \@ifstar\bbl@getproperty@s\bbl@getproperty@x}
3433 \def\bbl@getproperty@s#1#2#3{%
3434   \let#1\relax
3435   \def\bbl@elt##1##2##3{%
3436     \bbl@ifsamestring{##1/##2}{##3}%
3437     {\providecommand#1{##3}%
3438     \def\bbl@elt####1####2####3{}}}%
3439   {}}%
3440   \bbl@cs{inidata@#2}}%
3441 \def\bbl@getproperty@x#1#2#3{%
3442   \bbl@getproperty@s{#1}{#2}{#3}%
3443   \ifx#1\relax
3444     \bbl@error
3445     {Unknown key for locale '#2':\%
3446     #3\%
3447     \string#1 will be set to \relax}%
3448     {Perhaps you misspelled it.}%
3449   \fi}
3450 \let\bbl@ini@loaded\@empty
3451 \newcommand\LocaleForEach{\bbl@foreach\bbl@ini@loaded}

```

## 8 Adjusting the Babel bahavior

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

```
3452 \newcommand\babeladjust[1]{% TODO. Error handling.
3453   \bbl@forkv{#1}{%
3454     \bbl@ifunset{bbl@ADJ@##1@##2}%
3455     {\bbl@cs{ADJ@##1}{##2}}%
3456     {\bbl@cs{ADJ@##1@##2}}}
3457 %
3458 \def\bbl@adjust@lua#1#2{%
3459   \ifvmode
3460     \ifnum\currentgrouplevel=\z@
3461       \directlua{ Babel.#2 }%
3462       \expandafter\expandafter\expandafter\@gobble
3463     \fi
3464   \fi
3465   {\bbl@error   % The error is gobbled if everything went ok.
3466     {Currently, #1 related features can be adjusted only\\%
3467       in the main vertical list.}%
3468     {Maybe things change in the future, but this is what it is.}}}
3469 \@namedef{bbl@ADJ@bidi.mirroring@on}{%
3470   \bbl@adjust@lua{bidi}{mirroring_enabled=true}}
3471 \@namedef{bbl@ADJ@bidi.mirroring@off}{%
3472   \bbl@adjust@lua{bidi}{mirroring_enabled=false}}
3473 \@namedef{bbl@ADJ@bidi.text@on}{%
3474   \bbl@adjust@lua{bidi}{bidi_enabled=true}}
3475 \@namedef{bbl@ADJ@bidi.text@off}{%
3476   \bbl@adjust@lua{bidi}{bidi_enabled=false}}
3477 \@namedef{bbl@ADJ@bidi.mapdigits@on}{%
3478   \bbl@adjust@lua{bidi}{digits_mapped=true}}
3479 \@namedef{bbl@ADJ@bidi.mapdigits@off}{%
3480   \bbl@adjust@lua{bidi}{digits_mapped=false}}
3481 %
3482 \@namedef{bbl@ADJ@linebreak.sea@on}{%
3483   \bbl@adjust@lua{linebreak}{sea_enabled=true}}
3484 \@namedef{bbl@ADJ@linebreak.sea@off}{%
3485   \bbl@adjust@lua{linebreak}{sea_enabled=false}}
3486 \@namedef{bbl@ADJ@linebreak.cjk@on}{%
3487   \bbl@adjust@lua{linebreak}{cjk_enabled=true}}
3488 \@namedef{bbl@ADJ@linebreak.cjk@off}{%
3489   \bbl@adjust@lua{linebreak}{cjk_enabled=false}}
3490 \@namedef{bbl@ADJ@justify.arabic@on}{%
3491   \bbl@adjust@lua{linebreak}{arabic.justify_enabled=true}}
3492 \@namedef{bbl@ADJ@justify.arabic@off}{%
3493   \bbl@adjust@lua{linebreak}{arabic.justify_enabled=false}}
3494 %
3495 \def\bbl@adjust@layout#1{%
3496   \ifvmode
3497     #1%
3498     \expandafter\expandafter\@gobble
3499   \fi
3500   {\bbl@error   % The error is gobbled if everything went ok.
3501     {Currently, layout related features can be adjusted only\\%
3502       in vertical mode.}%
3503     {Maybe things change in the future, but this is what it is.}}}
3504 \@namedef{bbl@ADJ@layout.tabular@on}{%
3505   \bbl@adjust@layout{\let\@tabular\bbl@NL@tabular}}
3506 \@namedef{bbl@ADJ@layout.tabular@off}{%
3507   \bbl@adjust@layout{\let\@tabular\bbl@OL@tabular}}
3508 \@namedef{bbl@ADJ@layout.lists@on}{%
3509   \bbl@adjust@layout{\let\list\bbl@NL@list}}
3510 \@namedef{bbl@ADJ@layout.lists@off}{%
3511   \bbl@adjust@layout{\let\list\bbl@OL@list}}
```

```

3512 \@namedef{bbl@ADJ@hyphenation.extra@on}{%
3513   \bbl@activateposthyphen}
3514 %
3515 \@namedef{bbl@ADJ@autoload.bcp47@on}{%
3516   \bbl@bcpallowedtrue}
3517 \@namedef{bbl@ADJ@autoload.bcp47@off}{%
3518   \bbl@bcpallowedfalse}
3519 \@namedef{bbl@ADJ@autoload.bcp47.prefix}#1{%
3520   \def\bbl@bcp@prefix{#1}}
3521 \def\bbl@bcp@prefix{bcp47-}
3522 \@namedef{bbl@ADJ@autoload.options}#1{%
3523   \def\bbl@autoload@options{#1}}
3524 \let\bbl@autoload@bcptoptions\empty
3525 \@namedef{bbl@ADJ@autoload.bcp47.options}#1{%
3526   \def\bbl@autoload@bcptoptions{#1}}
3527 \newif\ifbbl@bcptname
3528 \@namedef{bbl@ADJ@bcp47.toname@on}{%
3529   \bbl@bcptonametrue}
3530 \BabelEnsureInfo}
3531 \@namedef{bbl@ADJ@bcp47.toname@off}{%
3532   \bbl@bcptonamefalse}
3533 \@namedef{bbl@ADJ@prehyphenation.disable@nohyphenation}{%
3534   \directlua{ Babel.ignore_pre_char = function(node)
3535     return (node.lang == \the\csname l@nohyphenation\endcsname)
3536   end }}
3537 \@namedef{bbl@ADJ@prehyphenation.disable@off}{%
3538   \directlua{ Babel.ignore_pre_char = function(node)
3539     return false
3540   end }}
3541 \@namedef{bbl@ADJ@select.write@shift}{%
3542   \let\bbl@restorelastskip\relax
3543   \def\bbl@savelastskip{%
3544     \let\bbl@restorelastskip\relax
3545     \ifvmode
3546       \ifdim\lastskip=\z@
3547         \let\bbl@restorelastskip\nobreak
3548       \else
3549         \bbl@exp{%
3550           \def\\bbl@restorelastskip{%
3551             \skip@=\the\lastskip
3552             \\nobreak \vskip-\skip@ \vskip\skip@}}%
3553       \fi
3554     \fi}}
3555 \@namedef{bbl@ADJ@select.write@keep}{%
3556   \let\bbl@restorelastskip\relax
3557   \let\bbl@savelastskip\relax}
3558 \@namedef{bbl@ADJ@select.write@omit}{%
3559   \let\bbl@restorelastskip\relax
3560   \def\bbl@savelastskip##1\bbl@restorelastskip{}}

  As the final task, load the code for lua. TODO: use babel name, override
3561 \ifx\directlua\undefined\else
3562   \ifx\bbl@luapatterns\undefined
3563     \input luababel.def
3564   \fi
3565 \fi

  Continue with  $\LaTeX$ .
3566 \end{package}
3567 \end{package}

```

## 8.1 Cross referencing macros

The  $\LaTeX$  book states:

The *key* argument is any sequence of letters, digits, and punctuation symbols; upper- and lowercase letters are regarded as different.

When the above quote should still be true when a document is typeset in a language that has active characters, special care has to be taken of the category codes of these characters when they appear in an argument of the cross referencing macros.

When a cross referencing command processes its argument, all tokens in this argument should be character tokens with category ‘letter’ or ‘other’.

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

```
3568 <(*More package options)> ≡
3569 \DeclareOption{safe=none}{\let\bbl@opt@safe\empty}
3570 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
3571 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
3572 \DeclareOption{safe=refbib}{\def\bbl@opt@safe{BR}}
3573 \DeclareOption{safe=bibref}{\def\bbl@opt@safe{BR}}
3574 <(/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.

```
3575 \bbl@trace{Cross referencing macros}
3576 \ifx\bbl@opt@safe\empty\else % ie, if 'ref' and/or 'bib'
3577   \def\@newl@bel#1#2#3{%
3578     \@safe@activestrue
3579     \bbl@ifunset{#1@#2}%
3580     \relax
3581     {\gdef\@multiplelabels{%
3582       \@latex@warning@no@line{There were multiply-defined labels}}%
3583       \@latex@warning@no@line{Label `#2' multiply defined}}%
3584     \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.

```
3585 \CheckCommand*\@testdef[3]{%
3586   \def\reserved@a{#3}%
3587   \expandafter\ifx\csname#1@#2\endcsname\reserved@a
3588   \else
3589     \@tempwattrue
3590   \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.

```
3591 \def\@testdef#1#2#3{% TODO. With @samestring?
3592   \@safe@activestrue
3593   \expandafter\let\expandafter\bbl@tempa\csname #1@#2\endcsname
3594   \def\bbl@tempb{#3}%
3595   \@safe@activesfalse
3596   \ifx\bbl@tempa\relax
3597   \else
3598     \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
3599   \fi
3600   \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
3601   \ifx\bbl@tempa\bbl@tempb
3602   \else
3603     \@tempwattrue
3604   \fi
3605 \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.

```

3606 \bbl@xin@{R}\bbl@opt@safe
3607 \ifin@
3608 \edef\bbl@tempc{\expandafter\string\csname ref code\endcsname}%
3609 \bbl@xin@{\expandafter\strip@prefix\meaning\bbl@tempc}%
3610 {\expandafter\strip@prefix\meaning\ref}%
3611 \ifin@
3612 \bbl@redefine\@kernel@ref#1{%
3613 \@safe@activetrue\org@@kernel@ref{#1}\@safe@activesfalse}
3614 \bbl@redefine\@kernel@pageref#1{%
3615 \@safe@activetrue\org@@kernel@pageref{#1}\@safe@activesfalse}
3616 \bbl@redefine\@kernel@sref#1{%
3617 \@safe@activetrue\org@@kernel@sref{#1}\@safe@activesfalse}
3618 \bbl@redefine\@kernel@spageref#1{%
3619 \@safe@activetrue\org@@kernel@spageref{#1}\@safe@activesfalse}
3620 \else
3621 \bbl@redefineroast\ref#1{%
3622 \@safe@activetrue\org@ref{#1}\@safe@activesfalse}
3623 \bbl@redefineroast\pageref#1{%
3624 \@safe@activetrue\org@pageref{#1}\@safe@activesfalse}
3625 \fi
3626 \else
3627 \let\org@ref\ref
3628 \let\org@pageref\pageref
3629 \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.

```

3630 \bbl@xin@{B}\bbl@opt@safe
3631 \ifin@
3632 \bbl@redefine\@citex[#1]#2{%
3633 \@safe@activetrue\edef\@tempa{#2}\@safe@activesfalse
3634 \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.

```

3635 \AtBeginDocument{%
3636 \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.)

```

3637 \def\@citex[#1][#2]#3{%
3638 \@safe@activetrue\edef\@tempa{#3}\@safe@activesfalse
3639 \org@@citex[#1][#2]{\@tempa}}%
3640 }{}

```

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

```

3641 \AtBeginDocument{%
3642 \ifpackageloaded{cite}{%
3643 \def\@citex[#1]#2{%
3644 \@safe@activetrue\org@@citex[#1][#2]\@safe@activesfalse}%
3645 }{}

```

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

```

3646 \bbl@redefine\nocite#1{%
3647 \@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.

```

3648 \bbl@redefine\bibcite{%
3649   \bbl@cite@choice
3650   \bibcite}

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

3651 \def\bbl@bibcite#1#2{%
3652   \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.

3653 \def\bbl@cite@choice{%
3654   \global\let\bibcite\bbl@bibcite
3655   \@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{}%
3656   \@ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{}%
3657   \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.

3658 \AtBeginDocument{\bbl@cite@choice}

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

3659 \bbl@redefine\@bibitem#1{%
3660   \@safe@activetrue\org@bibitem{#1}\@safe@activesfalse}
3661 \else
3662   \let\org@nocite\nocite
3663   \let\org@@citex\@citex
3664   \let\org@bibcite\bibcite
3665   \let\org@@bibitem\@bibitem
3666 \fi

```

## 8.2 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.

```

3667 \bbl@trace{Marks}
3668 \IfBabelLayout{sectioning}
3669   {\ifx\bbl@opt@headfoot\@nnil
3670     \g@addto@macro\@resetactivechars{%
3671       \set@typeset@protect
3672       \expandafter\select@language\x\expandafter{\bbl@main@language}%
3673       \let\protect\noexpand
3674       \ifcase\bbl@bidimode\else % Only with bidi. See also above
3675         \edef\thepage{%
3676           \noexpand\babelsublr{\unexpanded\expandafter{\thepage}}}%
3677       \fi}%
3678   \fi}
3679 {\ifbbl@single\else
3680   \bbl@ifunset{markright }{\bbl@redefine\bbl@redefineroobust
3681     \markright#1{%
3682       \bbl@ifblank{#1}%

```

```

3683      {\org@markright{}}}%
3684      {\toks@{#1}}%
3685      \bbl@exp{%
3686          \org@markright{\protect\foreignlanguage{\language\name}%
3687              {\protect\bbl@restore@actives\the\toks@}}}%
\markboth The definition of \markboth is equivalent to that of \markright, except that we need two token
\@mkboth registers. The documentclasses report and book define and set the headings for the page. While
doing so they also store a copy of \markboth in \@mkboth. Therefore we need to check whether
\@mkboth has already been set. If so we need to do that again with the new definition of \markboth.
(As of Oct 2019,  $\TeX$  stores the definition in an intermediate macro, so it's not necessary anymore,
but it's preserved for older versions.)
3688      \ifx\@mkboth\markboth
3689          \def\bbl@tempc{\let\@mkboth\markboth}
3690      \else
3691          \def\bbl@tempc{}
3692      \fi
3693      \bbl@ifunset{markboth }{\bbl@redefine\bbl@redefineroobust
3694          \markboth#1#2{%
3695              \protected@edef\bbl@tempb##1{%
3696                  \protect\foreignlanguage
3697                      {\language\name}{\protect\bbl@restore@actives##1}}}%
3698              \bbl@ifblank{#1}%
3699                  {\toks@{}}%
3700                  {\toks@\expandafter{\bbl@tempb{#1}}}%
3701              \bbl@ifblank{#2}%
3702                  {\@temptokena{}}%
3703                  {\@temptokena\expandafter{\bbl@tempb{#2}}}%
3704              \bbl@exp{\org@markboth{\the\toks@}{\the\@temptokena}}%
3705              \bbl@tempc
3706          \fi} % end ifbbl@single, end \IfBabelLayout

```

## 8.3 Preventing clashes with other packages

### 8.3.1 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.

```

3707 \bbl@trace{Preventing clashes with other packages}
3708 \ifx\org@ref\undefined\else
3709     \bbl@xin{R}\bbl@opt@safe
3710     \ifin@
3711         \AtBeginDocument{%
3712             \@ifpackageloaded{ifthen}{%
3713                 \bbl@redefine@long\ifthenelse#1#2#3{%
3714                     \let\bbl@temp@pref\pageref
3715                     \let\pageref\org@pageref
3716                     \let\bbl@temp@ref\ref
3717                     \let\ref\org@ref

```

```

3718         \@safe@activetrue
3719         \org@ifthenelse{#1}%
3720         {\let\pageref\bbl@temp@pref
3721          \let\ref\bbl@temp@ref
3722          \@safe@activesfalse
3723          #2}%
3724         {\let\pageref\bbl@temp@pref
3725          \let\ref\bbl@temp@ref
3726          \@safe@activesfalse
3727          #3}%
3728     }%
3729 }{}%
3730 }
3731 \fi

```

### 8.3.2 varioref

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

```

3732 \AtBeginDocument{%
3733   \@ifpackageloaded{varioref}{%
3734     \bbl@redefine\@vpageref#1[#2]#3{%
3735       \@safe@activetrue
3736       \org@@@vpageref{#1}[#2]#3}%
3737     \@safe@activesfalse}%
3738   \bbl@redefine\vrefpagenum#1#2{%
3739     \@safe@activetrue
3740     \org@vrefpagenum{#1}#2}%
3741   \@safe@activesfalse}%

```

The package `varioref` defines `\Ref` to be a robust command which uppercases the first character of the reference text. In order to be able to do that it needs to access the expandable form of `\ref`. So we employ a little trick here. We redefine the (internal) command `\Ref␣` to call `\org@ref` instead of `\ref`. The disadvantage of this solution is that whenever the definition of `\Ref` changes, this definition needs to be updated as well.

```

3742   \expandafter\def\csname Ref\endcsname#1{%
3743     \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
3744   }{}%
3745 }
3746 \fi

```

### 8.3.3 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.

```

3747 \AtEndOfPackage{%
3748   \AtBeginDocument{%
3749     \@ifpackageloaded{hhline}%
3750     {\expandafter\ifx\csname normal@char\string\endcsname\relax
3751      \else
3752        \makeatletter
3753        \def\@currname{hhline}\input{hhline.sty}\makeatother
3754        \fi}%
3755     {}}}

```

`\substitutefontfamily` Deprecated. Use the tools provided by  $\TeX$ . 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.

```

3756 \def\substitutefontfamily#1#2#3{%
3757   \lowercase{\immediate\openout15=#1#2.fd\relax}%
3758   \immediate\write15{%
3759     \string\ProvidesFile{#1#2.fd}%
3760     [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
3761     \space generated font description file]^^J
3762     \string\DeclareFontFamily{#1}{#2}{^^J
3763     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{^^J
3764     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{^^J
3765     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{^^J
3766     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{^^J
3767     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{^^J
3768     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{^^J
3769     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{^^J
3770     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{^^J
3771     }%
3772   \closeout15
3773 }
3774 \@onlypreamble\substitutefontfamily

```

## 8.4 Encoding and fonts

Because documents may use non-ASCII font encodings, we make sure that the logos of  $\TeX$  and  $\LaTeX$  always come out in the right encoding. There is a list of non-ASCII encodings. Requested encodings are currently stored in `\fontenc@load@list`. If a non-ASCII has been loaded, we define versions of `\TeX` and `\LaTeX` for them using `\ensureascii`. The default ASCII encoding is set, too (in reverse order): the “main” encoding (when the document begins), the last loaded, or OT1.

`\ensureascii`

```

3775 \bbl@trace{Encoding and fonts}
3776 \newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU}
3777 \newcommand\BabelNonText{TS1,T3,TS3}
3778 \let\org@TeX\TeX
3779 \let\org@LaTeX\LaTeX
3780 \let\ensureascii@firstofone
3781 \AtBeginDocument{%
3782   \def\@elt#1{,#1,}%
3783   \edef\bbl@tempa{\expandafter\@gobbletwo\fontenc@load@list}%
3784   \let\@elt\relax
3785   \let\bbl@tempb\@empty
3786   \def\bbl@tempc{OT1}%
3787   \bbl@foreach\BabelNonASCII{% LGR loaded in a non-standard way
3788     \bbl@ifunset{T#1}{\def\bbl@tempb{#1}}}%
3789   \bbl@foreach\bbl@tempa{%
3790     \bbl@xin@{#1}{\BabelNonASCII}%
3791     \ifin@
3792       \def\bbl@tempb{#1}% Store last non-ascii
3793     \else\bbl@xin@{#1}{\BabelNonText}% Pass
3794     \ifin@
3795       \def\bbl@tempc{#1}% Store last ascii
3796     \fi
3797   \fi}%
3798   \ifx\bbl@tempb\@empty\else
3799     \bbl@xin@{,\cf@encoding,}{,\BabelNonASCII,\BabelNonText,}%
3800     \ifin@
3801       \edef\bbl@tempc{\cf@encoding}% The default if ascii wins
3802     \fi
3803     \edef\ensureascii#1{%
3804       {\noexpand\fontencoding{\bbl@tempc}\noexpand\selectfont#1}}%
3805     \DeclareTextCommandDefault{\TeX}{\ensureascii{\org@TeX}}%
3806     \DeclareTextCommandDefault{\LaTeX}{\ensureascii{\org@LaTeX}}%
3807   \fi}

```

Now comes the old deprecated stuff (with a little change in 3.9l, for fontspec). The first thing we need to do is to determine, at `\begin{document}`, which latin fontencoding to use.

`\latinencoding` When text is being typeset in an encoding other than ‘latin’ (OT1 or T1), it would be nice to still have Roman numerals come out in the Latin encoding. So we first assume that the current encoding at the end of processing the package is the Latin encoding.

```
3808 \AtEndOfPackage{\edef\latinencoding{\cf@encoding}}
```

But this might be overruled with a later loading of the package fontenc. Therefore we check at the execution of `\begin{document}` whether it was loaded with the T1 option. The normal way to do this (using `\ifpackageloaded`) is disabled for this package. Now we have to revert to parsing the internal macro `\@filelist` which contains all the filenames loaded.

```
3809 \AtBeginDocument{%
3810   \@ifpackageloaded{fontspec}%
3811     {\xdef\latinencoding{%
3812       \ifx\UTFencname\@undefined
3813         EU\ifcase\bbl@engine\or2\or1\fi
3814       \else
3815         \UTFencname
3816       \fi}}%
3817   {\gdef\latinencoding{OT1}%
3818     \ifx\cf@encoding\bbl@t@one
3819       \xdef\latinencoding{\bbl@t@one}%
3820     \else
3821       \def\@elt#1{,#1,}%
3822       \edef\bbl@tempa{\expandafter\@gobbletwo\@fontenc@load@list}%
3823       \let\@elt\relax
3824       \bbl@xin@{,T1,}\bbl@tempa
3825       \ifin@
3826         \xdef\latinencoding{\bbl@t@one}%
3827       \fi
3828     \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.

```
3829 \DeclareRobustCommand{\latintext}{%
3830   \fontencoding{\latinencoding}\selectfont
3831   \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.

```
3832 \ifx\@undefined\DeclareTextFontCommand
3833   \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
3834 \else
3835   \DeclareTextFontCommand{\textlatin}{\latintext}
3836 \fi
```

For several functions, we need to execute some code with `\selectfont`. With  $\TeX$  2021-06-01, there is a hook for this purpose, but in older versions the  $\TeX$  command is patched (the latter solution will be eventually removed).

```
3837 \def\bbl@patchfont#1{\AddToHook{selectfont}{#1}}
```

## 8.5 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).

- pdf<sub>tex</sub> provides a minimal support for bidi text, and it must be done by hand. Vertical typesetting is not possible.
- x<sub>etex</sub> 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.
- lua<sub>tex</sub> can provide the most complete solution, as we can manipulate almost freely the node list, the generated lines, and so on, but bidi text does not work out of the box and some development is necessary. It also provides tools to properly set left-to-right and right-to-left page layouts. As Lua $\TeX$ -ja shows, vertical typesetting is possible, too.

```

3838 \bbl@trace{Loading basic (internal) bidi support}
3839 \ifodd\bbl@engine
3840 \else % TODO. Move to txtbabel
3841   \ifnum\bbl@bidimode>100 \ifnum\bbl@bidimode<200
3842     \bbl@error
3843       {The bidi method 'basic' is available only in\\%
3844       luatex. I'll continue with 'bidi=default', so\\%
3845       expect wrong results}%
3846       {See the manual for further details.}%
3847   \let\bbl@beforeforeign\leavevmode
3848   \AtEndOfPackage{%
3849     \EnableBabelHook{babel-bidi}%
3850     \bbl@xebidipar}
3851 \fi\fi
3852 \def\bbl@loadxebidi#1{%
3853   \ifx\RTLfootnotetext\@undefined
3854     \AtEndOfPackage{%
3855       \EnableBabelHook{babel-bidi}%
3856       \ifx\fontspec\@undefined
3857         \bbl@loadfontspec % bidi needs fontspec
3858       \fi
3859       \usepackage#1{bidi}}%
3860   \fi}
3861 \ifnum\bbl@bidimode>200
3862   \ifcase\expandafter\@gobbletwo\the\bbl@bidimode\or
3863     \bbl@tentative{bidi=bidi}
3864     \bbl@loadxebidi{}
3865   \or
3866     \bbl@loadxebidi{[rldocument]}
3867   \or
3868     \bbl@loadxebidi{}
3869   \fi
3870 \fi
3871 \fi
3872 % TODO? Separate:
3873 \ifnum\bbl@bidimode=\@ne
3874   \let\bbl@beforeforeign\leavevmode
3875   \ifodd\bbl@engine
3876     \newattribute\bbl@attr@dir
3877     \directlua{ Babel.attr_dir = luatexbase.registernumber'bbl@attr@dir' }
3878     \bbl@exp{\output{\bodydir\pagedir\the\output}}
3879   \fi
3880   \AtEndOfPackage{%
3881     \EnableBabelHook{babel-bidi}%
3882     \ifodd\bbl@engine\else
3883       \bbl@xebidipar
3884     \fi}
3885 \fi

```

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

```

3886 \bbl@trace{Macros to switch the text direction}
3887 \def\bbl@alscripts{,Arabic,Syriac,Thaana,}

```

```

3888 \def\bb@rscripts{% TODO. Base on codes ??
3889   ,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
3890   Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaeen,%
3891   Manichaeen,Meroitic Cursive,Meroitic,Old North Arabian,%
3892   Nabataean,N'Ko,Orkhon,Palmyrene,Inscriptional Pahlavi,%
3893   Psalter Pahlavi,Phoenician,Inscriptional Parthian,Samaritan,%
3894   Old South Arabian,}%
3895 \def\bb@provide@dirs#1{%
3896   \bb@xin@{\csname bbl@sname@#1\endcsname}{\bb@alscripts\bb@rscripts}%
3897   \ifin@
3898     \global\bb@csarg\chardef{wdir@#1}\@ne
3899     \bb@xin@{\csname bbl@sname@#1\endcsname}{\bb@alscripts}%
3900     \ifin@
3901       \global\bb@csarg\chardef{wdir@#1}\tw@ % useless in xetex
3902     \fi
3903   \else
3904     \global\bb@csarg\chardef{wdir@#1}\z@
3905   \fi
3906   \ifodd\bb@engine
3907     \bb@csarg\ifcase{wdir@#1}%
3908       \directlua{ Babel.locale_props[\the\localeid].texdir = 'l' }%
3909     \or
3910       \directlua{ Babel.locale_props[\the\localeid].texdir = 'r' }%
3911     \or
3912       \directlua{ Babel.locale_props[\the\localeid].texdir = 'al' }%
3913     \fi
3914   \fi}
3915 \def\bb@switchdir{%
3916   \bb@ifunset{bb@lsys@\languagename}{\bb@provide@lsys{\languagename}}{%
3917     \bb@ifunset{bb@wdir@\languagename}{\bb@provide@dirs{\languagename}}{%
3918       \bb@exp{\bb@setdirs\bb@cl{wdir}}}%
3919 \def\bb@setdirs#1{% TODO - math
3920   \ifcase\bb@select@type % TODO - strictly, not the right test
3921     \bb@bodydir{#1}%
3922     \bb@pardir{#1}%
3923   \fi
3924   \bb@texdir{#1}}
3925 % TODO. Only if \bb@bidimode > 0:
3926 \AddBabelHook{babel-bidi}{afterextras}{\bb@switchdir}
3927 \DisableBabelHook{babel-bidi}

```

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

```

3928 \ifodd\bb@engine % luatex=1
3929 \else % pdftex=0, xetex=2
3930   \newcount\bb@dirlevel
3931   \chardef\bb@thetexdir\z@
3932   \chardef\bb@thepardir\z@
3933   \def\bb@texdir#1{%
3934     \ifcase#1\relax
3935       \chardef\bb@thetexdir\z@
3936       \bb@texdir@i\beginL\endL
3937     \else
3938       \chardef\bb@thetexdir\@ne
3939       \bb@texdir@i\beginR\endR
3940     \fi}
3941   \def\bb@texdir@i#1#2{%
3942     \ifhmode
3943       \ifnum\currentgrouplevel>\z@
3944         \ifnum\currentgrouplevel=\bb@dirlevel
3945           \bb@error{Multiple bidi settings inside a group}%
3946           {I'll insert a new group, but expect wrong results.}%
3947           \bgroup\aftergroup#2\aftergroup\egroup
3948         \else

```

```

3949 \ifcase\currentgroup\type\or % 0 bottom
3950 \aftergroup#2% 1 simple {}
3951 \or
3952 \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
3953 \or
3954 \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
3955 \or\or\or % vbox vtop align
3956 \or
3957 \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
3958 \or\or\or\or\or\or % output math disc insert vcent mathchoice
3959 \or
3960 \aftergroup#2% 14 \beginR
3961 \else
3962 \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
3963 \fi
3964 \fi
3965 \bbl@dirlevel\currentgrouplevel
3966 \fi
3967 #1%
3968 \fi}
3969 \def\bbl@pardir#1{\chardef\bbl@thepardir#1\relax}
3970 \let\bbl@bodydir\@gobble
3971 \let\bbl@pagedir\@gobble
3972 \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).

```

3973 \def\bbl@xebidipar{%
3974 \let\bbl@xebidipar\relax
3975 \TeXeTstate\@ne
3976 \def\bbl@xeverypar{%
3977 \ifcase\bbl@thepardir
3978 \ifcase\bbl@thetextdir\else\beginR\fi
3979 \else
3980 {\setbox\z@\lastbox\beginR\box\z@}%
3981 \fi}%
3982 \let\bbl@severypar\everypar
3983 \newtoks\everypar
3984 \everypar=\bbl@severypar
3985 \bbl@severypar{\bbl@xeverypar\the\everypar}}
3986 \ifnum\bbl@bidimode>200
3987 \let\bbl@textdir\i\@gobbletwo
3988 \let\bbl@xebidipar\@empty
3989 \AddBabelHook{bidi}{foreign}{%
3990 \def\bbl@tempa{\def\BabelText####1}%
3991 \ifcase\bbl@thetextdir
3992 \expandafter\bbl@tempa\expandafter{\BabelText{\LR{####1}}}%
3993 \else
3994 \expandafter\bbl@tempa\expandafter{\BabelText{\RL{####1}}}%
3995 \fi}
3996 \def\bbl@pardir#1{\ifcase#1\relax\setLR\else\setRL\fi}
3997 \fi
3998 \fi

```

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

```

3999 \DeclareRobustCommand\babelsublr[1]{\leavevmode{\bbl@textdir\z@#1}}
4000 \AtBeginDocument{%
4001 \ifx\pdfstringdefDisableCommands\undefined\else
4002 \ifx\pdfstringdefDisableCommands\relax\else
4003 \pdfstringdefDisableCommands{\let\babelsublr\@firstofone}%
4004 \fi
4005 \fi}

```



## 8.6 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`.

```
4006 \bbl@trace{Local Language Configuration}
4007 \ifx\loadlocalcfg\undefined
4008   \@ifpackagewith{babel}{noconfigs}%
4009   {\let\loadlocalcfg@gobble}%
4010   {\def\loadlocalcfg#1{%
4011     \InputIfFileExists{#1.cfg}%
4012     {\typeout{*****^J%
4013               * Local config file #1.cfg used^^J%
4014               *}}%
4015     \@empty}}
4016 \fi
```

## 8.7 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).

```
4017 \bbl@trace{Language options}
4018 \let\bbl@afterlang\relax
4019 \let\BabelModifiers\relax
4020 \let\bbl@loaded\@empty
4021 \def\bbl@load@language#1{%
4022   \InputIfFileExists{#1.ldf}%
4023   {\edef\bbl@loaded{\CurrentOption
4024     \ifx\bbl@loaded\@empty\else,\bbl@loaded\fi}%
4025     \expandafter\let\expandafter\bbl@afterlang
4026     \csname\CurrentOption.ldf-h@@k\endcsname
4027     \expandafter\let\expandafter\BabelModifiers
4028     \csname\bbl@mod@\CurrentOption\endcsname}%
4029   {\bbl@error{%
4030     Unknown option '\CurrentOption'. Either you misspelled it\\%
4031     or the language definition file \CurrentOption.ldf was not found}%
4032     Valid options are, among others: shorthands=, KeepShorthandsActive,\\%
4033     activeacute, activegrave, noconfigs, safe=, main=, math=\\%
4034     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.

```
4035 \def\bbl@try@load@lang#1#2#3{%
4036   \IfFileExists{\CurrentOption.ldf}%
4037   {\bbl@load@language{\CurrentOption}}%
4038   {\#1\bbl@load@language{\#2}\#3}}
4039 %
4040 \DeclareOption{hebrew}{%
4041   \input{rlbabel.def}%
4042   \bbl@load@language{hebrew}}
4043 \DeclareOption{hungarian}{\bbl@try@load@lang{}{magyar}{}}
4044 \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}{}}
4045 \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}{}}
4046 \DeclareOption{polutonikogreek}{%
4047   \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}}
4048 \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}{}}
4049 \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}{}}
4050 \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.

```

4051 \ifx\bbbl@opt@config\@nnil
4052 \@ifpackagewith{babel}{noconfigs}}{}%
4053 {\InputIfFileExists{bblopts.cfg}%
4054 {\typeout{*****^J%
4055          * Local config file bblopts.cfg used^^J%
4056          *}}}%
4057 {}}}%
4058 \else
4059 \InputIfFileExists{\bbbl@opt@config.cfg}%
4060 {\typeout{*****^J%
4061          * Local config file \bbbl@opt@config.cfg used^^J%
4062          *}}}%
4063 {\bbbl@error{%
4064   Local config file '\bbbl@opt@config.cfg' not found}%
4065   Perhaps you misspelled it.}}%
4066 \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 `bbbl@language@opts` are assumed to be languages. If not declared above, the names of the option and the file are the same. We first pre-process the class and package options to determine the main language, which is processed in the third ‘main’ pass, *except* if all files are `ldf` and there is no main key. In the latter case (`\bbbl@opt@main` is still `\@nnil`), the traditional way to set the main language is kept — the last loaded is the main language.

```

4067 \ifx\bbbl@opt@main\@nnil
4068 \ifnum\bbbl@iniflag>\z@ % if all ldf's: set implicitly, no main pass
4069 \let\bbbl@tempb\@empty
4070 \edef\bbbl@tempa{\@classoptionslist,\bbbl@language@opts}%
4071 \bbbl@foreach\bbbl@tempa{\edef\bbbl@tempb{#1,\bbbl@tempb}}}%
4072 \bbbl@foreach\bbbl@tempb{% \bbbl@tempb is a reversed list
4073 \ifx\bbbl@opt@main\@nnil % ie, if not yet assigned
4074 \ifodd\bbbl@iniflag % = *=
4075 \IfFileExists{babel-#1.tex}{\def\bbbl@opt@main{#1}}}%
4076 \else % n +=
4077 \IfFileExists{#1.ldf}{\def\bbbl@opt@main{#1}}}%
4078 \fi
4079 \fi}%
4080 \fi
4081 \else
4082 \bbbl@info{Main language set with 'main='. Except if you have\\
4083           problems, prefer the default mechanism for setting\\
4084           the main language. Reported}
4085 \fi

```

A few languages are still defined explicitly. They are stored in case they are needed in the ‘main’ pass (the value can be `\relax`).

```

4086 \ifx\bbbl@opt@main\@nnil\else
4087 \bbbl@csarg\let\loadmain\expandafter\csname ds@\bbbl@opt@main\endcsname
4088 \expandafter\let\csname ds@\bbbl@opt@main\endcsname\relax
4089 \fi

```

Now define the corresponding loaders. With package options, assume the language exists. With class options, check if the option is a language by checking if the correspondin file exists.

```

4090 \bbbl@foreach\bbbl@language@opts{%
4091 \def\bbbl@tempa{#1}%
4092 \ifx\bbbl@tempa\bbbl@opt@main\else
4093 \ifnum\bbbl@iniflag<\tw@ % 0 0 (other = ldf)
4094 \bbbl@ifunset{ds@#1}%
4095 {\DeclareOption{#1}{\bbbl@load@language{#1}}}%

```

```

4096     }%
4097     \else % + * (other = ini)
4098         \DeclareOption{#1}{%
4099             \bbl@ldfinit
4100             \babelprovide[import]{#1}%
4101             \bbl@afterldf{}}%
4102     \fi
4103 \fi}
4104 \bbl@foreach\@classoptionslist{%
4105     \def\bbl@tempa{#1}%
4106     \ifx\bbl@tempa\bbl@opt@main\else
4107         \ifnum\bbl@iniflag<\tw@ % 0 0 (other = ldf)
4108             \bbl@ifunset{ds@#1}%
4109             {\IfFileExists{#1.ldf}%
4110              {\DeclareOption{#1}{\bbl@load@language{#1}}}%
4111              {}}%
4112         }%
4113     \else % + * (other = ini)
4114         \IfFileExists{babel-#1.tex}%
4115         {\DeclareOption{#1}{%
4116             \bbl@ldfinit
4117             \babelprovide[import]{#1}%
4118             \bbl@afterldf{}}}%
4119         {}}%
4120     \fi
4121 \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 (but remember class options are processed before):

```

4122 \def\AfterBabelLanguage#1{%
4123     \bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang{}}
4124 \DeclareOption*{
4125 \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. With some options in provide, the package luatexbase is loaded (and immediately used), and therefore \babelprovide can't go inside a \DeclareOption; this explains why it's executed directly, with a dummy declaration. Then all languages have been loaded, so we deactivate \AfterBabelLanguage.

```

4126 \bbl@trace{Option 'main'}
4127 \ifx\bbl@opt@main\@nnil
4128     \edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
4129     \let\bbl@tempc\@empty
4130     \bbl@for\bbl@tempb\bbl@tempa{%
4131         \bbl@xin@{,\bbl@tempb,}{,\bbl@loaded,}%
4132         \ifin@ \edef\bbl@tempc{\bbl@tempb}\fi}
4133     \def\bbl@tempa#1,#2\@nnil{\def\bbl@tempb{#1}}
4134     \expandafter\bbl@tempa\bbl@loaded,\@nnil
4135     \ifx\bbl@tempb\bbl@tempc\else
4136         \bbl@warning{%
4137             Last declared language option is '\bbl@tempc',\%
4138             but the last processed one was '\bbl@tempb'.\%
4139             The main language can't be set as both a global\%
4140             and a package option. Use 'main=\bbl@tempc' as\%
4141             option. Reported}
4142     \fi
4143 \else
4144     \ifodd\bbl@iniflag % case 1,3 (main is ini)
4145         \bbl@ldfinit
4146         \let\CurrentOption\bbl@opt@main
4147         \bbl@exp{% \bbl@opt@provide = empty if *

```

```

4148     \\\babelprovide[\bbl@opt@provide,import,main]{\bbl@opt@main}}%
4149     \bbl@afterldf{}
4150     \DeclareOption{\bbl@opt@main}{}
4151 \else % case 0,2 (main is ldf)
4152     \ifx\bbl@loadmain\relax
4153         \DeclareOption{\bbl@opt@main}{\bbl@load@language{\bbl@opt@main}}
4154     \else
4155         \DeclareOption{\bbl@opt@main}{\bbl@loadmain}
4156     \fi
4157     \ExecuteOptions{\bbl@opt@main}
4158     \@namedef{ds@\bbl@opt@main}{}%
4159 \fi
4160 \DeclareOption*{}
4161 \ProcessOptions*
4162 \fi
4163 \def\AfterBabelLanguage{%
4164     \bbl@error
4165     {Too late for \string\AfterBabelLanguage}%
4166     {Languages have been loaded, so I can do nothing}}

In order to catch the case where the user didn't specify a language we check whether
\bbl@main@language, has become defined. If not, the nil language is loaded.

4167 \ifx\bbl@main@language\@undefined
4168     \bbl@info{%
4169         You haven't specified a language. I'll use 'nil'\%
4170         as the main language. Reported}
4171     \bbl@load@language{nil}
4172 \fi
4173 \</package>

```

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

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

A proxy file for switch.def

```

4174 \*kernel>
4175 \let\bbl@onlyswitch\@empty
4176 \input babel.def
4177 \let\bbl@onlyswitch\@undefined
4178 \</kernel>
4179 \*patterns>

```

## 10 Loading hyphenation patterns

The following code is meant to be read by iniT<sub>E</sub>X because it should instruct T<sub>E</sub>X to read hyphenation patterns. To this end the docstrip option patterns is used to include this code in the file hyphen.cfg. Code is written with lower level macros.

```

4180 \<<Make sure ProvidesFile is defined>>
4181 \ProvidesFile{hyphen.cfg}[\<<date>> \<<version>> Babel hyphens]
4182 \xdef\bbl@format{\jobname}
4183 \def\bbl@version{\<<version>>}
4184 \def\bbl@date{\<<date>>}
4185 \ifx\AtBeginDocument\@undefined

```

```

4186 \def\@empty{}
4187 \fi
4188 \langle Define core switching macros \rangle

```

`\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.

```

4189 \def\process@line#1#2 #3 #4 {%
4190   \ifx=#1%
4191     \process@synonym{#2}%
4192   \else
4193     \process@language{#1#2}{#3}{#4}%
4194   \fi
4195   \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.

```

4196 \toks@{}
4197 \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.

```

4198 \def\process@synonym#1{%
4199   \ifnum\last@language=\m@ne
4200     \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
4201   \else
4202     \expandafter\chardef\csname l@#1\endcsname\last@language
4203     \wlog{\string\l@#1=\string\language\the\last@language}%
4204     \expandafter\let\csname #1hyphenmins\expandafter\endcsname
4205       \csname\language\hyphenmins\endcsname
4206     \let\bbl@elt\relax
4207     \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}{}}%
4208   \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`. T<sub>E</sub>X does not keep track of these assignments. Therefore we try to detect such assignments and store them in the `\langle lang \rangle hyphenmins` macro. When no assignments were made we provide a default setting.

Some pattern files contain changes to the `\lccode` `\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.

```

4209 \def\process@language#1#2#3{%
4210   \expandafter\addlanguage\csname l@#1\endcsname
4211   \expandafter\language\csname l@#1\endcsname
4212   \edef\language{#1}%
4213   \bbl@hook@everylanguage{#1}%
4214   % > luatex
4215   \bbl@get@enc#1::\@@@
4216   \begingroup
4217     \lefthyphenmin@m@ne
4218     \bbl@hook@loadpatterns{#2}%
4219     % > luatex
4220     \ifnum\lefthyphenmin=\m@ne
4221     \else
4222       \expandafter\xdef\csname #1hyphenmins\endcsname{%
4223         \the\lefthyphenmin\the\righthyphenmin}%
4224     \fi
4225   \endgroup
4226   \def\bbl@tempa{#3}%
4227   \ifx\bbl@tempa\@empty\else
4228     \bbl@hook@loadexceptions{#3}%
4229     % > luatex
4230   \fi
4231   \let\bbl@elt\relax
4232   \edef\bbl@languages{%
4233     \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
4234   \ifnum\the\language=\z@
4235     \expandafter\ifx\csname #1hyphenmins\endcsname\relax
4236       \set@hyphenmins\tw@\thr@@\relax
4237     \else
4238       \expandafter\expandafter\expandafter\set@hyphenmins
4239       \csname #1hyphenmins\endcsname
4240     \fi
4241     \the\toks@
4242     \toks@{}%
4243   \fi}

```

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

```

4244 \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.

```

4245 \def\bbl@hook@everylanguage#1{}
4246 \def\bbl@hook@loadpatterns#1{\input #1\relax}
4247 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
4248 \def\bbl@hook@loadkernel#1{%
4249   \def\addlanguage{\csname newlanguage\endcsname}%
4250   \def\adddialect##1##2{%
4251     \global\chardef##1##2\relax
4252     \wlog{\string##1 = a dialect from \string\language##2}}%
4253   \def\iflanguage##1{%
4254     \expandafter\ifx\csname l@##1\endcsname\relax
4255       \nolannerr{##1}%
4256     \else
4257       \ifnum\csname l@##1\endcsname=\language
4258         \expandafter\expandafter\expandafter\@firstoftwo
4259       \else
4260         \expandafter\expandafter\expandafter\@secondoftwo
4261       \fi
4262     \fi}%

```

```

4263 \def\providehyphenmins##1##2{%
4264   \expandafter\ifx\csname ##1hyphenmins\endcsname\relax
4265     \@namedef{##1hyphenmins}{##2}%
4266   \fi}%
4267 \def\set@hyphenmins##1##2{%
4268   \lefthyphenmin##1\relax
4269   \righthyphenmin##2\relax}%
4270 \def\selectlanguage{%
4271   \errhelp{Selecting a language requires a package supporting it}%
4272   \errmessage{Not loaded}}%
4273 \let\foreignlanguage\selectlanguage
4274 \let\otherlanguage\selectlanguage
4275 \expandafter\let\csname otherlanguage*\endcsname\selectlanguage
4276 \def\bbl@usehooks##1##2{% TODO. Temporary!!
4277 \def\setlocale{%
4278   \errhelp{Find an armchair, sit down and wait}%
4279   \errmessage{Not yet available}}%
4280 \let\uselocale\setlocale
4281 \let\locale\setlocale
4282 \let\selectlocale\setlocale
4283 \let\localename\setlocale
4284 \let\textlocale\setlocale
4285 \let\textlanguage\setlocale
4286 \let\languagetext\setlocale}
4287 \begingroup
4288 \def\AddBabelHook#1#2{%
4289   \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
4290     \def\next{\toks1}%
4291   \else
4292     \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname####1}%
4293   \fi
4294   \next}
4295 \ifx\directlua\@undefined
4296   \ifx\XeTeXinputencoding\@undefined\else
4297     \input xebabel.def
4298   \fi
4299 \else
4300   \input luababel.def
4301 \fi
4302 \openin1 = babel-\bbl@format.cfg
4303 \ifeof1
4304 \else
4305   \input babel-\bbl@format.cfg\relax
4306 \fi
4307 \closein1
4308 \endgroup
4309 \bbl@hook@loadkernel{switch.def}

```

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

```

4310 \openin1 = language.dat

```

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

```

4311 \def\languagename{english}%
4312 \ifeof1
4313   \message{I couldn't find the file language.dat,\space
4314     I will try the file hyphen.tex}
4315   \input hyphen.tex\relax
4316   \chardef\l@english\z@
4317 \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`.

```
4318 \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.

```
4319 \loop
4320 \endlinechar\m@ne
4321 \read1 to \bbl@line
4322 \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.

```
4323 \if T\ifeof1F\fi T\relax
4324 \ifx\bbl@line\empty\else
4325 \edef\bbl@line{\bbl@line\space\space\space}%
4326 \expandafter\process@line\bbl@line\relax
4327 \fi
4328 \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.

```
4329 \begingroup
4330 \def\bbl@elt#1#2#3#4{%
4331 \global\language=#2\relax
4332 \gdef\language#1}%
4333 \def\bbl@elt##1##2##3##4{}}%
4334 \bbl@languages
4335 \endgroup
4336 \fi
4337 \closein1
```

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

```
4338 \if/\the\toks@/\else
4339 \errhelp{language.dat loads no language, only synonyms}
4340 \errmessage{Orphan language synonym}
4341 \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.

```
4342 \let\bbl@line\@undefined
4343 \let\process@line\@undefined
4344 \let\process@synonym\@undefined
4345 \let\process@language\@undefined
4346 \let\bbl@get@enc\@undefined
4347 \let\bbl@hyph@enc\@undefined
4348 \let\bbl@tempa\@undefined
4349 \let\bbl@hook@loadkernel\@undefined
4350 \let\bbl@hook@everylanguage\@undefined
4351 \let\bbl@hook@loadpatterns\@undefined
4352 \let\bbl@hook@loadexceptions\@undefined
4353 \</patterns>
```

Here the code for `iniTeX` ends.

## 11 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].

```
4354 <(*More package options)> ≡
```



```

4355 \chardef\bbl@bidimode\z@
4356 \DeclareOption{bidi=default}{\chardef\bbl@bidimode=\@ne}
4357 \DeclareOption{bidi=basic}{\chardef\bbl@bidimode=101 }
4358 \DeclareOption{bidi=basic-r}{\chardef\bbl@bidimode=102 }
4359 \DeclareOption{bidi=bidi}{\chardef\bbl@bidimode=201 }
4360 \DeclareOption{bidi=bidi-r}{\chardef\bbl@bidimode=202 }
4361 \DeclareOption{bidi=bidi-l}{\chardef\bbl@bidimode=203 }
4362 <</More package options>>

```

With explicit languages, we could define the font at once, but we don't. Just wait and see if the language is actually activated. `bbl@font` replaces hardcoded font names inside `\.family` by the corresponding macro `\.default`.

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

```

4363 <<{*Font selection}>> ≡
4364 \bbl@trace{Font handling with fontspec}
4365 \ifx\ExplSyntaxOn\@undefined\else
4366   \ExplSyntaxOn
4367   \catcode`\ =10
4368   \def\bbl@loadfontspec{%
4369     \usepackage{fontspec}% TODO. Apply patch always
4370     \expandafter
4371     \def\csname msg~text~>~fontspec/language-not-exist\endcsname##1##2##3##4{%
4372       Font '\l_fontspec_fontname_tl' is using the\\%
4373       default features for language '##1'.\\%
4374       That's usually fine, because many languages\\%
4375       require no specific features, but if the output is\\%
4376       not as expected, consider selecting another font.}
4377     \expandafter
4378     \def\csname msg~text~>~fontspec/no-script\endcsname##1##2##3##4{%
4379       Font '\l_fontspec_fontname_tl' is using the\\%
4380       default features for script '##2'.\\%
4381       That's not always wrong, but if the output is\\%
4382       not as expected, consider selecting another font.}}
4383   \ExplSyntaxOff
4384 \fi
4385 \@onlypreamble\babelfont
4386 \newcommand\babelfont[2][{}]{% 1=langs/scripts 2=fam
4387   \bbl@foreach{#1}{%
4388     \expandafter\ifx\csname date##1\endcsname\relax
4389       \IfFileExists{babel-##1.tex}%
4390       {\babelprovide{##1}}%
4391       {}%
4392     \fi}%
4393   \edef\bbl@tempa{#1}%
4394   \def\bbl@tempb{#2}% Used by \bbl@bblfont
4395   \ifx\fontspec\@undefined
4396     \bbl@loadfontspec
4397   \fi
4398   \EnableBabelHook{babel-fontspec}% Just calls \bbl@switchfont
4399   \bbl@bblfont}
4400 \newcommand\bbl@bblfont[2][{}]{% 1=features 2=fontname, @font=rm|sf|tt
4401   \bbl@ifunset{\bbl@tempb family}%
4402   {\bbl@providefam{\bbl@tempb}}%
4403   {}%
4404   % For the default font, just in case:
4405   \bbl@ifunset{\bbl@lsys@\languagename}{\bbl@provide@lsys{\languagename}}{%
4406     \expandafter\bbl@ifblank\expandafter{\bbl@tempa}%
4407     {\bbl@csarg\edef{\bbl@tempb dflt@}{<{#1}{#2}}% save \bbl@rmdflt@
4408     \bbl@exp{%
4409       \let<\bbl@tempb dflt@\languagename>\<\bbl@tempb dflt@>%
4410       \bbl@font@set<\bbl@tempb dflt@\languagename>%

```

```

4411         \<\bbl@tempb default>\<\bbl@tempb family>}}%
4412     {\bbl@foreach\bbl@tempa{% ie bbl@rmdflt@lang / *scrt
4413         \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:

```

4414 \def\bbl@providfam#1{%
4415     \bbl@exp{%
4416         \\\newcommand\<#1default>{% Just define it
4417         \\\bbl@add@list\\bbl@font@fams{#1}%
4418         \\\DeclareRobustCommand\<#1family>{%
4419             \\\not@math@alphabet\<#1family>\relax
4420             % \\\prepare@family@series@update{#1}\<#1default>% TODO. Fails
4421             \\\fontfamily\<#1default>%
4422             \<ifx>\\UseHooks\\@undefined\<else>\\UseHook{#1family}\<fi>%
4423             \\\selectfont}%
4424             \\\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.

```

4425 \def\bbl@nostdfont#1{%
4426     \bbl@ifunset{\bbl@WFF@f@family}%
4427     {\bbl@csarg\gdef{\bbl@WFF@f@family}{% Flag, to avoid dupl warns
4428         \bbl@infowarn{The current font is not a babel standard family:\%
4429             #1%
4430             \fontname\font\\%
4431             There is nothing intrinsically wrong with this warning, and\\%
4432             you can ignore it altogether if you do not need these\\%
4433             families. But if they are used in the document, you should be\\%
4434             aware 'babel' will no set Script and Language for them, so\\%
4435             you may consider defining a new family with \string\babelfont.\\%
4436             See the manual for further details about \string\babelfont.\\%
4437             Reported}}
4438     }}%
4439 \gdef\bbl@switchfont{%
4440     \bbl@ifunset{\bbl@lsys@language}{\bbl@provide@lsys{language}}}%
4441     \bbl@exp{% eg Arabic -> arabic
4442     \lowercase{\edef\\bbl@tempa{\bbl@cl{sname}}}}%
4443     \bbl@foreach\bbl@font@fams{%
4444         \bbl@ifunset{\bbl@##1dflt@language}% (1) language?
4445         {\bbl@ifunset{\bbl@##1dflt@*bbl@tempa}% (2) from script?
4446         {\bbl@ifunset{\bbl@##1dflt@}% 2=F - (3) from generic?
4447         {% 123=F - nothing!
4448         {\bbl@exp{% 3=T - from generic
4449             \global\let\<bbl@##1dflt@language>%
4450             \<bbl@##1dflt@>}}}%
4451             {\bbl@exp{% 2=T - from script
4452                 \global\let\<bbl@##1dflt@language>%
4453                 \<bbl@##1dflt@*bbl@tempa>}}}%
4454             }}% 1=T - language, already defined
4455     \def\bbl@tempa{\bbl@nostdfont}}}%
4456     \bbl@foreach\bbl@font@fams{% don't gather with prev for
4457         \bbl@ifunset{\bbl@##1dflt@language}%
4458         {\bbl@cs{famrst@##1}%
4459         \global\bbl@csarg\let{famrst@##1}\relax}%
4460         {\bbl@exp{% order is relevant. TODO: but sometimes wrong!
4461             \\\bbl@add\\originalTeX{%
4462             \\\bbl@font@rst{\bbl@cl{##1dflt}}}%
4463             \<##1default>\<##1family>{##1}}}%
4464             \\\bbl@font@set\<bbl@##1dflt@language>% the main part!
4465             \<##1default>\<##1family>}}}%
4466     \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.

```

4467 \ifx\fbfamily\undefined\else % if latex
4468 \ifcase\bbbl@engine % if pdftex
4469 \let\bbbl@ckeckstdfonts\relax
4470 \else
4471 \def\bbbl@ckeckstdfonts{%
4472 \begingroup
4473 \global\let\bbbl@ckeckstdfonts\relax
4474 \let\bbbl@tempa\@empty
4475 \bbbl@foreach\bbbl@font@fams{%
4476 \bbbl@ifunset{\bbbl@##1dflt@}%
4477 {\@nameuse{##1family}}%
4478 \bbbl@csarg\gdef{WFF@\fbfamily}}}% Flag
4479 \bbbl@exp{\bbbl@add{\bbbl@tempa{* \<##1family}>= \fbfamily\\}%
4480 \space\space\fontname\font\\}%
4481 \bbbl@csarg\xdef{##1dflt@}{\fbfamily}%
4482 \expandafter\xdef\csname ##1default\endcsname{\fbfamily}}%
4483 {}}%
4484 \ifx\bbbl@tempa\@empty\else
4485 \bbbl@infowarn{The following font families will use the default\\%
4486 settings for all or some languages:\\%
4487 \bbbl@tempa
4488 There is nothing intrinsically wrong with it, but\\%
4489 'babel' will no set Script and Language, which could\\%
4490 be relevant in some languages. If your document uses\\%
4491 these families, consider redefining them with \string\babelfont.\\%
4492 Reported}%
4493 \fi
4494 \endgroup}
4495 \fi
4496 \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 `\bbbl@mapselect` because `\selectfont` is called internally when a font is defined.

```

4497 \def\bbbl@font@set#1#2#3{% eg \bbbl@rmdflt@lang \rmdefault \rmfamily
4498 \bbbl@xin@{<>}{#1}%
4499 \ifin@
4500 \bbbl@exp{\bbbl@fontspec@set\#1\expandafter\@gobbletwo#1\#3}%
4501 \fi
4502 \bbbl@exp{%
4503 \def\#2{#1}% eg, \rmdefault{\bbbl@rmdflt@lang}
4504 \bbbl@ifsamestring{#2}{\fbfamily}%
4505 {\#3%
4506 \bbbl@ifsamestring{\fbseries}{\bfdefault}{\bbbfseries}}}%
4507 \let\bbbl@tempa\relax}%
4508 {}}}
4509 % TODO - next should be global?, but even local does its job. I'm
4510 % still not sure -- must investigate:
4511 \def\bbbl@fontspec@set#1#2#3#4{% eg \bbbl@rmdflt@lang fnt-opt fnt-nme \xxfamily
4512 \let\bbbl@tempe\bbbl@mapselect
4513 \let\bbbl@mapselect\relax
4514 \let\bbbl@temp@fam#4% eg, 'rmfamily', to be restored below
4515 \let#4\@empty % Make sure \renewfontfamily is valid
4516 \bbbl@exp{%
4517 \let\bbbl@temp@pfam\<\bbbl@stripslash#4\space>% eg, 'rmfamily '
4518 \<keys_if_exist:nnF>{fontspec-opentype}{Script/\bbbl@cl{sname}}}%
4519 {\bbbl@newfontscript{\bbbl@cl{sname}}{\bbbl@cl{sotf}}}%
4520 \<keys_if_exist:nnF>{fontspec-opentype}{Language/\bbbl@cl{lname}}}%
4521 {\bbbl@newfontlanguage{\bbbl@cl{lname}}{\bbbl@cl{lotf}}}%
4522 \renewfontfamily\#4%
4523 [\bbbl@cl{lsys},#2]{#3}% ie \bbbl@exp{..}{#3}
4524 \begingroup

```

```

4525      #4%
4526      \xdef#1{\f@family}%      eg, \bbl@rmdflt@lang{FreeSerif(0)}
4527      \endgroup
4528      \let#4\bbl@temp@fam
4529      \bbl@exp{\let\<\bbl@stripslash#4\space>}\bbl@temp@pfam
4530      \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.

```

4531 \def\bbl@font@rst#1#2#3#4{%
4532   \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.

```

4533 \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 :-).

```

4534 \newcommand\babelFSstore[2][{%
4535   \bbl@ifblank{#1}%
4536   {\bbl@csarg\def{sname@#2}{Latin}}%
4537   {\bbl@csarg\def{sname@#2}{#1}}%
4538   \bbl@provide@dirs{#2}%
4539   \bbl@csarg\ifnum{wdir@#2}>\z@
4540     \let\bbl@beforeforeign\leavevmode
4541     \EnableBabelHook{babel-bidi}%
4542   \fi
4543   \bbl@foreach{#2}{%
4544     \bbl@FSstore{##1}{rm}\rmdefault\bbl@save@rmdefault
4545     \bbl@FSstore{##1}{sf}\sfdefault\bbl@save@sfdefault
4546     \bbl@FSstore{##1}{tt}\ttdefault\bbl@save@ttdefault}}
4547 \def\bbl@FSstore#1#2#3#4{%
4548   \bbl@csarg\edef{#2default#1}{#3}%
4549   \expandafter\addto\csname extras#1\endcsname{%
4550     \let#4#3%
4551     \ifx#3\f@family
4552       \edef#3{\csname bbl@#2default#1\endcsname}%
4553       \fontfamily{#3}\selectfont
4554     \else
4555       \edef#3{\csname bbl@#2default#1\endcsname}%
4556       \fi}%
4557   \expandafter\addto\csname noextras#1\endcsname{%
4558     \ifx#3\f@family
4559       \fontfamily{#4}\selectfont
4560     \fi
4561     \let#3#4}}
4562 \let\bbl@langfeatures\@empty
4563 \def\babelFSfeatures{% make sure \fontspec is redefined once
4564   \let\bbl@ori@fontspec\fontspec
4565   \renewcommand\fontspec[1][{%
4566     \bbl@ori@fontspec[\bbl@langfeatures##1]}
4567   \let\babelFSfeatures\bbl@FSfeatures
4568   \babelFSfeatures}
4569 \def\bbl@FSfeatures#1#2{%
4570   \expandafter\addto\csname extras#1\endcsname{%
4571     \babel@save\bbl@langfeatures
4572     \edef\bbl@langfeatures{#2,}}%
4573 \</Font selection>

```

## 12 Hooks for XeTeX and LuaTeX

### 12.1 XeTeX

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

```
4574 <(*Footnote changes)> ≡
4575 \bbl@trace{Bidi footnotes}
4576 \ifnum\bbl@bidimode>\z@
4577   \def\bbl@footnote#1#2#3{%
4578     \ifnextchar[%
4579       {\bbl@footnote@o{#1}{#2}{#3}}%
4580       {\bbl@footnote@x{#1}{#2}{#3}}}
4581   \long\def\bbl@footnote@x#1#2#3#4{%
4582     \bgroup
4583     \select@language@x{\bbl@main@language}%
4584     \bbl@fn@footnote{#2#1{\ignorespaces#4}#3}%
4585     \egroup}
4586   \long\def\bbl@footnote@o#1#2#3[#4]#5{%
4587     \bgroup
4588     \select@language@x{\bbl@main@language}%
4589     \bbl@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}%
4590     \egroup}
4591   \def\bbl@footnotetext#1#2#3{%
4592     \ifnextchar[%
4593       {\bbl@footnotetext@o{#1}{#2}{#3}}%
4594       {\bbl@footnotetext@x{#1}{#2}{#3}}}
4595   \long\def\bbl@footnotetext@x#1#2#3#4{%
4596     \bgroup
4597     \select@language@x{\bbl@main@language}%
4598     \bbl@fn@footnotetext{#2#1{\ignorespaces#4}#3}%
4599     \egroup}
4600   \long\def\bbl@footnotetext@o#1#2#3[#4]#5{%
4601     \bgroup
4602     \select@language@x{\bbl@main@language}%
4603     \bbl@fn@footnotetext[#4]{#2#1{\ignorespaces#5}#3}%
4604     \egroup}
4605   \def\BabelFootnote#1#2#3#4{%
4606     \ifx\bbl@fn@footnote\undefined
4607       \let\bbl@fn@footnote\footnote
4608       \fi
4609     \ifx\bbl@fn@footnotetext\undefined
4610       \let\bbl@fn@footnotetext\footnotetext
4611       \fi
4612     \bbl@ifblank{#2}%
4613     {\def#1{\bbl@footnote{\@firstofone}{#3}{#4}}
4614      \namedef{\bbl@stripslash#1text}%
4615      {\bbl@footnotetext{\@firstofone}{#3}{#4}}}%
4616     {\def#1{\bbl@exp{\bbl@footnote{\foreignlanguage{#2}}}{#3}{#4}}%
4617      \namedef{\bbl@stripslash#1text}%
4618      {\bbl@exp{\bbl@footnotetext{\foreignlanguage{#2}}}{#3}{#4}}}%
4619   \fi
4620 <(/Footnote changes)>
```

Now, the code.

```
4621 <(*xetex)
4622 \def\BabelStringsDefault{unicode}
4623 \let\xebbl@stop\relax
4624 \AddBabelHook{xetex}{encodedcommands}{%
4625   \def\bbl@tempa{#1}%
4626   \ifx\bbl@tempa\empty
4627     \XeTeXinputencoding"bytes"%
4628   \else
```

```

4629 \XeTeXinputencoding"#1"%
4630 \fi
4631 \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
4632 \AddBabelHook{xetex}{stopcommands}{%
4633 \xebbl@stop
4634 \let\xebbl@stop\relax}
4635 \def\bbl@intraspace#1 #2 #3\@@{%
4636 \bbl@csarg\gdef{xeisp@\language}%
4637 {\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}}
4638 \def\bbl@intrapenalty#1\@@{%
4639 \bbl@csarg\gdef{xeipn@\language}%
4640 {\XeTeXlinebreakpenalty #1\relax}}
4641 \def\bbl@provide@intraspace{%
4642 \bbl@xin@{/s}{/bbl@cl{lnbrk}}}%
4643 \ifin@else\bbl@xin@{/c}{/bbl@cl{lnbrk}}\fi
4644 \ifin@
4645 \bbl@ifunset{bbl@intsp@\language}{%
4646 {\expandafter\ifx\csname bbl@intsp@\language\endcsname\empty\else
4647 \ifx\bbl@KVP@intraspace\@nil
4648 \bbl@exp{%
4649 \bbl@intraspace\bbl@cl{intsp}\@@}%
4650 \fi
4651 \ifx\bbl@KVP@intrapenalty\@nil
4652 \bbl@intrapenalty0\@@
4653 \fi
4654 \fi
4655 \ifx\bbl@KVP@intraspace\@nil\else % We may override the ini
4656 \expandafter\bbl@intraspace\bbl@KVP@intraspace\@@
4657 \fi
4658 \ifx\bbl@KVP@intrapenalty\@nil\else
4659 \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@@
4660 \fi
4661 \bbl@exp{%
4662 % TODO. Execute only once (but redundant):
4663 \bbl@add<extras\language>%
4664 \XeTeXlinebreaklocale "\bbl@cl{tbcpr}"%
4665 \<bbl@xeisp@\language>%
4666 \<bbl@xeipn@\language>%
4667 \bbl@tglobal\<extras\language>%
4668 \bbl@add<noextras\language>%
4669 \XeTeXlinebreaklocale "en"%
4670 \bbl@tglobal\<noextras\language>%
4671 \ifx\bbl@ispacesize\undefined
4672 \gdef\bbl@ispacesize{\bbl@cl{xeisp}}%
4673 \ifx\AtBeginDocument\@notprerr
4674 \expandafter\@secondoftwo % to execute right now
4675 \fi
4676 \AtBeginDocument{\bbl@patchfont{\bbl@ispacesize}}%
4677 \fi}%
4678 \fi}
4679 \ifx\DisableBabelHook\undefined\endinput\fi
4680 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
4681 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ckeckstdfonts}
4682 \DisableBabelHook{babel-fontspec}
4683 <<Font selection>>
4684 \input txtbabel.def
4685 </xetex>

```

## 12.2 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>te</sub>x and xet<sub>ex</sub>.

```

4686 \*texet
4687 \providecommand\bbl@provide@intraspace{}
4688 \bbl@trace{Redefinitions for bidi layout}
4689 \def\bbl@sspre@caption{%
4690   \bbl@exp{\everyhbox{\bbl@textdir\bbl@cs{wdir}\bbl@main@language}}}%
4691   \ifx\bbl@opt@layout\@nnil\endinput\fi % No layout
4692   \def\bbl@startskip{\ifcase\bbl@thepardir\leftskip\else\rightskip\fi}
4693   \def\bbl@endskip{\ifcase\bbl@thepardir\rightskip\else\leftskip\fi}
4694   \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
4695     \def\@hangfrom#1{%
4696       \setbox\@tempboxa\hbox{#1}%
4697       \hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
4698       \noindent\box\@tempboxa}
4699     \def\raggedright{%
4700       \let\@centercr
4701       \bbl@startskip\z@skip
4702       \@rightskip\@flushglue
4703       \bbl@endskip\@rightskip
4704       \parindent\z@
4705       \parfillskip\bbl@startskip}
4706     \def\raggedleft{%
4707       \let\@centercr
4708       \bbl@startskip\@flushglue
4709       \bbl@endskip\z@skip
4710       \parindent\z@
4711       \parfillskip\bbl@endskip}
4712 \fi
4713 \IfBabelLayout{lists}
4714   {\bbl@sreplace\list
4715     {\@totalleftmargin\leftmargin}{\@totalleftmargin\bbl@listleftmargin}%
4716     \def\bbl@listleftmargin{%
4717       \ifcase\bbl@thepardir\leftmargin\else\rightmargin\fi}%
4718     \ifcase\bbl@engine
4719       \def\labelenumii{}\theenumii{}\pdfTeX doesn't reverse ()
4720       \def\p@enumiii{\p@enumii}\theenumii{}\fi
4721     \bbl@sreplace\@verbatim
4722     {\leftskip\@totalleftmargin}%
4723     {\bbl@startskip\textwidth
4724     \advance\bbl@startskip-\linewidth}%
4725     \bbl@sreplace\@verbatim
4726     {\rightskip\z@skip}%
4727     {\bbl@endskip\z@skip}}%
4728   {}
4729 {}
4730 \IfBabelLayout{contents}
4731   {\bbl@sreplace\@dottedtocline{\leftskip}{\bbl@startskip}%
4732     \bbl@sreplace\@dottedtocline{\rightskip}{\bbl@endskip}}
4733   {}
4734 \IfBabelLayout{columns}
4735   {\bbl@sreplace\@outputdblcol{\hb@xt@\textwidth}{\bbl@outputbox}%
4736     \def\bbl@outputbox#1{%
4737       \hb@xt@\textwidth{%
4738         \hskip\columnwidth
4739         \hfil
4740         {\normalcolor\vrule \@width\columnseprule}%
4741         \hfil
4742         \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
4743         \hskip-\textwidth
4744         \hb@xt@\columnwidth{\box\@outputbox \hss}%

```

```

4745 \hskip\columnsep
4746 \hskip\columnwidth}}}%
4747 {}
4748 <Footnote changes>
4749 \IfBabelLayout{footnotes}%
4750 {\BabelFootnote\footnote\language\language{}{}}%
4751 \BabelFootnote\localfootnote\language\language{}{}}%
4752 \BabelFootnote\mainfootnote{}{}}{}
4753 {}

```

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.

```

4754 \IfBabelLayout{counters}%
4755 {\let\bbl@latinarabic=@arabic
4756 \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}}%
4757 \let\bbl@asciroman=@roman
4758 \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
4759 \let\bbl@asciiRoman=@Roman
4760 \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}{}}{}
4761 />texet>

```

## 12.3 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).

```

4762 <*luatex>
4763 \ifx\AddBabelHook\undefined % When plain.def, babel.sty starts
4764 \bbl@trace{Read language.dat}
4765 \ifx\bbl@readstream\undefined
4766 \csname newread\endcsname\bbl@readstream
4767 \fi
4768 \begingroup
4769 \toks@{}

```



```

4770 \count@\z@ % 0=start, 1=0th, 2=normal
4771 \def\bbl@process@line#1#2 #3 #4 {%
4772   \ifx=#1%
4773     \bbl@process@synonym{#2}%
4774   \else
4775     \bbl@process@language{#1#2}{#3}{#4}%
4776   \fi
4777   \ignorespaces}
4778 \def\bbl@manylang{%
4779   \ifnum\bbl@last>\@ne
4780     \bbl@info{Non-standard hyphenation setup}%
4781   \fi
4782   \let\bbl@manylang\relax}
4783 \def\bbl@process@language#1#2#3{%
4784   \ifcase\count@
4785     \@ifundefined{zth@#1}{\count@\tw@}{\count@\@ne}%
4786   \or
4787     \count@\tw@
4788   \fi
4789   \ifnum\count@=\tw@
4790     \expandafter\addlanguage\csname l@#1\endcsname
4791     \language\allocationnumber
4792     \chardef\bbl@last\allocationnumber
4793     \bbl@manylang
4794     \let\bbl@elt\relax
4795     \xdef\bbl@languages{%
4796       \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
4797   \fi
4798   \the\toks@
4799   \toks@{}}
4800 \def\bbl@process@synonym@aux#1#2{%
4801   \global\expandafter\chardef\csname l@#1\endcsname#2\relax
4802   \let\bbl@elt\relax
4803   \xdef\bbl@languages{%
4804     \bbl@languages\bbl@elt{#1}{#2}{}}}%
4805 \def\bbl@process@synonym#1{%
4806   \ifcase\count@
4807     \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
4808   \or
4809     \@ifundefined{zth@#1}{\bbl@process@synonym@aux{#1}{0}}}%
4810   \else
4811     \bbl@process@synonym@aux{#1}{\the\bbl@last}%
4812   \fi}
4813 \ifx\bbl@languages\@undefined % Just a (sensible?) guess
4814   \chardef\l@english\z@
4815   \chardef\l@USenglish\z@
4816   \chardef\bbl@last\z@
4817   \global\@namedef{bbl@hyphendata@0}{\hyphen.tex}{}
4818   \gdef\bbl@languages{%
4819     \bbl@elt{english}{0}{\hyphen.tex}{}%
4820     \bbl@elt{USenglish}{0}{}}
4821 \else
4822   \global\let\bbl@languages@format\bbl@languages
4823   \def\bbl@elt#1#2#3#4{% Remove all except language 0
4824     \ifnum#2>\z@
4825       \noexpand\bbl@elt{#1}{#2}{#3}{#4}%
4826     \fi}%
4827   \xdef\bbl@languages{\bbl@languages}%
4828   \fi
4829   \def\bbl@elt#1#2#3#4{\@namedef{zth@#1}} % Define flags
4830   \bbl@languages
4831   \openin\bbl@readstream=language.dat
4832   \ifeof\bbl@readstream

```

```

4833 \bbl@warning{I couldn't find language.dat. No additional\\%
4834 patterns loaded. Reported}%
4835 \else
4836 \loop
4837 \endlinechar\m@ne
4838 \read\bbl@readstream to \bbl@line
4839 \endlinechar\^^M
4840 \if T\ifeof\bbl@readstream F\fi T\relax
4841 \ifx\bbl@line\empty\else
4842 \edef\bbl@line{\bbl@line\space\space\space}%
4843 \expandafter\bbl@process@line\bbl@line\relax
4844 \fi
4845 \repeat
4846 \fi
4847 \endgroup
4848 \bbl@trace{Macros for reading patterns files}
4849 \def\bbl@get@enc#1:#2:#3\@@@{\def\bbl@hyph@enc{#2}}
4850 \ifx\babelcatcodetablenum\undefined
4851 \ifx\newcatcodetable\undefined
4852 \def\babelcatcodetablenum{5211}
4853 \def\bbl@pattcodes{\numexpr\babelcatcodetablenum+1\relax}
4854 \else
4855 \newcatcodetable\babelcatcodetablenum
4856 \newcatcodetable\bbl@pattcodes
4857 \fi
4858 \else
4859 \def\bbl@pattcodes{\numexpr\babelcatcodetablenum+1\relax}
4860 \fi
4861 \def\bbl@luapatterns#1#2{%
4862 \bbl@get@enc#1::\@@@
4863 \setbox\z@\hbox\bgroup
4864 \begingroup
4865 \savecatcodetable\babelcatcodetablenum\relax
4866 \initcatcodetable\bbl@pattcodes\relax
4867 \catcodetable\bbl@pattcodes\relax
4868 \catcode`\#=6 \catcode`\$=3 \catcode`\&=4 \catcode`\^=7
4869 \catcode`\_ =8 \catcode`\{=1 \catcode`\}=2 \catcode`\~ =13
4870 \catcode`\@=11 \catcode`\^^I=10 \catcode`\^^J=12
4871 \catcode`\<=12 \catcode`\>=12 \catcode`\*=12 \catcode`\.=12
4872 \catcode`\-=12 \catcode`\/=12 \catcode`\[=12 \catcode`\]=12
4873 \catcode`\`=12 \catcode`\'=12 \catcode`\`=12
4874 \input #1\relax
4875 \catcodetable\babelcatcodetablenum\relax
4876 \endgroup
4877 \def\bbl@tempa{#2}%
4878 \ifx\bbl@tempa\empty\else
4879 \input #2\relax
4880 \fi
4881 \egroup}%
4882 \def\bbl@patterns@lua#1{%
4883 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
4884 \csname l@#1\endcsname
4885 \edef\bbl@tempa{#1}%
4886 \else
4887 \csname l@#1:\f@encoding\endcsname
4888 \edef\bbl@tempa{#1:\f@encoding}%
4889 \fi\relax
4890 \@namedef{lu@texhyphen@loaded@the\language}{} Temp
4891 \@ifundefined{bbl@hyphendata@the\language}%
4892 {\def\bbl@elt##1##2##3##4{%
4893 \ifnum##2=\csname l@bbl@tempa\endcsname % #2=spanish, dutch:OT1...
4894 \def\bbl@tempb{##3}%
4895 \ifx\bbl@tempb\empty\else % if not a synonymous

```

```

4896         \def\bbl@tempc{{##3}{##4}}%
4897         \fi
4898         \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4899         \fi}%
4900     \bbl@languages
4901     \@ifundefined{bbl@hyphendata@the\language}%
4902     {\bbl@info{No hyphenation patterns were set for\%
4903         language '\bbl@tempa'. Reported}}%
4904     {\expandafter\expandafter\expandafter\bbl@luapatterns
4905         \csname bbl@hyphendata@the\language\endcsname}}}%
4906 \endinput\fi
4907 % Here ends \ifx\AddBabelHook\@undefined
4908 % A few lines are only read by hyphen.cfg
4909 \ifx\DisableBabelHook\@undefined
4910     \AddBabelHook{luatex}{everylanguage}{%
4911         \def\process@language##1##2##3{%
4912             \def\process@line####1####2 ####3 ####4 {}}%
4913     \AddBabelHook{luatex}{loadpatterns}{%
4914         \input #1\relax
4915         \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
4916             {{#1}}}%
4917     \AddBabelHook{luatex}{loadexceptions}{%
4918         \input #1\relax
4919         \def\bbl@tempb##1##2{{##1}{#1}}%
4920         \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
4921             {\expandafter\expandafter\expandafter\bbl@tempb
4922                 \csname bbl@hyphendata@the\language\endcsname}}%
4923 \endinput\fi
4924 % Here stops reading code for hyphen.cfg
4925 % The following is read the 2nd time it's loaded
4926 \begingroup % TODO - to a lua file
4927 \catcode`\%=12
4928 \catcode`\'=12
4929 \catcode`\\"=12
4930 \catcode`\:=12
4931 \directlua{
4932     Babel = Babel or {}
4933     function Babel.bytes(line)
4934         return line:gsub("(.)",
4935             function (chr) return unicode.utf8.char(string.byte(chr)) end)
4936     end
4937     function Babel.begin_process_input()
4938         if luatexbase and luatexbase.add_to_callback then
4939             luatexbase.add_to_callback('process_input_buffer',
4940                 Babel.bytes, 'Babel.bytes')
4941         else
4942             Babel.callback = callback.find('process_input_buffer')
4943             callback.register('process_input_buffer', Babel.bytes)
4944         end
4945     end
4946     function Babel.end_process_input ()
4947         if luatexbase and luatexbase.remove_from_callback then
4948             luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
4949         else
4950             callback.register('process_input_buffer', Babel.callback)
4951         end
4952     end
4953     function Babel.addpatterns(pp, lg)
4954         local lg = lang.new(lg)
4955         local pats = lang.patterns(lg) or ''
4956         lang.clear_patterns(lg)
4957         for p in pp:gmatch('[^%s]+') do
4958             ss = ''

```

```

4959     for i in string.utfcharacters(p:gsub('%d', '')) do
4960         ss = ss .. '%d?' .. i
4961     end
4962     ss = ss:gsub('^%%d%?%.', '%%.') .. '%d?'
4963     ss = ss:gsub('%.%d%?$', '%%.')
4964     pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
4965     if n == 0 then
4966         tex.sprint(
4967             [[\string\csname\space bbl@info\endcsname{New pattern: }]]
4968             .. p .. [[{}]])
4969         pats = pats .. ' ' .. p
4970     else
4971         tex.sprint(
4972             [[\string\csname\space bbl@info\endcsname{Renew pattern: }]]
4973             .. p .. [[{}]])
4974     end
4975 end
4976 lang.patterns(lg, pats)
4977 end
4978 function Babel.hlist_has_bidi(head)
4979     local has_bidi = false
4980     for item in node.traverse(head) do
4981         if item.id == node.id'glyph' then
4982             local itemchar = item.char
4983             local chardata = Babel.characters[itemchar]
4984             local dir = chardata and chardata.d or nil
4985             if not dir then
4986                 for nn, et in ipairs(Babel.ranges) do
4987                     if itemchar < et[1] then
4988                         break
4989                     elseif itemchar <= et[2] then
4990                         dir = et[3]
4991                         break
4992                     end
4993                 end
4994             end
4995             if dir and (dir == 'al' or dir == 'r') then
4996                 has_bidi = true
4997             end
4998         end
4999     end
5000     return has_bidi
5001 end
5002 function Babel.set_chranges_b (script, chrng)
5003     if chrng == '' then return end
5004     texio.write('Replacing ' .. script .. ' script ranges')
5005     Babel.script_blocks[script] = {}
5006     for s, e in string.gmatch(chrng..' ', '(-)%.%(-)%s') do
5007         table.insert(
5008             Babel.script_blocks[script], {tonumber(s,16), tonumber(e,16)})
5009     end
5010 end
5011 }
5012 \endgroup
5013 \ifx\newattribute\undefined\else
5014     \newattribute\bbl@attr@locale
5015     \directlua{ Babel.attr_locale = luatexbase.registernumber'bbl@attr@locale' }
5016     \AddBabelHook{luatex}{beforeextras}{%
5017         \setattribute\bbl@attr@locale\localeid}
5018 \fi
5019 \def\BabelStringsDefault{unicode}
5020 \let\luabbl@stop\relax
5021 \AddBabelHook{luatex}{encodedcommands}{%

```

```

5022 \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
5023 \ifx\bbl@tempa\bbl@tempb\else
5024   \directlua{Babel.begin_process_input()}%
5025   \def\luabbl@stop{%
5026     \directlua{Babel.end_process_input()}}%
5027   \fi}%
5028 \AddBabelHook{luatex}{stopcommands}{%
5029   \luabbl@stop
5030   \let\luabbl@stop\relax}
5031 \AddBabelHook{luatex}{patterns}{%
5032   \@ifundefined{bbl@hyphendata@the\language}%
5033   {\def\bbl@elt##1##2##3##4{%
5034     \ifnum##2=\csname l@##2\endcsname % #2=spanish, dutch:OT1...
5035     \def\bbl@tempb{##3}%
5036     \ifx\bbl@tempb\@empty\else % if not a synonymous
5037       \def\bbl@tempc{##3}{##4}}%
5038     \fi
5039     \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
5040     \fi}%
5041   \bbl@languages
5042   \@ifundefined{bbl@hyphendata@the\language}%
5043   {\bbl@info{No hyphenation patterns were set for\%
5044     language '#2'. Reported}}%
5045   {\expandafter\expandafter\expandafter\bbl@luapatterns
5046     \csname bbl@hyphendata@the\language\endcsname}}}%
5047 \@ifundefined{bbl@patterns@}{}%
5048 \begingroup
5049   \bbl@xin@{, \number\language, }{, \bbl@pttnlist}%
5050   \ifin@velse
5051     \ifx\bbl@patterns@\@empty\else
5052       \directlua{ Babel.addpatterns(
5053         [[\bbl@patterns@]], \number\language) }%
5054       \fi
5055       \@ifundefined{bbl@patterns@#1}%
5056       {\@empty
5057         {\directlua{ Babel.addpatterns(
5058           [[\space\csname bbl@patterns@#1\endcsname]],
5059           \number\language) }}%
5060         \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
5061         \fi
5062       \endgroup}%
5063 \bbl@exp{%
5064   \bbl@ifunset{bbl@prehc@languagename}{}%
5065   {\bbl@ifblank{\bbl@cs{prehc@languagename}}}%
5066   {\prehyphenchar=\bbl@cl{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.

```

5067 \@onlypreamble\babelpatterns
5068 \AtEndOfPackage{%
5069   \newcommand\babelpatterns[2][\@empty]{%
5070     \ifx\bbl@patterns@\relax
5071       \let\bbl@patterns@\@empty
5072     \fi
5073     \ifx\bbl@pttnlist@\@empty\else
5074       \bbl@warning{%
5075         You must not intermingle \string\selectlanguage\space and\%
5076         \string\babelpatterns\space or some patterns will not\%
5077         be taken into account. Reported}%
5078       \fi
5079     \ifx\@empty#1%
5080       \protected@edef\bbl@patterns@{\bbl@patterns@\space#2}%

```

```

5081 \else
5082 \edef\bbl@tempb{\zap@space#1 \@empty}%
5083 \bbl@for\bbl@tempa\bbl@tempb{%
5084 \bbl@fixname\bbl@tempa
5085 \bbl@iflanguage\bbl@tempa{%
5086 \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
5087 \@ifundefined{bbl@patterns@\bbl@tempa}%
5088 \@empty
5089 {\csname bbl@patterns@\bbl@tempa\endcsname\space}%
5090 #2}}}%
5091 \fi}}

```

## 12.4 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.

```

5092% TODO - to a lua file
5093\directlua{
5094  Babel = Babel or {}
5095  Babel.linebreaking = Babel.linebreaking or {}
5096  Babel.linebreaking.before = {}
5097  Babel.linebreaking.after = {}
5098  Babel.locale = {} % Free to use, indexed by \localeid
5099  function Babel.linebreaking.add_before(func)
5100    tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
5101    table.insert(Babel.linebreaking.before, func)
5102  end
5103  function Babel.linebreaking.add_after(func)
5104    tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
5105    table.insert(Babel.linebreaking.after, func)
5106  end
5107}
5108\def\bbl@intraspace#1 #2 #3\@@{%
5109 \directlua{
5110   Babel = Babel or {}
5111   Babel.intraspaces = Babel.intraspaces or {}
5112   Babel.intraspaces['\csname bbl@sbc@language\endcsname'] = %
5113     {b = #1, p = #2, m = #3}
5114   Babel.locale_props[\the\localeid].intraspace = %
5115     {b = #1, p = #2, m = #3}
5116 }}
5117\def\bbl@intrapenalty#1\@@{%
5118 \directlua{
5119   Babel = Babel or {}
5120   Babel.intrapenalties = Babel.intrapenalties or {}
5121   Babel.intrapenalties['\csname bbl@sbc@language\endcsname'] = #1
5122   Babel.locale_props[\the\localeid].intrapenalty = #1
5123 }}
5124\begingroup
5125\catcode`\%=12
5126\catcode`\^=14
5127\catcode`\'=12
5128\catcode`\~=12
5129\gdef\bbl@seaintraspace{^
5130 \let\bbl@seaintraspace\relax
5131 \directlua{
5132   Babel = Babel or {}
5133   Babel.sea_enabled = true
5134   Babel.sea_ranges = Babel.sea_ranges or {}
5135   function Babel.set_chranges (script, chrng)
5136     local c = 0

```

```

5137     for s, e in string.gmatch(chrng..' ', '(.-%.(-)%s') do
5138         Babel.sea_ranges[script..c]={tonumber(s,16), tonumber(e,16)}
5139         c = c + 1
5140     end
5141 end
5142 function Babel.sea_disc_to_space (head)
5143     local sea_ranges = Babel.sea_ranges
5144     local last_char = nil
5145     local quad = 655360      ^% 10 pt = 655360 = 10 * 65536
5146     for item in node.traverse(head) do
5147         local i = item.id
5148         if i == node.id'glyph' then
5149             last_char = item
5150         elseif i == 7 and item.subtype == 3 and last_char
5151             and last_char.char > 0x0C99 then
5152             quad = font.getfont(last_char.font).size
5153             for lg, rg in pairs(sea_ranges) do
5154                 if last_char.char > rg[1] and last_char.char < rg[2] then
5155                     lg = lg:sub(1, 4)  ^% Remove trailing number of, eg, Cyril1
5156                     local intraspace = Babel.intraspaces[lg]
5157                     local intrapenalty = Babel.intrapenalties[lg]
5158                     local n
5159                     if intrapenalty ~= 0 then
5160                         n = node.new(14, 0)      ^% penalty
5161                         n.penalty = intrapenalty
5162                         node.insert_before(head, item, n)
5163                     end
5164                     n = node.new(12, 13)      ^% (glue, spaceskip)
5165                     node.setglue(n, intraspace.b * quad,
5166                                 intraspace.p * quad,
5167                                 intraspace.m * quad)
5168                     node.insert_before(head, item, n)
5169                     node.remove(head, item)
5170                 end
5171             end
5172         end
5173     end
5174 end
5175 }^^
5176 \bbl@luahyphenate}

```

## 12.5 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.

```

5177 \catcode`\%=14
5178 \gdef\bbl@cjkintraspaces{%
5179     \let\bbl@cjkintraspaces\relax
5180     \directlua{
5181         Babel = Babel or {}
5182         require('babel-data-cjk.lua')
5183         Babel.cjk_enabled = true
5184         function Babel.cjk_linebreak(head)
5185             local GLYPH = node.id'glyph'
5186             local last_char = nil
5187             local quad = 655360      % 10 pt = 655360 = 10 * 65536
5188             local last_class = nil
5189             local last_lang = nil
5190

```

```

5191     for item in node.traverse(head) do
5192         if item.id == GLYPH then
5193
5194             local lang = item.lang
5195
5196             local LOCALE = node.get_attribute(item,
5197                 Babel.attr_locale)
5198             local props = Babel.locale_props[LOCALE]
5199
5200             local class = Babel.cjk_class[item.char].c
5201
5202             if props.cjk_quotes and props.cjk_quotes[item.char] then
5203                 class = props.cjk_quotes[item.char]
5204             end
5205
5206             if class == 'cp' then class = 'cl' end % ]] as CL
5207             if class == 'id' then class = 'I' end
5208
5209             local br = 0
5210             if class and last_class and Babel.cjk_breaks[last_class][class] then
5211                 br = Babel.cjk_breaks[last_class][class]
5212             end
5213
5214             if br == 1 and props.linebreak == 'c' and
5215                 lang ~= \the\l@nohyphenation\space and
5216                 last_lang ~= \the\l@nohyphenation then
5217                 local intrapenalty = props.intrapenalty
5218                 if intrapenalty ~= 0 then
5219                     local n = node.new(14, 0)      % penalty
5220                     n.penalty = intrapenalty
5221                     node.insert_before(head, item, n)
5222                 end
5223                 local intraspace = props.intraspace
5224                 local n = node.new(12, 13)        % (glue, spaceskip)
5225                 node.setglue(n, intraspace.b * quad,
5226                     intraspace.p * quad,
5227                     intraspace.m * quad)
5228                 node.insert_before(head, item, n)
5229             end
5230
5231             if font.getfont(item.font) then
5232                 quad = font.getfont(item.font).size
5233             end
5234             last_class = class
5235             last_lang = lang
5236         else % if penalty, glue or anything else
5237             last_class = nil
5238         end
5239     end
5240     lang.hyphenate(head)
5241 end
5242 }%
5243 \bbl@luahyphenate}
5244 \gdef\bbl@luahyphenate{%
5245 \let\bbl@luahyphenate\relax
5246 \directlua{
5247     luatexbase.add_to_callback('hyphenate',
5248     function (head, tail)
5249         if Babel.linebreaking.before then
5250             for k, func in ipairs(Babel.linebreaking.before) do
5251                 func(head)
5252             end
5253         end

```



```

5254     if Babel.cjk_enabled then
5255         Babel.cjk_linebreak(head)
5256     end
5257     lang.hyphenate(head)
5258     if Babel.linebreaking.after then
5259         for k, func in ipairs(Babel.linebreaking.after) do
5260             func(head)
5261         end
5262     end
5263     if Babel.sea_enabled then
5264         Babel.sea_disc_to_space(head)
5265     end
5266 end,
5267 'Babel.hyphenate')
5268 }
5269 }
5270 \endgroup
5271 \def\bbl@provide@intraspace{%
5272   \bbl@ifunset\bbl@intsp@\language\language\{}%
5273   {\expandafter\ifx\csname\bbl@intsp@\language\endcsname\@empty\else
5274     \bbl@xin{/c}{/\bbl@cl{lnbrk}}}%
5275   \ifin@           % cjk
5276   \bbl@cjk@intraspace
5277   \directlua{
5278     Babel = Babel or {}
5279     Babel.locale_props = Babel.locale_props or {}
5280     Babel.locale_props[\the\localeid].linebreak = 'c'
5281   }%
5282   \bbl@exp{\bbl@intraspace\bbl@cl{intsp}\@}%
5283   \ifx\bbl@KVP@intrapenalty\@nil
5284     \bbl@intrapenalty0\@
5285   \fi
5286   \else           % sea
5287   \bbl@sea@intraspace
5288   \bbl@exp{\bbl@intraspace\bbl@cl{intsp}\@}%
5289   \directlua{
5290     Babel = Babel or {}
5291     Babel.sea_ranges = Babel.sea_ranges or {}
5292     Babel.set_chranges('bbl@cl{sbc}',
5293                       'bbl@cl{chrng}')
5294   }%
5295   \ifx\bbl@KVP@intrapenalty\@nil
5296     \bbl@intrapenalty0\@
5297   \fi
5298   \fi
5299   \fi
5300   \ifx\bbl@KVP@intrapenalty\@nil\else
5301     \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@
5302   \fi}}

```

## 12.6 Arabic justification

```

5303 \ifnum\bbl@bidimode>100 \ifnum\bbl@bidimode<200
5304 \def\bblar@chars{%
5305   0628,0629,062A,062B,062C,062D,062E,062F,0630,0631,0632,0633,%
5306   0634,0635,0636,0637,0638,0639,063A,063B,063C,063D,063E,063F,%
5307   0640,0641,0642,0643,0644,0645,0646,0647,0649}
5308 \def\bblar@elongated{%
5309   0626,0628,062A,062B,0633,0634,0635,0636,063B,%
5310   063C,063D,063E,063F,0641,0642,0643,0644,0646,%
5311   0649,064A}
5312 \begin{group}
5313   \catcode`_ =11 \catcode`:=11

```

```

5314 \gdef\bblar@nofswarn{\gdef\msg_warning:nx##1##2##3{}}
5315 \endgroup
5316 \gdef\bbl@arabicjust{%
5317   \let\bbl@arabicjust\relax
5318   \newattribute\bblar@kashida
5319   \directlua{ Babel.attr_kashida = luatexbase.registernumber'bblar@kashida' }%
5320   \bblar@kashida=\z@
5321   \bbl@patchfont{{\bbl@parsejalt}}}%
5322   \directlua{
5323     Babel.arabic.elong_map = Babel.arabic.elong_map or {}
5324     Babel.arabic.elong_map[\the\localeid] = {}
5325     luatexbase.add_to_callback('post_linebreak_filter',
5326       Babel.arabic.justify, 'Babel.arabic.justify')
5327     luatexbase.add_to_callback('hpack_filter',
5328       Babel.arabic.justify_hbox, 'Babel.arabic.justify_hbox')
5329   }}%
5330 % Save both node lists to make replacement. TODO. Save also widths to
5331 % make computations
5332 \def\bblar@fetchjalt#1#2#3#4{%
5333   \bbl@exp{\bbl@foreach{#1}}{%
5334     \bbl@ifunset{bblar@JE@##1}%
5335     {\setbox\z@\hbox{^^^^200d\char"##1#2}}%
5336     {\setbox\z@\hbox{^^^^200d\char"@nameuse{bblar@JE@##1}#2}}%
5337     \directlua{%
5338       local last = nil
5339       for item in node.traverse(tex.box[0].head) do
5340         if item.id == node.id'glyph' and item.char > 0x600 and
5341           not (item.char == 0x200D) then
5342           last = item
5343         end
5344       end
5345       Babel.arabic.#3['##1#4'] = last.char
5346     }}%
5347 % Brute force. No rules at all, yet. The ideal: look at jalt table. And
5348 % perhaps other tables (falt?, csw?). What about kaf? And diacritic
5349 % positioning?
5350 \gdef\bbl@parsejalt{%
5351   \ifx\addfontfeature\undefined\else
5352     \bbl@xin@{/e}{/\bbl@c1{lbrk}}%
5353     \ifin@
5354       \directlua{%
5355         if Babel.arabic.elong_map[\the\localeid][\fontid\font] == nil then
5356           Babel.arabic.elong_map[\the\localeid][\fontid\font] = {}
5357           tex.print([[string\csname\space bbl@parsejalti\endcsname]])
5358         end
5359       }%
5360     \fi
5361   \fi}
5362 \gdef\bbl@parsejalti{%
5363   \begingroup
5364     \let\bbl@parsejalt\relax % To avoid infinite loop
5365     \edef\bbl@tempb{\fontid\font}%
5366     \bblar@nofswarn
5367     \bblar@fetchjalt\bblar@elongated{}{from}}%
5368     \bblar@fetchjalt\bblar@chars{^^^^064a}{from}{a}% Alef maksura
5369     \bblar@fetchjalt\bblar@chars{^^^^0649}{from}{y}% Yeh
5370     \addfontfeature{RawFeature=+jalt}%
5371     % \@namedef{bblar@JE@0643}{06AA}% todo: catch medial kaf
5372     \bblar@fetchjalt\bblar@elongated{}{dest}}%
5373     \bblar@fetchjalt\bblar@chars{^^^^064a}{dest}{a}%
5374     \bblar@fetchjalt\bblar@chars{^^^^0649}{dest}{y}%
5375     \directlua{%
5376       for k, v in pairs(Babel.arabic.from) do

```

```

5377         if Babel.arabic.dest[k] and
5378             not (Babel.arabic.from[k] == Babel.arabic.dest[k]) then
5379             Babel.arabic.elong_map[\the\localeid][\bbl@tempb]
5380             [Babel.arabic.from[k]] = Babel.arabic.dest[k]
5381         end
5382     end
5383 }%
5384 \endgroup}
5385 %
5386 \begingroup
5387 \catcode`#=11
5388 \catcode`~=11
5389 \directlua{
5390
5391 Babel.arabic = Babel.arabic or {}
5392 Babel.arabic.from = {}
5393 Babel.arabic.dest = {}
5394 Babel.arabic.justify_factor = 0.95
5395 Babel.arabic.justify_enabled = true
5396
5397 function Babel.arabic.justify(head)
5398     if not Babel.arabic.justify_enabled then return head end
5399     for line in node.traverse_id(node.id'hlist', head) do
5400         Babel.arabic.justify_hlist(head, line)
5401     end
5402     return head
5403 end
5404
5405 function Babel.arabic.justify_hbox(head, gc, size, pack)
5406     local has_inf = false
5407     if Babel.arabic.justify_enabled and pack == 'exactly' then
5408         for n in node.traverse_id(12, head) do
5409             if n.stretch_order > 0 then has_inf = true end
5410         end
5411         if not has_inf then
5412             Babel.arabic.justify_hlist(head, nil, gc, size, pack)
5413         end
5414     end
5415     return head
5416 end
5417
5418 function Babel.arabic.justify_hlist(head, line, gc, size, pack)
5419     local d, new
5420     local k_list, k_item, pos_inline
5421     local width, width_new, full, k_curr, wt_pos, goal, shift
5422     local subst_done = false
5423     local elong_map = Babel.arabic.elong_map
5424     local last_line
5425     local GLYPH = node.id'glyph'
5426     local KASHIDA = Babel.attr_kashida
5427     local LOCALE = Babel.attr_locale
5428
5429     if line == nil then
5430         line = {}
5431         line.glue_sign = 1
5432         line.glue_order = 0
5433         line.head = head
5434         line.shift = 0
5435         line.width = size
5436     end
5437
5438     % Exclude last line. todo. But-- it discards one-word lines, too!
5439     % ? Look for glue = 12:15

```

```

5440 if (line.glue_sign == 1 and line.glue_order == 0) then
5441     elongs = {}      % Stores elongated candidates of each line
5442     k_list = {}      % And all letters with kashida
5443     pos_inline = 0   % Not yet used
5444
5445     for n in node.traverse_id(GLYPH, line.head) do
5446         pos_inline = pos_inline + 1 % To find where it is. Not used.
5447
5448         % Elongated glyphs
5449         if elong_map then
5450             local locale = node.get_attribute(n, LOCALE)
5451             if elong_map[locale] and elong_map[locale][n.font] and
5452                 elong_map[locale][n.font][n.char] then
5453                 table.insert(elongs, {node = n, locale = locale} )
5454                 node.set_attribute(n.prev, KASHIDA, 0)
5455             end
5456         end
5457
5458         % Tatwil
5459         if Babel.kashida_wts then
5460             local k_wt = node.get_attribute(n, KASHIDA)
5461             if k_wt > 0 then % todo. parameter for multi inserts
5462                 table.insert(k_list, {node = n, weight = k_wt, pos = pos_inline})
5463             end
5464         end
5465
5466     end % of node.traverse_id
5467
5468     if #elongs == 0 and #k_list == 0 then goto next_line end
5469     full = line.width
5470     shift = line.shift
5471     goal = full * Babel.arabic.justify_factor % A bit crude
5472     width = node.dimensions(line.head)      % The 'natural' width
5473
5474     % == Elongated ==
5475     % Original idea taken from 'chickenize'
5476     while (#elongs > 0 and width < goal) do
5477         subst_done = true
5478         local x = #elongs
5479         local curr = elongs[x].node
5480         local oldchar = curr.char
5481         curr.char = elong_map[elongs[x].locale][curr.font][curr.char]
5482         width = node.dimensions(line.head) % Check if the line is too wide
5483         % Substitute back if the line would be too wide and break:
5484         if width > goal then
5485             curr.char = oldchar
5486             break
5487         end
5488         % If continue, pop the just substituted node from the list:
5489         table.remove(elongs, x)
5490     end
5491
5492     % == Tatwil ==
5493     if #k_list == 0 then goto next_line end
5494
5495     width = node.dimensions(line.head)      % The 'natural' width
5496     k_curr = #k_list
5497     wt_pos = 1
5498
5499     while width < goal do
5500         subst_done = true
5501         k_item = k_list[k_curr].node
5502         if k_list[k_curr].weight == Babel.kashida_wts[wt_pos] then

```

```

5503     d = node.copy(k_item)
5504     d.char = 0x0640
5505     line.head, new = node.insert_after(line.head, k_item, d)
5506     width_new = node.dimensions(line.head)
5507     if width > goal or width == width_new then
5508         node.remove(line.head, new) % Better compute before
5509         break
5510     end
5511     width = width_new
5512 end
5513 if k_curr == 1 then
5514     k_curr = #k_list
5515     wt_pos = (wt_pos >= table.getn(Babel.kashida_wts)) and 1 or wt_pos+1
5516 else
5517     k_curr = k_curr - 1
5518 end
5519 end
5520
5521 ::next_line::
5522
5523 % Must take into account marks and ins, see luatex manual.
5524 % Have to be executed only if there are changes. Investigate
5525 % what's going on exactly.
5526 if subst_done and not gc then
5527     d = node.hpack(line.head, full, 'exactly')
5528     d.shift = shift
5529     node.insert_before(head, line, d)
5530     node.remove(head, line)
5531 end
5532 end % if process line
5533 end
5534 }
5535 \endgroup
5536 \fi\fi % Arabic just block

```

## 12.7 Common stuff

```

5537 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
5538 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@cckstdfont}
5539 \DisableBabelHook{babel-fontspec}
5540 <{Font selection}>

```

## 12.8 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.

```

5541 % TODO - to a lua file
5542 \directlua{
5543 Babel.script_blocks = {
5544   ['dflt'] = {},
5545   ['Arab'] = {{0x0600, 0x06FF}, {0x08A0, 0x08FF}, {0x0750, 0x077F},
5546               {0xFE70, 0xFEFF}, {0xFB50, 0xFDFF}, {0x1EE00, 0x1EEFF}},
5547   ['Armn'] = {{0x0530, 0x058F}},
5548   ['Beng'] = {{0x0980, 0x09FF}},
5549   ['Cher'] = {{0x13A0, 0x13FF}, {0xAB70, 0xABBF}},
5550   ['Copt'] = {{0x03E2, 0x03EF}, {0x2C80, 0x2CFF}, {0x102E0, 0x102FF}},
5551   ['Cyr1'] = {{0x0400, 0x04FF}, {0x0500, 0x052F}, {0x1C80, 0x1C8F},
5552               {0x2DE0, 0x2DFF}, {0xA640, 0xA69F}},
5553   ['Deva'] = {{0x0900, 0x097F}, {0xA8E0, 0xA8FF}},

```

```

5554 ['Ethi'] = {{0x1200, 0x137F}, {0x1380, 0x139F}, {0x2D80, 0x2DDF},
5555             {0xAB00, 0xAB2F}},
5556 ['Geor'] = {{0x10A0, 0x10FF}, {0x2D00, 0x2D2F}},
5557 % Don't follow strictly Unicode, which places some Coptic letters in
5558 % the 'Greek and Coptic' block
5559 ['Grek'] = {{0x0370, 0x03E1}, {0x03F0, 0x03FF}, {0x1F00, 0x1FFF}},
5560 ['Hans'] = {{0x2E80, 0x2EFF}, {0x3000, 0x303F}, {0x31C0, 0x31EF},
5561             {0x3300, 0x33FF}, {0x3400, 0x4DBF}, {0x4E00, 0x9FFF},
5562             {0xF900, 0xFAFF}, {0xFE30, 0xFE4F}, {0xFF00, 0xFFEF},
5563             {0x20000, 0x2A6DF}, {0x2A700, 0x2B73F},
5564             {0x2B740, 0x2B81F}, {0x2B820, 0x2CEAF},
5565             {0x2CEB0, 0x2EBEF}, {0x2F800, 0x2FA1F}},
5566 ['Hebr'] = {{0x0590, 0x05FF}},
5567 ['Jpan'] = {{0x3000, 0x303F}, {0x3040, 0x309F}, {0x30A0, 0x30FF},
5568             {0x4E00, 0x9FAF}, {0xFF00, 0xFFEF}},
5569 ['Khmr'] = {{0x1780, 0x17FF}, {0x19E0, 0x19FF}},
5570 ['Knda'] = {{0x0C80, 0x0CFF}},
5571 ['Kore'] = {{0x1100, 0x11FF}, {0x3000, 0x303F}, {0x3130, 0x318F},
5572             {0x4E00, 0x9FAF}, {0xA960, 0xA97F}, {0xAC00, 0xD7AF},
5573             {0xD7B0, 0xD7FF}, {0xFF00, 0xFFEF}},
5574 ['Lao'] = {{0x0E80, 0x0EFF}},
5575 ['Latn'] = {{0x0000, 0x007F}, {0x0080, 0x00FF}, {0x0100, 0x017F},
5576             {0x0180, 0x024F}, {0x1E00, 0x1EFF}, {0x2C60, 0x2C7F},
5577             {0xA720, 0xA7FF}, {0xAB30, 0xAB6F}},
5578 ['Mahj'] = {{0x11150, 0x1117F}},
5579 ['Mlym'] = {{0x0D00, 0x0D7F}},
5580 ['Mymr'] = {{0x1000, 0x109F}, {0xAA60, 0xAA7F}, {0xA9E0, 0xA9FF}},
5581 ['Orya'] = {{0x0B00, 0x0B7F}},
5582 ['Sinh'] = {{0x0D80, 0x0DFF}, {0x111E0, 0x111FF}},
5583 ['Sycr'] = {{0x0700, 0x074F}, {0x0860, 0x086F}},
5584 ['Taml'] = {{0x0B80, 0x0BFF}},
5585 ['Telu'] = {{0x0C00, 0x0C7F}},
5586 ['Tfng'] = {{0x2D30, 0x2D7F}},
5587 ['Thai'] = {{0x0E00, 0x0E7F}},
5588 ['Tibt'] = {{0x0F00, 0x0FFF}},
5589 ['Vaii'] = {{0xA500, 0xA63F}},
5590 ['Yiii'] = {{0xA000, 0xA48F}, {0xA490, 0xA4CF}}
5591 }
5592
5593 Babel.script_blocks.Cyrs = Babel.script_blocks.Cyrl
5594 Babel.script_blocks.Hant = Babel.script_blocks.Hans
5595 Babel.script_blocks.Kana = Babel.script_blocks.Jpan
5596
5597 function Babel.locale_map(head)
5598   if not Babel.locale_mapped then return head end
5599
5600   local LOCALE = Babel.attr_locale
5601   local GLYPH = node.id('glyph')
5602   local inmath = false
5603   local toloc_save
5604   for item in node.traverse(head) do
5605     local toloc
5606     if not inmath and item.id == GLYPH then
5607       % Optimization: build a table with the chars found
5608       if Babel.chr_to_loc[item.char] then
5609         toloc = Babel.chr_to_loc[item.char]
5610       else
5611         for lc, maps in pairs(Babel.loc_to_scr) do
5612           for _, rg in pairs(maps) do
5613             if item.char >= rg[1] and item.char <= rg[2] then
5614               Babel.chr_to_loc[item.char] = lc
5615               toloc = lc
5616               break

```

```

5617         end
5618     end
5619 end
5620 end
5621 % Now, take action, but treat composite chars in a different
5622 % fashion, because they 'inherit' the previous locale. Not yet
5623 % optimized.
5624 if not toloc and
5625     (item.char >= 0x0300 and item.char <= 0x036F) or
5626     (item.char >= 0x1AB0 and item.char <= 0x1AFF) or
5627     (item.char >= 0x1DC0 and item.char <= 0x1DFF) then
5628     toloc = toloc_save
5629 end
5630 if toloc and toloc > -1 then
5631     if Babel.locale_props[toloc].lg then
5632         item.lang = Babel.locale_props[toloc].lg
5633         node.set_attribute(item, LOCALE, toloc)
5634     end
5635     if Babel.locale_props[toloc]['/'..item.font] then
5636         item.font = Babel.locale_props[toloc]['/'..item.font]
5637     end
5638     toloc_save = toloc
5639 end
5640 elseif not inmath and item.id == 7 then
5641     item.replace = item.replace and Babel.locale_map(item.replace)
5642     item.pre      = item.pre and Babel.locale_map(item.pre)
5643     item.post     = item.post and Babel.locale_map(item.post)
5644 elseif item.id == node.id'math' then
5645     inmath = (item.subtype == 0)
5646 end
5647 end
5648 return head
5649 end
5650 }

```

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

```

5651 \newcommand\babelcharproperty[1]{%
5652   \count@=#1\relax
5653   \ifvmode
5654     \expandafter\bbl@chprop
5655   \else
5656     \bbl@error{\string\babelcharproperty\space can be used only in\%
5657       vertical mode (preamble or between paragraphs)}%
5658     {See the manual for futher info}%
5659   \fi}
5660 \newcommand\bbl@chprop[3][\the\count@]{%
5661   \@tempcnta=#1\relax
5662   \bbl@ifunset\bbl@chprop@#2}%
5663   {\bbl@error{No property named '#2'. Allowed values are\%
5664     direction (bc), mirror (bmg), and linebreak (lb)}%
5665     {See the manual for futher info}}%
5666   {%
5667   \loop
5668     \bbl@cs{chprop@#2}{#3}%
5669     \ifnum\count@<\@tempcnta
5670       \advance\count@\@ne
5671     \repeat}
5672 \def\bbl@chprop@direction#1{%
5673   \directlua{
5674     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
5675     Babel.characters[\the\count@]['d'] = '#1'
5676   }}

```

```

5677 \let\bbl@chprop@bc\bbl@chprop@direction
5678 \def\bbl@chprop@mirror#1{%
5679   \directlua{
5680     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
5681     Babel.characters[\the\count@]['m'] = '\number#1'
5682   }}
5683 \let\bbl@chprop@bmg\bbl@chprop@mirror
5684 \def\bbl@chprop@linebreak#1{%
5685   \directlua{
5686     Babel.cjk_characters[\the\count@] = Babel.cjk_characters[\the\count@] or {}
5687     Babel.cjk_characters[\the\count@]['c'] = '#1'
5688   }}
5689 \let\bbl@chprop@lb\bbl@chprop@linebreak
5690 \def\bbl@chprop@locale#1{%
5691   \directlua{
5692     Babel.chr_to_loc = Babel.chr_to_loc or {}
5693     Babel.chr_to_loc[\the\count@] =
5694       \bbl@ifblank{#1}{-1000}{\the\bbl@cs{id@#1}}\space
5695   }}

```

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). The Lua code is below.

```

5696 \directlua{
5697   Babel.nohyphenation = \the\l@nohyphenation
5698 }

```

Now the T<sub>E</sub>X 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, 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).

```

5699 \begingroup
5700 \catcode`\~ = 12
5701 \catcode`\% = 12
5702 \catcode`\& = 14
5703 \catcode`\| = 12
5704 \gdef\babelprehyphenation{%&
5705   \@ifnextchar[{\bbl@settransform{0}}{\bbl@settransform{0}}{]}
5706 \gdef\babelposthyphenation{%&
5707   \@ifnextchar[{\bbl@settransform{1}}{\bbl@settransform{1}}{]}
5708 \gdef\bbl@settransform#1[#2]#3#4#5{%&
5709   \ifcase#1
5710     \bbl@activateprehyphen
5711   \else
5712     \bbl@activateposthyphen
5713   \fi
5714 \begingroup
5715   \def\babeltempa{\bbl@add@list\babeltempb}%&
5716   \let\babeltempb\@empty
5717   \def\bbl@tempa{#5}%&
5718   \bbl@replace\bbl@tempa{,}{ ,}%& TODO. Ugly trick to preserve {}
5719   \expandafter\bbl@foreach\expandafter{\bbl@tempa}{%&
5720     \bbl@ifsamestring{##1}{remove}%&
5721     {\bbl@add@list\babeltempb{nil}}}%&
5722     {\directlua{
5723       local rep = {[#1]}
5724       rep = rep:gsub('^%s*(remove)%s*$', 'remove = true')
5725       rep = rep:gsub('^%s*(insert)%s*', 'insert = true, ')
5726       rep = rep:gsub('(string)%s*=%s*([%s,]*)', Babel.capture_func)
5727       if #1 == 0 then

```



```

5728         rep = rep:gsub('(space)%s*=%s*([%d%.]+)%s+([%d%.]+)%s+([%d%.]+)',
5729         'space = {' .. '%2, %3, %4' .. '}')
5730         rep = rep:gsub('(spacefactor)%s*=%s*([%d%.]+)%s+([%d%.]+)%s+([%d%.]+)',
5731         'spacefactor = {' .. '%2, %3, %4' .. '}')
5732         rep = rep:gsub('(kashida)%s*=%s*([^\s,]*)', Babel.capture_kashida)
5733     else
5734         rep = rep:gsub(' (no)%s*=%s*([^\s,]*)', Babel.capture_func)
5735         rep = rep:gsub(' (pre)%s*=%s*([^\s,]*)', Babel.capture_func)
5736         rep = rep:gsub(' (post)%s*=%s*([^\s,]*)', Babel.capture_func)
5737     end
5738     tex.print([[string\babeltempa{[]} .. rep .. [{}]])
5739     }&%
5740 \let\bbl@kv@attribute\relax
5741 \let\bbl@kv@label\relax
5742 \bbl@forkv{#2}{\bbl@csarg\edef{kv@##1}{##2}}&%
5743 \ifx\bbl@kv@attribute\relax\else
5744     \edef\bbl@kv@attribute{\expandafter\bbl@stripslash\bbl@kv@attribute}&%
5745 \fi
5746 \directlua{
5747     local lbkr = Babel.linebreaking.replacements[#1]
5748     local u = unicode.utf8
5749     local id, attr, label
5750     if #1 == 0 then
5751         id = \the\csname bbl@id@#3\endcsname\space
5752     else
5753         id = \the\csname l@#3\endcsname\space
5754     end
5755     \ifx\bbl@kv@attribute\relax
5756         attr = -1
5757     \else
5758         attr = luatexbase.registernumber'\bbl@kv@attribute'
5759     \fi
5760     \ifx\bbl@kv@label\relax\else    %% Same refs:
5761         label = [==[\bbl@kv@label]==]
5762     \fi
5763     %% Convert pattern:
5764     local patt = string.gsub([==[#4]==], '%s', '')
5765     if #1 == 0 then
5766         patt = string.gsub(patt, '|', ' ')
5767     end
5768     if not u.find(patt, '()', nil, true) then
5769         patt = '()' .. patt .. '()'
5770     end
5771     if #1 == 1 then
5772         patt = string.gsub(patt, '%(%)^', '^()')
5773         patt = string.gsub(patt, '%$%(%)', '()$')
5774     end
5775     patt = u.gsub(patt, '{(.)}',
5776     function (n)
5777         return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
5778     end)
5779     patt = u.gsub(patt, '{(%x%x%x%x%+)}',
5780     function (n)
5781         return u.gsub(u.char(tonumber(n, 16)), '(%p)', '%%1')
5782     end)
5783     lbkr[id] = lbkr[id] or {}
5784     table.insert(lbkr[id],
5785     { label=label, attr=attr, pattern=patt, replace={\babeltempb} })
5786 }&%
5787 \endgroup}
5788 \endgroup
5789 \def\bbl@activateposthyphen{%
5790     \let\bbl@activateposthyphen\relax

```

```

5791 \directlua{
5792     require('babel-transforms.lua')
5793     Babel.linebreaking.add_after(Babel.post_hyphenate_replace)
5794 }}
5795 \def\bbl@activateprehyphen{%
5796     \let\bbl@activateprehyphen\relax
5797     \directlua{
5798         require('babel-transforms.lua')
5799         Babel.linebreaking.add_before(Babel.pre_hyphenate_replace)
5800     }}

```

## 12.9 Bidi

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 L<sup>A</sup>T<sub>E</sub>X. Just in case, consider the possibility it has not been loaded.

```

5801 \def\bbl@activate@preotf{%
5802     \let\bbl@activate@preotf\relax % only once
5803     \directlua{
5804         Babel = Babel or {}
5805         %
5806         function Babel.pre_otfload_v(head)
5807             if Babel.numbers and Babel.digits_mapped then
5808                 head = Babel.numbers(head)
5809             end
5810             if Babel.bidi_enabled then
5811                 head = Babel.bidi(head, false, dir)
5812             end
5813             return head
5814         end
5815         %
5816         function Babel.pre_otfload_h(head, gc, sz, pt, dir)
5817             if Babel.numbers and Babel.digits_mapped then
5818                 head = Babel.numbers(head)
5819             end
5820             if Babel.bidi_enabled then
5821                 head = Babel.bidi(head, false, dir)
5822             end
5823             return head
5824         end
5825         %
5826         luatexbase.add_to_callback('pre_linebreak_filter',
5827             Babel.pre_otfload_v,
5828             'Babel.pre_otfload_v',
5829             luatexbase.priority_in_callback('pre_linebreak_filter',
5830                 'luaotfload.node_processor') or nil)
5831         %
5832         luatexbase.add_to_callback('hpack_filter',
5833             Babel.pre_otfload_h,
5834             'Babel.pre_otfload_h',
5835             luatexbase.priority_in_callback('hpack_filter',
5836                 'luaotfload.node_processor') or nil)
5837     }}

```

The basic setup. The output is modified at a very low level to set the `\bodydir` to the `\pagedir`. Sadly, we have to deal with boxes in math with basic, so the `\bbl@mathboxdir` hack is activated every math with the package option `bidi=`.

```

5838 \ifnum\bbl@bidimode>100 \ifnum\bbl@bidimode<200
5839     \let\bbl@beforeforeign\leavevmode
5840     \AtEndOfPackage{\EnableBabelHook{babel-bidi}}
5841     \RequirePackage{luatexbase}
5842     \bbl@activate@preotf
5843     \directlua{

```

```

5844     require('babel-data-bidi.lua')
5845     \ifcase\expandafter\@gobbletwo\the\bbl@bidimode\or
5846     require('babel-bidi-basic.lua')
5847     \or
5848     require('babel-bidi-basic-r.lua')
5849     \fi}
5850 % TODO - to locale_props, not as separate attribute
5851 \newattribute\bbl@attr@dir
5852 \directlua{ Babel.attr_dir = luatexbase.registernumber'bbl@attr@dir' }
5853 % TODO. I don't like it, hackish:
5854 \bbl@exp{\output{\bodydir\pagedir\the\output}}
5855 \AtEndOfPackage{\EnableBabelHook{babel-bidi}}
5856 \fi\fi
5857 \chardef\bbl@thetextdir\z@
5858 \chardef\bbl@thepardir\z@
5859 \def\bbl@getluadir#1{%
5860   \directlua{
5861     if tex.#1dir == 'TLT' then
5862       tex.sprint('0')
5863     elseif tex.#1dir == 'TRT' then
5864       tex.sprint('1')
5865     end}}
5866 \def\bbl@setluadir#1#2#3{% 1=text/par.. 2=\textdir.. 3=0 lr/1 rl
5867   \ifcase#3\relax
5868     \ifcase\bbl@getluadir{#1}\relax\else
5869       #2 TLT\relax
5870     \fi
5871   \else
5872     \ifcase\bbl@getluadir{#1}\relax
5873       #2 TRT\relax
5874     \fi
5875   \fi}
5876 \def\bbl@thedir{0}
5877 \def\bbl@textdir#1{%
5878   \bbl@setluadir{text}\textdir{#1}%
5879   \chardef\bbl@thetextdir#1\relax
5880   \edef\bbl@thedir{\the\numexpr\bbl@thepardir*3+#1}%
5881   \setattribute\bbl@attr@dir{\numexpr\bbl@thepardir*3+#1}}
5882 \def\bbl@pardir#1{%
5883   \bbl@setluadir{par}\pardir{#1}%
5884   \chardef\bbl@thepardir#1\relax}
5885 \def\bbl@bodydir{\bbl@setluadir{body}\bodydir}
5886 \def\bbl@pagedir{\bbl@setluadir{page}\pagedir}
5887 \def\bbl@dirparastext{\pardir\the\textdir\relax}%   %%%
5888 %
5889 \ifnum\bbl@bidimode>\z@
5890   \def\bbl@insidemath{0}%
5891   \def\bbl@everymath{\def\bbl@insidemath{1}}
5892   \def\bbl@everydisplay{\def\bbl@insidemath{2}}
5893   \frozen@everymath\expandafter{%
5894     \expandafter\bbl@everymath\the\frozen@everymath}
5895   \frozen@everydisplay\expandafter{%
5896     \expandafter\bbl@everydisplay\the\frozen@everydisplay}
5897   \AtBeginDocument{
5898     \directlua{
5899       function Babel.math_box_dir(head)
5900         if not (token.get_macro('bbl@insidemath') == '0') then
5901           if Babel.hlist_has_bidi(head) then
5902             local d = node.new(node.id'dir')
5903             d.dir = '+TRT'
5904             node.insert_before(head, node.has_glyph(head), d)
5905             for item in node.traverse(head) do
5906               node.set_attribute(item,

```

```

5907             Babel.attr_dir, token.get_macro('bbl@thedir'))
5908         end
5909     end
5910 end
5911 return head
5912 end
5913 luatexbase.add_to_callback("hpack_filter", Babel.math_box_dir,
5914     "Babel.math_box_dir", 0)
5915 }}%
5916 \fi

```

## 12.10 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.

```

5917 \bbl@trace{Redefinitions for bidi layout}
5918 %
5919 <<(*More package options)>> ≡
5920 \chardef\bbl@eqnpos\z@
5921 \DeclareOption{leqno}{\chardef\bbl@eqnpos\@ne}
5922 \DeclareOption{fleqn}{\chardef\bbl@eqnpos\@tw@}
5923 <</More package options>>
5924 %
5925 \def\BabelNoAMSMath{\let\bbl@noamsmath\relax}
5926 \ifnum\bbl@bidimode>\z@
5927   \ifx\matheqdirmode\undefined\else
5928     \matheqdirmode\@ne
5929   \fi
5930   \let\bbl@eqnodir\relax
5931   \def\bbl@eqdel{()}
5932   \def\bbl@eqnum{%
5933     {\normalfont\normalcolor
5934       \expandafter\@firstoftwo\bbl@eqdel
5935       \theequation
5936       \expandafter\@secondoftwo\bbl@eqdel}}
5937   \def\bbl@puteqno#1{\leqno\hbox{#1}}
5938   \def\bbl@putleqno#1{\leqno\hbox{#1}}
5939   \def\bbl@eqno@flip#1{%
5940     \ifdim\predisplaysize=-\maxdimen
5941       \leqno
5942       \hb@xt@.01pt{\hb@xt@\displaywidth{\hss{#1}}\hss}%
5943     \else
5944       \leqno\hbox{#1}%
5945     \fi}
5946   \def\bbl@leqno@flip#1{%
5947     \ifdim\predisplaysize=-\maxdimen
5948       \leqno
5949       \hb@xt@.01pt{\hss\hb@xt@\displaywidth{{#1}\hss}}%
5950     \else
5951       \leqno\hbox{#1}%
5952     \fi}
5953   \AtBeginDocument{%
5954     \ifx\maketag@@@\undefined % Normal equation, eqnarray

```

```

5955 \AddToHook{env/equation/begin}{%
5956   \ifnum\bb1@thetextdir>\z@
5957     \let\@eqnnum\bb1@eqnum
5958     \edef\bb1@eqnodir{\noexpand\bb1@textdir{\the\bb1@thetextdir}}%
5959     \chardef\bb1@thetextdir\z@
5960     \bb1@add\normalfont{\bb1@eqnodir}%
5961     \ifcase\bb1@eqnpos
5962       \let\bb1@puteqno\bb1@eqno@flip
5963     \or
5964       \let\bb1@puteqno\bb1@leqno@flip
5965     \fi
5966   \fi}%
5967 \ifnum\bb1@eqnpos=\tw@%else
5968   \def\endequation{\bb1@puteqno{\@eqnnum}$$\@ignoretrue}%
5969 \fi
5970 \AddToHook{env/eqnarray/begin}{%
5971   \ifnum\bb1@thetextdir>\z@
5972     \edef\bb1@eqnodir{\noexpand\bb1@textdir{\the\bb1@thetextdir}}%
5973     \chardef\bb1@thetextdir\z@
5974     \bb1@add\normalfont{\bb1@eqnodir}%
5975     \ifnum\bb1@eqnpos=\@ne
5976       \def\@eqnnum{%
5977         \setbox\z@\hbox{\bb1@eqnum}%
5978         \hbox to 0.01pt{\hss\hbox to\displaywidth{\box\z@\hss}}}%
5979       \else
5980         \let\@eqnnum\bb1@eqnum
5981       \fi
5982     \fi}
5983 % Hack. YA luatex bug?:
5984 \expandafter\bb1@sreplace\csname] \endcsname{${$}\eqno\kern.001pt${$}%
5985 \else % amstex
5986   \ifx\bb1@noamsmath\@undefined
5987     \ifnum\bb1@eqnpos=\@ne
5988       \let\bb1@ams@lap\hbox
5989     \else
5990       \let\bb1@ams@lap\llap
5991     \fi
5992     \ExplSyntaxOn
5993     \bb1@sreplace\intertext@{\normalbaselines}%
5994     {\normalbaselines
5995       \ifx\bb1@eqnodir\relax\else\bb1@pardir\@ne\bb1@eqnodir\fi}%
5996     \ExplSyntaxOff
5997     \def\bb1@ams@tagbox#1#2{#1{\bb1@eqnodir#2}}% #1=hbox|@lap|flip
5998     \ifx\bb1@ams@lap\hbox % leqno
5999       \def\bb1@ams@flip#1{%
6000         \hbox to 0.01pt{\hss\hbox to\displaywidth{#{1}\hss}}}%
6001     \else % eqno
6002       \def\bb1@ams@flip#1{%
6003         \hbox to 0.01pt{\hbox to\displaywidth{\hss{#1}\hss}}}%
6004     \fi
6005     \def\bb1@ams@preset#1{%
6006       \ifnum\bb1@thetextdir>\z@
6007         \edef\bb1@eqnodir{\noexpand\bb1@textdir{\the\bb1@thetextdir}}%
6008         \bb1@sreplace\textdef@{\hbox}{\bb1@ams@tagbox\hbox}%
6009         \bb1@sreplace\maketag@@@{\hbox}{\bb1@ams@tagbox#1}%
6010       \fi}%
6011     \ifnum\bb1@eqnpos=\tw@%else
6012       \def\bb1@ams@equation{%
6013         \ifnum\bb1@thetextdir>\z@
6014           \edef\bb1@eqnodir{\noexpand\bb1@textdir{\the\bb1@thetextdir}}%
6015           \chardef\bb1@thetextdir\z@
6016           \bb1@add\normalfont{\bb1@eqnodir}%
6017           \ifcase\bb1@eqnpos

```

```

6018         \def\veqno##1##2{\bbl@eqno@flip{##1##2}}%
6019         \or
6020         \def\veqno##1##2{\bbl@leqno@flip{##1##2}}%
6021         \fi
6022     \fi}%
6023     \AddToHook{env/equation/begin}{\bbl@ams@equation}%
6024     \AddToHook{env/equation*/begin}{\bbl@ams@equation}%
6025     \fi
6026     \AddToHook{env/cases/begin}{\bbl@ams@preset\bbl@ams@lap}%
6027     \AddToHook{env/multline/begin}{\bbl@ams@preset\hbox}%
6028     \AddToHook{env/gather/begin}{\bbl@ams@preset\bbl@ams@lap}%
6029     \AddToHook{env/gather*/begin}{\bbl@ams@preset\bbl@ams@lap}%
6030     \AddToHook{env/align/begin}{\bbl@ams@preset\bbl@ams@lap}%
6031     \AddToHook{env/align*/begin}{\bbl@ams@preset\bbl@ams@lap}%
6032     \AddToHook{env/eqnalign/begin}{\bbl@ams@preset\hbox}%
6033     % Hackish, for proper alignment. Don't ask me why it works!:
6034     \bbl@exp{% Avoid a 'visible' conditional
6035         \\\AddToHook{env/align*/end}{\<iftag@>\<else>\\tag*{}<fi>}}%
6036     \AddToHook{env/flalign/begin}{\bbl@ams@preset\hbox}%
6037     \AddToHook{env/split/before}{%
6038         \ifnum\bbl@thetextdir>\z@
6039             \bbl@ifsamestring\@currentvir{equation}%
6040             {\ifx\bbl@ams@lap\hbox % leqno
6041                 \def\bbl@ams@flip#1{%
6042                     \hbox to 0.01pt{\hbox to\displaywidth{{#1}\hss}\hss}}%
6043                 \else
6044                     \def\bbl@ams@flip#1{%
6045                         \hbox to 0.01pt{\hss\hbox to\displaywidth{\hss{#1}}}%
6046                     \fi}%
6047                 {}%
6048             \fi}%
6049         \fi
6050     \fi}
6051 \fi
6052 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
6053 \ifnum\bbl@bidimode>\z@
6054     \def\bbl@nextfake#1{% non-local changes, use always inside a group!
6055         \bbl@exp{%
6056             \def\\bbl@insidemath{0}%
6057             \mathdir\the\bodydir
6058             #1% Once entered in math, set boxes to restore values
6059             \<ifmmode>%
6060                 \everyvbox{%
6061                     \the\everyvbox
6062                     \bodydir\the\bodydir
6063                     \mathdir\the\mathdir
6064                     \everyhbox{\the\everyhbox}%
6065                     \everyvbox{\the\everyvbox}}%
6066                 \everyhbox{%
6067                     \the\everyhbox
6068                     \bodydir\the\bodydir
6069                     \mathdir\the\mathdir
6070                     \everyhbox{\the\everyhbox}%
6071                     \everyvbox{\the\everyvbox}}%
6072             \<fi>}}%
6073     \def\@hangfrom#1{%
6074         \setbox\@tempboxa\hbox{{#1}}%
6075         \hangindent\wd\@tempboxa
6076         \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
6077             \shapemode\@ne
6078         \fi
6079         \noindent\box\@tempboxa}
6080 \fi

```

```

6081 \IfBabelLayout{tabular}
6082   {\let\bbl@OL@@tabular\@tabular
6083     \bbl@replace\@tabular{$}\bbl@nextfake$}%
6084     \let\bbl@NL@@tabular\@tabular
6085     \AtBeginDocument{%
6086       \ifx\bbl@NL@@tabular\@tabular\else
6087         \bbl@replace\@tabular{$}\bbl@nextfake$}%
6088         \let\bbl@NL@@tabular\@tabular
6089       \fi}}
6090   {}
6091 \IfBabelLayout{lists}
6092   {\let\bbl@OL@list\list
6093     \bbl@sreplace\list{\parshape}\bbl@listparshape}%
6094     \let\bbl@NL@list\list
6095     \def\bbl@listparshape#1#2#3{%
6096       \parshape #1 #2 #3 %
6097       \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
6098         \shapemode\tw@
6099       \fi}}
6100   {}
6101 \IfBabelLayout{graphics}
6102   {\let\bbl@pictresetdir\relax
6103     \def\bbl@pictsetdir#1{%
6104       \ifcase\bbl@thetextdir
6105         \let\bbl@pictresetdir\relax
6106       \else
6107         \ifcase#1\bodydir TLT % Remember this sets the inner boxes
6108           \or\textdir TLT
6109           \else\bodydir TLT \textdir TLT
6110         \fi
6111         % \ (text|par)dir required in pgf:
6112         \def\bbl@pictresetdir{\bodydir TRT\pardir TRT\textdir TRT\relax}%
6113       \fi}%
6114     \AddToHook{env/picture/begin}{\bbl@pictsetdir\tw@}%
6115     \directlua{
6116       Babel.get_picture_dir = true
6117       Babel.picture_has_bidi = 0
6118       %
6119       function Babel.picture_dir (head)
6120         if not Babel.get_picture_dir then return head end
6121         if Babel.hlist_has_bidi(head) then
6122           Babel.picture_has_bidi = 1
6123         end
6124         return head
6125       end
6126       luatexbase.add_to_callback("hpack_filter", Babel.picture_dir,
6127         "Babel.picture_dir")
6128     }%
6129     \AtBeginDocument{%
6130       \long\def\put(#1,#2)#3{%
6131         \@killglue
6132         % Try:
6133         \ifx\bbl@pictresetdir\relax
6134           \def\bbl@tempc{0}%
6135         \else
6136           \directlua{
6137             Babel.get_picture_dir = true
6138             Babel.picture_has_bidi = 0
6139           }%
6140           \setbox\z@\hb@xt@\z@{%
6141             \@defaultunitsset\@tempdimc{#1}\unitlength
6142             \kern\@tempdimc
6143             #3\hss}% TODO: #3 executed twice (below). That's bad.

```

```

6144     \edef\bbl@tempc{\directlua{tex.print(Babel.picture_has_bidi)}}%
6145     \fi
6146     % Do:
6147     \@defaultunitsset\@tempdimc{#2}\unitlength
6148     \raise\@tempdimc\hb@xt@\z@{%
6149       \@defaultunitsset\@tempdimc{#1}\unitlength
6150       \kern\@tempdimc
6151       {\ifnum\bbl@tempc>\z@\bbl@pictresetdir\fi#3}\hss}%
6152     \ignorespaces}%
6153     \MakeRobust\put}%
6154     \AtBeginDocument
6155     {\AddToHook{cmd/diagbox@pict/before}{\let\bbl@pictsetdir\@gobble}%
6156     \ifx\pgfpicture\@undefined\else % TODO. Allow deactivate?
6157       \AddToHook{env/pgfpicture/begin}{\bbl@pictsetdir\@ne}%
6158       \bbl@add\pgfinterruptpicture{\bbl@pictresetdir}%
6159       \bbl@add\pgfsys@beginpicture{\bbl@pictsetdir\z@}%
6160     \fi
6161     \ifx\tikzpicture\@undefined\else
6162       \AddToHook{env/tikzpicture/begin}{\bbl@pictsetdir\z@}%
6163       \bbl@add\tikz@atbegin@node{\bbl@pictresetdir}%
6164       \bbl@sreplace\tikz{\begingroup}{\begingroup\bbl@pictsetdir\tw@}%
6165     \fi
6166     \ifx\tcolorbox\@undefined\else
6167       \AddToHook{env/tcolorbox/begin}{\bbl@pictsetdir\@ne}%
6168       \bbl@sreplace\tcb@savebox
6169       {\ignorespaces}{\ignorespaces\bbl@pictresetdir}%
6170       \ifx\tikzpicture\tcb@hooked\@undefined\else
6171         \bbl@sreplace\tikzpicture\tcb@hooked{\noexpand\tikzpicture}%
6172         {\textdir TLT\noexpand\tikzpicture}%
6173       \fi
6174     \fi
6175   }}
6176 {}

```

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`.

```

6177 \IfBabelLayout{counters}%
6178 {\let\bbl@OL@@textsuperscript\@textsuperscript
6179  \bbl@sreplace\@textsuperscript{\m@th}{\m@th\mathdir\pagedir}%
6180  \let\bbl@latinarabic=\@arabic
6181  \let\bbl@OL@@arabic\@arabic
6182  \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
6183  \@ifpackagewith{babel}{bidi=default}%
6184  {\let\bbl@asciroman=\@roman
6185   \let\bbl@OL@@roman\@roman
6186   \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
6187   \let\bbl@asciiRoman=\@Roman
6188   \let\bbl@OL@@roman\@Roman
6189   \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}%
6190   \let\bbl@OL@labelenumii\labelenumii
6191   \def\labelenumii{}\theenumii}%
6192   \let\bbl@OL@p@enumiii\p@enumiii
6193   \def\p@enumiii{\p@enumii}\theenumii{}\{}\{}\}
6194 }{<Footnote changes>}
6195 \IfBabelLayout{footnotes}%
6196 {\let\bbl@OL@footnote\footnote
6197  \BabelFootnote\footnote\languagename{}\{}\}%
6198  \BabelFootnote\localfootnote\languagename{}\{}\}%
6199  \BabelFootnote\mainfootnote{}\{}\{}\}
6200 {}

```

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.



```

6201 \IfBabelLayout{extras}%
6202   {\let\bbl@OL@underline\underline
6203    \bbl@sreplace\underline{$\@@underline}{\bbl@nextfake$\@@underline}%
6204    \let\bbl@OL@LaTeX2e\LaTeX2e
6205    \DeclareRobustCommand{\LaTeXe}{\mbox{\m@th
6206     \if b\expandafter\@car\@series\@nil\boldmath\fi
6207     \babelsublr}%
6208     \LaTeX\kern.15em2\bbl@nextfake$_{\textstyle\varepsilon}}}}
6209   {}
6210 \end{luatex}

```

## 12.11 Lua: transforms

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.

```

6211 (*transforms)
6212 Babel.linebreaking.replacements = {}
6213 Babel.linebreaking.replacements[0] = {} -- pre
6214 Babel.linebreaking.replacements[1] = {} -- post
6215
6216 -- Discretionaries contain strings as nodes
6217 function Babel.str_to_nodes(fn, matches, base)
6218   local n, head, last
6219   if fn == nil then return nil end
6220   for s in string.utfvalues(fn(matches)) do
6221     if base.id == 7 then
6222       base = base.replace
6223     end
6224     n = node.copy(base)
6225     n.char = s
6226     if not head then
6227       head = n
6228     else
6229       last.next = n
6230     end
6231     last = n
6232   end
6233   return head
6234 end
6235
6236 Babel.fetch_subtext = {}
6237
6238 Babel.ignore_pre_char = function(node)
6239   return (node.lang == Babel.nohyphenation)
6240 end
6241
6242 -- Merging both functions doesn't seem feasible, because there are too
6243 -- many differences.
6244 Babel.fetch_subtext[0] = function(head)
6245   local word_string = ''
6246   local word_nodes = {}
6247   local lang
6248   local item = head
6249   local inmath = false

```

```

6250
6251 while item do
6252
6253     if item.id == 11 then
6254         inmath = (item.subtype == 0)
6255     end
6256
6257     if inmath then
6258         -- pass
6259
6260     elseif item.id == 29 then
6261         local locale = node.get_attribute(item, Babel.attr_locale)
6262
6263         if lang == locale or lang == nil then
6264             lang = lang or locale
6265             if Babel.ignore_pre_char(item) then
6266                 word_string = word_string .. Babel.us_char
6267             else
6268                 word_string = word_string .. unicode.utf8.char(item.char)
6269             end
6270             word_nodes[#word_nodes+1] = item
6271         else
6272             break
6273         end
6274
6275     elseif item.id == 12 and item.subtype == 13 then
6276         word_string = word_string .. ' '
6277         word_nodes[#word_nodes+1] = item
6278
6279         -- Ignore leading unrecognized nodes, too.
6280     elseif word_string ~= '' then
6281         word_string = word_string .. Babel.us_char
6282         word_nodes[#word_nodes+1] = item -- Will be ignored
6283     end
6284
6285     item = item.next
6286 end
6287
6288 -- Here and above we remove some trailing chars but not the
6289 -- corresponding nodes. But they aren't accessed.
6290 if word_string:sub(-1) == ' ' then
6291     word_string = word_string:sub(1,-2)
6292 end
6293 word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
6294 return word_string, word_nodes, item, lang
6295 end
6296
6297 Babel.fetch_subtext[1] = function(head)
6298     local word_string = ''
6299     local word_nodes = {}
6300     local lang
6301     local item = head
6302     local inmath = false
6303
6304     while item do
6305
6306         if item.id == 11 then
6307             inmath = (item.subtype == 0)
6308         end
6309
6310         if inmath then
6311             -- pass
6312

```

```

6313     elseif item.id == 29 then
6314         if item.lang == lang or lang == nil then
6315             if (item.char ~= 124) and (item.char ~= 61) then -- not =, not |
6316                 lang = lang or item.lang
6317                 word_string = word_string .. unicode.utf8.char(item.char)
6318                 word_nodes[#word_nodes+1] = item
6319             end
6320         else
6321             break
6322         end
6323
6324     elseif item.id == 7 and item.subtype == 2 then
6325         word_string = word_string .. '='
6326         word_nodes[#word_nodes+1] = item
6327
6328     elseif item.id == 7 and item.subtype == 3 then
6329         word_string = word_string .. '|'
6330         word_nodes[#word_nodes+1] = item
6331
6332     -- (1) Go to next word if nothing was found, and (2) implicitly
6333     -- remove leading USs.
6334     elseif word_string == '' then
6335         -- pass
6336
6337     -- This is the responsible for splitting by words.
6338     elseif (item.id == 12 and item.subtype == 13) then
6339         break
6340
6341     else
6342         word_string = word_string .. Babel.us_char
6343         word_nodes[#word_nodes+1] = item -- Will be ignored
6344     end
6345
6346     item = item.next
6347 end
6348
6349 word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
6350 return word_string, word_nodes, item, lang
6351 end
6352
6353 function Babel.pre_hyphenate_replace(head)
6354     Babel.hyphenate_replace(head, 0)
6355 end
6356
6357 function Babel.post_hyphenate_replace(head)
6358     Babel.hyphenate_replace(head, 1)
6359 end
6360
6361 Babel.us_char = string.char(31)
6362
6363 function Babel.hyphenate_replace(head, mode)
6364     local u = unicode.utf8
6365     local lbkr = Babel.linebreaking.replacements[mode]
6366
6367     local word_head = head
6368
6369     while true do -- for each subtext block
6370
6371         local w, w_nodes, nw, lang = Babel.fetch_subtext[mode](word_head)
6372
6373         if Babel.debug then
6374             print()
6375             print((mode == 0) and '@@@<' or '@@@>', w)

```

```

6376     end
6377
6378     if nw == nil and w == '' then break end
6379
6380     if not lang then goto next end
6381     if not lbkr[lang] then goto next end
6382
6383     -- For each saved (pre|post)hyphenation. TODO. Reconsider how
6384     -- loops are nested.
6385     for k=1, #lbkr[lang] do
6386         local p = lbkr[lang][k].pattern
6387         local r = lbkr[lang][k].replace
6388         local attr = lbkr[lang][k].attr or -1
6389
6390         if Babel.debug then
6391             print('*****', p, mode)
6392         end
6393
6394         -- This variable is set in some cases below to the first *byte*
6395         -- after the match, either as found by u.match (faster) or the
6396         -- computed position based on sc if w has changed.
6397         local last_match = 0
6398         local step = 0
6399
6400         -- For every match.
6401         while true do
6402             if Babel.debug then
6403                 print('====')
6404             end
6405             local new -- used when inserting and removing nodes
6406
6407             local matches = { u.match(w, p, last_match) }
6408
6409             if #matches < 2 then break end
6410
6411             -- Get and remove empty captures (with ()'s, which return a
6412             -- number with the position), and keep actual captures
6413             -- (from (...)), if any, in matches.
6414             local first = table.remove(matches, 1)
6415             local last = table.remove(matches, #matches)
6416             -- Non re-fetched substrings may contain \31, which separates
6417             -- subsubstrings.
6418             if string.find(w:sub(first, last-1), Babel.us_char) then break end
6419
6420             local save_last = last -- with A()BC()D, points to D
6421
6422             -- Fix offsets, from bytes to unicode. Explained above.
6423             first = u.len(w:sub(1, first-1)) + 1
6424             last = u.len(w:sub(1, last-1)) -- now last points to C
6425
6426             -- This loop stores in a small table the nodes
6427             -- corresponding to the pattern. Used by 'data' to provide a
6428             -- predictable behavior with 'insert' (w_nodes is modified on
6429             -- the fly), and also access to 'remove'd nodes.
6430             local sc = first-1 -- Used below, too
6431             local data_nodes = {}
6432
6433             local enabled = true
6434             for q = 1, last-first+1 do
6435                 data_nodes[q] = w_nodes[sc+q]
6436                 if enabled
6437                     and attr > -1
6438                     and not node.has_attribute(data_nodes[q], attr)

```

```

6439         then
6440             enabled = false
6441         end
6442     end
6443
6444     -- This loop traverses the matched substring and takes the
6445     -- corresponding action stored in the replacement list.
6446     -- sc = the position in substr nodes / string
6447     -- rc = the replacement table index
6448     local rc = 0
6449
6450     while rc < last-first+1 do -- for each replacement
6451         if Babel.debug then
6452             print('.....', rc + 1)
6453         end
6454         sc = sc + 1
6455         rc = rc + 1
6456
6457         if Babel.debug then
6458             Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
6459             local ss = ''
6460             for itt in node.traverse(head) do
6461                 if itt.id == 29 then
6462                     ss = ss .. unicode.utf8.char(itt.char)
6463                 else
6464                     ss = ss .. '{' .. itt.id .. '}'
6465                 end
6466             end
6467             print('*****', ss)
6468
6469         end
6470
6471         local crep = r[rc]
6472         local item = w_nodes[sc]
6473         local item_base = item
6474         local placeholder = Babel.us_char
6475         local d
6476
6477         if crep and crep.data then
6478             item_base = data_nodes[crep.data]
6479         end
6480
6481         if crep then
6482             step = crep.step or 0
6483         end
6484
6485         if (not enabled) or (crep and next(crep) == nil) then -- = {}
6486             last_match = save_last -- Optimization
6487             goto next
6488
6489         elseif crep == nil or crep.remove then
6490             node.remove(head, item)
6491             table.remove(w_nodes, sc)
6492             w = u.sub(w, 1, sc-1) .. u.sub(w, sc+1)
6493             sc = sc - 1 -- Nothing has been inserted.
6494             last_match = utf8.offset(w, sc+1+step)
6495             goto next
6496
6497         elseif crep and crep.kashida then -- Experimental
6498             node.set_attribute(item,
6499                 Babel.attr_kashida,
6500                 crep.kashida)
6501             last_match = utf8.offset(w, sc+1+step)

```

```

6502         goto next
6503
6504     elseif crep and crep.string then
6505         local str = crep.string(matches)
6506         if str == '' then -- Gather with nil
6507             node.remove(head, item)
6508             table.remove(w_nodes, sc)
6509             w = u.sub(w, 1, sc-1) .. u.sub(w, sc+1)
6510             sc = sc - 1 -- Nothing has been inserted.
6511         else
6512             local loop_first = true
6513             for s in string.utfvalues(str) do
6514                 d = node.copy(item_base)
6515                 d.char = s
6516                 if loop_first then
6517                     loop_first = false
6518                     head, new = node.insert_before(head, item, d)
6519                     if sc == 1 then
6520                         word_head = head
6521                     end
6522                     w_nodes[sc] = d
6523                     w = u.sub(w, 1, sc-1) .. u.char(s) .. u.sub(w, sc+1)
6524                 else
6525                     sc = sc + 1
6526                     head, new = node.insert_before(head, item, d)
6527                     table.insert(w_nodes, sc, new)
6528                     w = u.sub(w, 1, sc-1) .. u.char(s) .. u.sub(w, sc)
6529                 end
6530                 if Babel.debug then
6531                     print('.....', 'str')
6532                     Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
6533                 end
6534             end -- for
6535             node.remove(head, item)
6536         end -- if ''
6537         last_match = utf8.offset(w, sc+1+step)
6538         goto next
6539
6540     elseif mode == 1 and crep and (crep.pre or crep.no or crep.post) then
6541         d = node.new(7, 0) -- (disc, discretionary)
6542         d.pre = Babel.str_to_nodes(crep.pre, matches, item_base)
6543         d.post = Babel.str_to_nodes(crep.post, matches, item_base)
6544         d.replace = Babel.str_to_nodes(crep.no, matches, item_base)
6545         d.attr = item_base.attr
6546         if crep.pre == nil then -- TeXbook p96
6547             d.penalty = crep.penalty or tex.hyphenpenalty
6548         else
6549             d.penalty = crep.penalty or tex.exhyphenpenalty
6550         end
6551         placeholder = '|'
6552         head, new = node.insert_before(head, item, d)
6553
6554     elseif mode == 0 and crep and (crep.pre or crep.no or crep.post) then
6555         -- ERROR
6556
6557     elseif crep and crep.penalty then
6558         d = node.new(14, 0) -- (penalty, userpenalty)
6559         d.attr = item_base.attr
6560         d.penalty = crep.penalty
6561         head, new = node.insert_before(head, item, d)
6562
6563     elseif crep and crep.space then
6564         -- 655360 = 10 pt = 10 * 65536 sp

```

```

6565         d = node.new(12, 13)      -- (glue, spaceskip)
6566         local quad = font.getfont(item_base.font).size or 655360
6567         node.setglue(d, crep.space[1] * quad,
6568                     crep.space[2] * quad,
6569                     crep.space[3] * quad)
6570         if mode == 0 then
6571             placeholder = ' '
6572         end
6573         head, new = node.insert_before(head, item, d)
6574
6575     elseif crep and crep.spacefactor then
6576         d = node.new(12, 13)      -- (glue, spaceskip)
6577         local base_font = font.getfont(item_base.font)
6578         node.setglue(d,
6579                     crep.spacefactor[1] * base_font.parameters['space'],
6580                     crep.spacefactor[2] * base_font.parameters['space_stretch'],
6581                     crep.spacefactor[3] * base_font.parameters['space_shrink'])
6582         if mode == 0 then
6583             placeholder = ' '
6584         end
6585         head, new = node.insert_before(head, item, d)
6586
6587     elseif mode == 0 and crep and crep.space then
6588         -- ERROR
6589
6590     end -- ie replacement cases
6591
6592     -- Shared by disc, space and penalty.
6593     if sc == 1 then
6594         word_head = head
6595     end
6596     if crep.insert then
6597         w = u.sub(w, 1, sc-1) .. placeholder .. u.sub(w, sc)
6598         table.insert(w_nodes, sc, new)
6599         last = last + 1
6600     else
6601         w_nodes[sc] = d
6602         node.remove(head, item)
6603         w = u.sub(w, 1, sc-1) .. placeholder .. u.sub(w, sc+1)
6604     end
6605
6606     last_match = utf8.offset(w, sc+1+step)
6607
6608     ::next::
6609
6610     end -- for each replacement
6611
6612     if Babel.debug then
6613         print('.....', '/')
6614         Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
6615     end
6616
6617     end -- for match
6618
6619     end -- for patterns
6620
6621     ::next::
6622     word_head = nw
6623 end -- for substring
6624 return head
6625 end
6626
6627 -- This table stores capture maps, numbered consecutively

```

```

6628 Babel.capture_maps = {}
6629
6630 -- The following functions belong to the next macro
6631 function Babel.capture_func(key, cap)
6632   local ret = "[" .. cap:gsub('{{[0-9]}}', "]]..m[%1]..[" .. "]"
6633   local cnt
6634   local u = unicode.utf8
6635   ret, cnt = ret:gsub('{{[0-9]}|([^\]]+)|(.-)}', Babel.capture_func_map)
6636   if cnt == 0 then
6637     ret = u.gsub(ret, '{(%x%x%x%x+)}',
6638       function (n)
6639         return u.char(tonumber(n, 16))
6640       end)
6641   end
6642   ret = ret:gsub("%[%[%]]%.", '')
6643   ret = ret:gsub("%.%.%.%[%[%]]%", '')
6644   return key .. "[=function(m) return ]] .. ret .. [[ end]]
6645 end
6646
6647 function Babel.capt_map(from, mapno)
6648   return Babel.capture_maps[mapno][from] or from
6649 end
6650
6651 -- Handle the {n|abc|ABC} syntax in captures
6652 function Babel.capture_func_map(capno, from, to)
6653   local u = unicode.utf8
6654   from = u.gsub(from, '{(%x%x%x%x+)}',
6655     function (n)
6656       return u.char(tonumber(n, 16))
6657     end)
6658   to = u.gsub(to, '{(%x%x%x%x+)}',
6659     function (n)
6660       return u.char(tonumber(n, 16))
6661     end)
6662   local froms = {}
6663   for s in string.utfcharacters(from) do
6664     table.insert(froms, s)
6665   end
6666   local cnt = 1
6667   table.insert(Babel.capture_maps, {})
6668   local mlen = table.getn(Babel.capture_maps)
6669   for s in string.utfcharacters(to) do
6670     Babel.capture_maps[mlen][froms[cnt]] = s
6671     cnt = cnt + 1
6672   end
6673   return "]]..Babel.capt_map(m[" .. capno .. "], " ..
6674     (mlen) .. ").." .. "["
6675 end
6676
6677 -- Create/Extend reversed sorted list of kashida weights:
6678 function Babel.capture_kashida(key, wt)
6679   wt = tonumber(wt)
6680   if Babel.kashida_wts then
6681     for p, q in ipairs(Babel.kashida_wts) do
6682       if wt == q then
6683         break
6684       elseif wt > q then
6685         table.insert(Babel.kashida_wts, p, wt)
6686         break
6687       elseif table.getn(Babel.kashida_wts) == p then
6688         table.insert(Babel.kashida_wts, wt)
6689       end
6690     end

```



```

6691 else
6692     Babel.kashida_wts = { wt }
6693 end
6694 return 'kashida = ' .. wt
6695 end
6696 </transforms>

```

## 12.12 Lua: Auto bidi with basic and basic-r

The file `babel-data-bidi.lua` currently only contains data. It is a large and boring file and it is not shown here (see the generated file), but here is a sample:

```

[0x25]={d='et'},
[0x26]={d='on'},
[0x27]={d='on'},
[0x28]={d='on', m=0x29},
[0x29]={d='on', m=0x28},
[0x2A]={d='on'},
[0x2B]={d='es'},
[0x2C]={d='cs'},

```

For the meaning of these codes, see the Unicode standard.

Now the `basic-r` bidi mode. One of the aims is to implement a fast and simple bidi algorithm, with a single loop. I managed to do it for R texts, with a second smaller loop for a special case. The code is still somewhat chaotic, but its behavior is essentially correct. I cannot resist copying the following text from Emacs `bidi.c` (which also attempts to implement the bidi algorithm with a single loop):

Arrrrgh!! The UAX#9 algorithm is too deeply entrenched in the assumption of batch-style processing [...]. May the fleas of a thousand camels infest the armpits of those who design supposedly general-purpose algorithms by looking at their own implementations, and fail to consider other possible implementations!

Well, it took me some time to guess what the batch rules in UAX#9 actually mean (in other word, *what* they do and *why*, and not only *how*), but I think (or I hope) I've managed to understand them. In some sense, there are two bidi modes, one for numbers, and the other for text. Furthermore, setting just the direction in R text is not enough, because there are actually *two* R modes (set explicitly in Unicode with RLM and ALM). In babel the dir is set by a higher protocol based on the language/script, which in turn sets the correct dir (<l>, <r> or <al>).

From UAX#9: “Where available, markup should be used instead of the explicit formatting characters”. So, this simple version just ignores formatting characters. Actually, most of that annex is devoted to how to handle them.

BD14-BD16 are not implemented. Unicode (and the W3C) are making a great effort to deal with some special problematic cases in “streamed” plain text. I don’t think this is the way to go – particular issues should be fixed by a high level interface taking into account the needs of the document. And here is where `luatex` excels, because everything related to bidi writing is under our control.

```

6697 <*basic-r>
6698 Babel = Babel or {}
6699
6700 Babel.bidi_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
6709 local function dir_mark(head, from, to, outer)
6710     dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
6711     local d = node.new(DIR)
6712     d.dir = '+' .. dir
6713     node.insert_before(head, from, d)
6714     d = node.new(DIR)

```

```

6715 d.dir = '-' .. dir
6716 node.insert_after(head, to, d)
6717 end
6718
6719 function Babel.bidi(head, ispar)
6720   local first_n, last_n          -- first and last char with nums
6721   local last_es                  -- an auxiliary 'last' used with nums
6722   local first_d, last_d          -- first and last char in L/R block
6723   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):

```

6724   local strong = ('TRT' == tex.pardir) and 'r' or 'l'
6725   local strong_lr = (strong == 'l') and 'l' or 'r'
6726   local outer = strong
6727
6728   local new_dir = false
6729   local first_dir = false
6730   local inmath = false
6731
6732   local last_lr
6733
6734   local type_n = ''
6735
6736   for item in node.traverse(head) do
6737
6738     -- three cases: glyph, dir, otherwise
6739     if item.id == node.id'glyph'
6740       or (item.id == 7 and item.subtype == 2) then
6741
6742       local itemchar
6743       if item.id == 7 and item.subtype == 2 then
6744         itemchar = item.replace.char
6745       else
6746         itemchar = item.char
6747       end
6748       local chardata = characters[itemchar]
6749       dir = chardata and chardata.d or nil
6750       if not dir then
6751         for nn, et in ipairs(ranges) do
6752           if itemchar < et[1] then
6753             break
6754           elseif itemchar <= et[2] then
6755             dir = et[3]
6756             break
6757           end
6758         end
6759       end
6760       dir = dir or 'l'
6761       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).

```

6762   if new_dir then
6763     attr_dir = 0
6764     for at in node.traverse(item.attr) do
6765       if at.number == Babel.attr_dir then
6766         attr_dir = at.value % 3
6767       end
6768     end

```

```

6769         if attr_dir == 1 then
6770             strong = 'r'
6771         elseif attr_dir == 2 then
6772             strong = 'al'
6773         else
6774             strong = 'l'
6775         end
6776         strong_lr = (strong == 'l') and 'l' or 'r'
6777         outer = strong_lr
6778         new_dir = false
6779     end
6780
6781     if dir == 'nsm' then dir = strong end          -- W1

```

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

```

6782     dir_real = dir          -- We need dir_real to set strong below
6783     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:

```

6784     if strong == 'al' then
6785         if dir == 'en' then dir = 'an' end          -- W2
6786         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
6787         strong_lr = 'r'          -- W3
6788     end

```

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

```

6789     elseif item.id == node.id'dir' and not inmath then
6790         new_dir = true
6791         dir = nil
6792     elseif item.id == node.id'math' then
6793         inmath = (item.subtype == 0)
6794     else
6795         dir = nil          -- Not a char
6796     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>.

```

6797     if dir == 'en' or dir == 'an' or dir == 'et' then
6798         if dir ~= 'et' then
6799             type_n = dir
6800         end
6801         first_n = first_n or item
6802         last_n = last_es or item
6803         last_es = nil
6804     elseif dir == 'es' and last_n then -- W3+W6
6805         last_es = item
6806     elseif dir == 'cs' then          -- it's right - do nothing
6807     elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
6808         if strong_lr == 'r' and type_n ~= '' then
6809             dir_mark(head, first_n, last_n, 'r')
6810         elseif strong_lr == 'l' and first_d and type_n == 'an' then
6811             dir_mark(head, first_n, last_n, 'r')
6812             dir_mark(head, first_d, last_d, outer)
6813             first_d, last_d = nil, nil
6814         elseif strong_lr == 'l' and type_n ~= '' then
6815             last_d = last_n
6816         end
6817         type_n = ''
6818         first_n, last_n = nil, nil
6819     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:

```

6820   if dir == 'l' or dir == 'r' then
6821     if dir ~= outer then
6822       first_d = first_d or item
6823       last_d = item
6824     elseif first_d and dir ~= strong_lr then
6825       dir_mark(head, first_d, last_d, outer)
6826       first_d, last_d = nil, nil
6827     end
6828   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.

```

6829   if dir and not last_lr and dir ~= 'l' and outer == 'r' then
6830     item.char = characters[item.char] and
6831       characters[item.char].m or item.char
6832   elseif (dir or new_dir) and last_lr ~= item then
6833     local mir = outer .. strong_lr .. (dir or outer)
6834     if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
6835       for ch in node.traverse(node.next(last_lr)) do
6836         if ch == item then break end
6837         if ch.id == node.id'glyph' and characters[ch.char] then
6838           ch.char = characters[ch.char].m or ch.char
6839         end
6840       end
6841     end
6842   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).

```

6843   if dir == 'l' or dir == 'r' then
6844     last_lr = item
6845     strong = dir_real -- Don't search back - best save now
6846     strong_lr = (strong == 'l') and 'l' or 'r'
6847   elseif new_dir then
6848     last_lr = nil
6849   end
6850 end

```

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

```

6851   if last_lr and outer == 'r' then
6852     for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
6853       if characters[ch.char] then
6854         ch.char = characters[ch.char].m or ch.char
6855       end
6856     end
6857   end
6858   if first_n then
6859     dir_mark(head, first_n, last_n, outer)
6860   end
6861   if first_d then
6862     dir_mark(head, first_d, last_d, outer)
6863   end

```

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

```

6864   return node.prev(head) or head

```

```

6865 end
6866 \</basic-r>

And here the Lua code for bidi=basic:

6867 \<*basic>
6868 Babel = Babel or {}
6869
6870 -- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
6871
6872 Babel.fontmap = Babel.fontmap or {}
6873 Babel.fontmap[0] = {} -- l
6874 Babel.fontmap[1] = {} -- r
6875 Babel.fontmap[2] = {} -- al/an
6876
6877 Babel.bidi_enabled = true
6878 Babel.mirroring_enabled = true
6879
6880 require('babel-data-bidi.lua')
6881
6882 local characters = Babel.characters
6883 local ranges = Babel.ranges
6884
6885 local DIR = node.id('dir')
6886 local GLYPH = node.id('glyph')
6887
6888 local function insert_implicit(head, state, outer)
6889   local new_state = state
6890   if state.sim and state.eim and state.sim ~= state.eim then
6891     dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
6892     local d = node.new(DIR)
6893     d.dir = '+' .. dir
6894     node.insert_before(head, state.sim, d)
6895     local d = node.new(DIR)
6896     d.dir = '-' .. dir
6897     node.insert_after(head, state.eim, d)
6898   end
6899   new_state.sim, new_state.eim = nil, nil
6900   return head, new_state
6901 end
6902
6903 local function insert_numeric(head, state)
6904   local new
6905   local new_state = state
6906   if state.san and state.ean and state.san ~= state.ean then
6907     local d = node.new(DIR)
6908     d.dir = '+TLT'
6909     _, new = node.insert_before(head, state.san, d)
6910     if state.san == state.sim then state.sim = new end
6911     local d = node.new(DIR)
6912     d.dir = '-TLT'
6913     _, new = node.insert_after(head, state.ean, d)
6914     if state.ean == state.eim then state.eim = new end
6915   end
6916   new_state.san, new_state.ean = nil, nil
6917   return head, new_state
6918 end
6919
6920 -- TODO - \hbox with an explicit dir can lead to wrong results
6921 -- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>. A small attempt
6922 -- was s made to improve the situation, but the problem is the 3-dir
6923 -- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
6924 -- well.
6925

```

```

6926 function Babel.bidi(head, ispar, hdir)
6927   local d -- d is used mainly for computations in a loop
6928   local prev_d = ''
6929   local new_d = false
6930
6931   local nodes = {}
6932   local outer_first = nil
6933   local inmath = false
6934
6935   local glue_d = nil
6936   local glue_i = nil
6937
6938   local has_en = false
6939   local first_et = nil
6940
6941   local ATDIR = Babel.attr_dir
6942
6943   local save_outer
6944   local temp = node.get_attribute(head, ATDIR)
6945   if temp then
6946     temp = temp % 3
6947     save_outer = (temp == 0 and 'l') or
6948                  (temp == 1 and 'r') or
6949                  (temp == 2 and 'al')
6950   elseif ispar then -- Or error? Shouldn't happen
6951     save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
6952   else -- Or error? Shouldn't happen
6953     save_outer = ('TRT' == hdir) and 'r' or 'l'
6954   end
6955   -- when the callback is called, we are just _after_ the box,
6956   -- and the texdir is that of the surrounding text
6957   -- if not ispar and hdir ~= tex.texdir then
6958   --   save_outer = ('TRT' == hdir) and 'r' or 'l'
6959   -- end
6960   local outer = save_outer
6961   local last = outer
6962   -- 'al' is only taken into account in the first, current loop
6963   if save_outer == 'al' then save_outer = 'r' end
6964
6965   local fontmap = Babel.fontmap
6966
6967   for item in node.traverse(head) do
6968
6969     -- In what follows, #node is the last (previous) node, because the
6970     -- current one is not added until we start processing the neutrals.
6971
6972     -- three cases: glyph, dir, otherwise
6973     if item.id == GLYPH
6974       or (item.id == 7 and item.subtype == 2) then
6975
6976       local d_font = nil
6977       local item_r
6978       if item.id == 7 and item.subtype == 2 then
6979         item_r = item.replace -- automatic discs have just 1 glyph
6980       else
6981         item_r = item
6982       end
6983       local chardata = characters[item_r.char]
6984       d = chardata and chardata.d or nil
6985       if not d or d == 'nsm' then
6986         for nn, et in ipairs(ranges) do
6987           if item_r.char < et[1] then
6988             break

```

```

6989         elseif item_r.char <= et[2] then
6990             if not d then d = et[3]
6991             elseif d == 'nsm' then d_font = et[3]
6992             end
6993             break
6994         end
6995     end
6996 end
6997 d = d or 'l'
6998
6999 -- A short 'pause' in bidi for mapfont
7000 d_font = d_font or d
7001 d_font = (d_font == 'l' and 0) or
7002           (d_font == 'nsm' and 0) or
7003           (d_font == 'r' and 1) or
7004           (d_font == 'al' and 2) or
7005           (d_font == 'an' and 2) or nil
7006 if d_font and fontmap and fontmap[d_font][item_r.font] then
7007     item_r.font = fontmap[d_font][item_r.font]
7008 end
7009
7010 if new_d then
7011     table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
7012     if inmath then
7013         attr_d = 0
7014     else
7015         attr_d = node.get_attribute(item, ATDIR)
7016         attr_d = attr_d % 3
7017     end
7018     if attr_d == 1 then
7019         outer_first = 'r'
7020         last = 'r'
7021     elseif attr_d == 2 then
7022         outer_first = 'r'
7023         last = 'al'
7024     else
7025         outer_first = 'l'
7026         last = 'l'
7027     end
7028     outer = last
7029     has_en = false
7030     first_et = nil
7031     new_d = false
7032 end
7033
7034 if glue_d then
7035     if (d == 'l' and 'l' or 'r') ~= glue_d then
7036         table.insert(nodes, {glue_i, 'on', nil})
7037     end
7038     glue_d = nil
7039     glue_i = nil
7040 end
7041
7042 elseif item.id == DIR then
7043     d = nil
7044     if head ~= item then new_d = true end
7045
7046 elseif item.id == node.id'glue' and item.subtype == 13 then
7047     glue_d = d
7048     glue_i = item
7049     d = nil
7050
7051 elseif item.id == node.id'math' then

```

```

7052     inmath = (item.subtype == 0)
7053
7054 else
7055     d = nil
7056 end
7057
7058 -- AL <= EN/ET/ES      -- W2 + W3 + W6
7059 if last == 'al' and d == 'en' then
7060     d = 'an'           -- W3
7061 elseif last == 'al' and (d == 'et' or d == 'es') then
7062     d = 'on'           -- W6
7063 end
7064
7065 -- EN + CS/ES + EN      -- W4
7066 if d == 'en' and #nodes >= 2 then
7067     if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
7068         and nodes[#nodes-1][2] == 'en' then
7069         nodes[#nodes][2] = 'en'
7070     end
7071 end
7072
7073 -- AN + CS + AN          -- W4 too, because uax9 mixes both cases
7074 if d == 'an' and #nodes >= 2 then
7075     if (nodes[#nodes][2] == 'cs')
7076         and nodes[#nodes-1][2] == 'an' then
7077         nodes[#nodes][2] = 'an'
7078     end
7079 end
7080
7081 -- ET/EN                  -- W5 + W7->1 / W6->on
7082 if d == 'et' then
7083     first_et = first_et or (#nodes + 1)
7084 elseif d == 'en' then
7085     has_en = true
7086     first_et = first_et or (#nodes + 1)
7087 elseif first_et then      -- d may be nil here !
7088     if has_en then
7089         if last == 'l' then
7090             temp = 'l'    -- W7
7091         else
7092             temp = 'en'   -- W5
7093         end
7094     else
7095         temp = 'on'      -- W6
7096     end
7097     for e = first_et, #nodes do
7098         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
7099     end
7100     first_et = nil
7101     has_en = false
7102 end
7103
7104 -- Force mathdir in math if ON (currently works as expected only
7105 -- with 'l')
7106 if inmath and d == 'on' then
7107     d = ('TRT' == tex.mathdir) and 'r' or 'l'
7108 end
7109
7110 if d then
7111     if d == 'al' then
7112         d = 'r'
7113         last = 'al'
7114     elseif d == 'l' or d == 'r' then

```



```

7115         last = d
7116     end
7117     prev_d = d
7118     table.insert(nodes, {item, d, outer_first})
7119 end
7120
7121     outer_first = nil
7122
7123 end
7124
7125 -- TODO -- repeated here in case EN/ET is the last node. Find a
7126 -- better way of doing things:
7127 if first_et then      -- dir may be nil here !
7128     if has_en then
7129         if last == 'l' then
7130             temp = 'l'      -- W7
7131         else
7132             temp = 'en'     -- W5
7133         end
7134     else
7135         temp = 'on'        -- W6
7136     end
7137     for e = first_et, #nodes do
7138         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
7139     end
7140 end
7141
7142 -- dummy node, to close things
7143 table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
7144
7145 ----- NEUTRAL -----
7146
7147 outer = save_outer
7148 last = outer
7149
7150 local first_on = nil
7151
7152 for q = 1, #nodes do
7153     local item
7154
7155     local outer_first = nodes[q][3]
7156     outer = outer_first or outer
7157     last = outer_first or last
7158
7159     local d = nodes[q][2]
7160     if d == 'an' or d == 'en' then d = 'r' end
7161     if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
7162
7163     if d == 'on' then
7164         first_on = first_on or q
7165     elseif first_on then
7166         if last == d then
7167             temp = d
7168         else
7169             temp = outer
7170         end
7171         for r = first_on, q - 1 do
7172             nodes[r][2] = temp
7173             item = nodes[r][1]      -- MIRRORING
7174             if Babel.mirroring_enabled and item.id == GLYPH
7175                 and temp == 'r' and characters[item.char] then
7176                 local font_mode = ''
7177                 if font.fonts[item.font].properties then

```

```

7178         font_mode = font.fonts[item.font].properties.mode
7179     end
7180     if font_mode ~= 'harf' and font_mode ~= 'plug' then
7181         item.char = characters[item.char].m or item.char
7182     end
7183 end
7184 end
7185     first_on = nil
7186 end
7187
7188     if d == 'r' or d == 'l' then last = d end
7189 end
7190
7191 ----- IMPLICIT, REORDER -----
7192
7193 outer = save_outer
7194 last = outer
7195
7196 local state = {}
7197 state.has_r = false
7198
7199 for q = 1, #nodes do
7200
7201     local item = nodes[q][1]
7202
7203     outer = nodes[q][3] or outer
7204
7205     local d = nodes[q][2]
7206
7207     if d == 'nsm' then d = last end          -- W1
7208     if d == 'en' then d = 'an' end
7209     local isdir = (d == 'r' or d == 'l')
7210
7211     if outer == 'l' and d == 'an' then
7212         state.san = state.san or item
7213         state.ean = item
7214     elseif state.san then
7215         head, state = insert_numeric(head, state)
7216     end
7217
7218     if outer == 'l' then
7219         if d == 'an' or d == 'r' then      -- im -> implicit
7220             if d == 'r' then state.has_r = true end
7221             state.sim = state.sim or item
7222             state.eim = item
7223         elseif d == 'l' and state.sim and state.has_r then
7224             head, state = insert_implicit(head, state, outer)
7225         elseif d == 'l' then
7226             state.sim, state.eim, state.has_r = nil, nil, false
7227         end
7228     else
7229         if d == 'an' or d == 'l' then
7230             if nodes[q][3] then -- nil except after an explicit dir
7231                 state.sim = item -- so we move sim 'inside' the group
7232             else
7233                 state.sim = state.sim or item
7234             end
7235             state.eim = item
7236         elseif d == 'r' and state.sim then
7237             head, state = insert_implicit(head, state, outer)
7238         elseif d == 'r' then
7239             state.sim, state.eim = nil, nil
7240         end
7241     end

```

```

7241     end
7242
7243     if isdir then
7244         last = d          -- Don't search back - best save now
7245     elseif d == 'on' and state.san then
7246         state.san = state.san or item
7247         state.ean = item
7248     end
7249
7250 end
7251
7252 return node.prev(head) or head
7253 end
7254 </basic>

```

## 13 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.

## 14 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.

```

7255 <*nil>
7256 \ProvidesLanguage{nil}[\<<date>> \<<version>> Nil language]
7257 \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.

```

7258 \ifx\l@nil\undefined
7259   \newlanguage\l@nil
7260   \namedef{bbl@hyphendata@the\l@nil}{}}}% Remove warning
7261   \let\bbl@elt\relax
7262   \edef\bbl@languages{% Add it to the list of languages
7263     \bbl@languages\bbl@elt{nil}{the\l@nil}{}}
7264 \fi

```

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

```

7265 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}

```

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

```

\captionnil
\datenil
7266 \let\captionnil\@empty
7267 \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.

```

7268 \ldf@finish{nil}
7269 </nil>

```

## 15 Support for Plain T<sub>E</sub>X (plain.def)

### 15.1 Not renaming hyphen.tex

As Don Knuth has declared that the filename `hyphen.tex` may only be used to designate *his* version of the american English hyphenation patterns, a new solution has to be found in order to be able to load hyphenation patterns for other languages in a plain-based T<sub>E</sub>X-format. When asked he responded:

That file name is “sacred”, and if anybody changes it they will cause severe upward/downward compatibility headaches.

People can have a file `localhyphen.tex` or whatever they like, but they mustn’t diddle with `hyphen.tex` (or `plain.tex` except to preload additional fonts).

The files `bplain.tex` and `blplain.tex` can be used as replacement wrappers around `plain.tex` and `lplain.tex` to achieve the desired effect, based on the `babel` package. If you load each of them with `iniTEX`, you will get a file called either `bplain.fmt` or `blplain.fmt`, which you can use as replacements for `plain.fmt` and `lplain.fmt`.

As these files are going to be read as the first thing `iniTEX` sees, we need to set some category codes just to be able to change the definition of `\input`.

```
7270 <(*bplain | blplain>
7271 \catcode`\{=1 % left brace is begin-group character
7272 \catcode`\}=2 % right brace is end-group character
7273 \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).

```
7274 \openin 0 hyphen.cfg
7275 \ifeof0
7276 \else
7277   \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.

```
7278   \def\input #1 {%
7279     \let\input\input
7280     \a hyphen.cfg
7281     \let\input\undefined
7282   }
7283 \fi
7284 </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`.

```
7285 <bplain>\a plain.tex
7286 <blplain>\a lplain.tex
```

Finally we change the contents of `\fmtname` to indicate that this is *not* the plain format, but a format based on plain with the `babel` package preloaded.

```
7287 <bplain>\def\fmtname{babel-plain}
7288 <blplain>\def\fmtname{babel-lplain}
```

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

### 15.2 Emulating some L<sup>A</sup>T<sub>E</sub>X features

The file `babel.def` expects some definitions made in the `LATEX 2ε` style file. So, in 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).

```
7289 <(*Emulate LATEX)> ≡
7290 \def\@empty{}
```

```

7291 \def\loadlocalcfg#1{%
7292   \openin0#1.cfg
7293   \ifeof0
7294     \closein0
7295   \else
7296     \closein0
7297     {\immediate\write16{*****}%
7298      \immediate\write16{* Local config file #1.cfg used}%
7299      \immediate\write16{*}%
7300     }
7301     \input #1.cfg\relax
7302   \fi
7303   \@endofldf}

```

### 15.3 General tools

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

```

7304 \long\def\@firstofone#1{#1}
7305 \long\def\@firstoftwo#1#2{#1}
7306 \long\def\@secondoftwo#1#2{#2}
7307 \def\@nnil{\@nil}
7308 \def\@gobbletwo#1#2{}
7309 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
7310 \def\@star@or@long#1{%
7311   \@ifstar
7312   {\let\l@ngrel@x\relax#1}%
7313   {\let\l@ngrel@x\long#1}}
7314 \let\l@ngrel@x\relax
7315 \def\@car#1#2\@nil{#1}
7316 \def\@cdr#1#2\@nil{#2}
7317 \let\@typeset@protect\relax
7318 \let\protected@edef\edef
7319 \long\def\@gobble#1{}
7320 \edef\@backslashchar{\expandafter\@gobble\string\}
7321 \def\strip@prefix#1>{}
7322 \def\g@addto@macro#1#2{{%
7323   \toks@\expandafter{#1#2}%
7324   \xdef#1{\the\toks@}}}
7325 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
7326 \def\@nameuse#1{\csname #1\endcsname}
7327 \def\@ifundefined#1{%
7328   \expandafter\ifx\csname#1\endcsname\relax
7329     \expandafter\@firstoftwo
7330   \else
7331     \expandafter\@secondoftwo
7332   \fi}
7333 \def\@expandtwoargs#1#2#3{%
7334   \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
7335 \def\zap@space#1 #2{%
7336   #1%
7337   \ifx#2\@empty\else\expandafter\zap@space\fi
7338   #2}
7339 \let\bbl@trace\@gobble
7340 \def\bbl@error#1#2{%
7341   \begingroup
7342     \newlinechar=`^^J
7343     \def\{^^J(babel) }%
7344     \errhelp{#2}\errmessage{\{#1}%
7345   \endgroup}
7346 \def\bbl@warning#1{%
7347   \begingroup
7348     \newlinechar=`^^J
7349     \def\{^^J(babel) }%

```

```

7350 \message{\#1}%
7351 \endgroup}
7352 \let\bbl@infowarn\bbl@warning
7353 \def\bbl@info#1{%
7354 \begingroup
7355 \newlinechar=`^^J
7356 \def\{^^J}%
7357 \wlog{#1}%
7358 \endgroup}

```

$\LaTeX 2\epsilon$  has the command `\onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

7359 \ifx\@preamblecmds\undefined
7360 \def\@preamblecmds{}
7361 \fi
7362 \def\onlypreamble#1{%
7363 \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
7364 \@preamblecmds\do#1}}
7365 \@onlypreamble\onlypreamble

```

Mimick  $\LaTeX$ 's `\AtBeginDocument`; for this to work the user needs to add `\begindocument` to his file.

```

7366 \def\begindocument{%
7367 \@begindocumenthook
7368 \global\let\@begindocumenthook\undefined
7369 \def\do##1{\global\let##1\undefined}%
7370 \@preamblecmds
7371 \global\let\do\noexpand}
7372 \ifx\@begindocumenthook\undefined
7373 \def\@begindocumenthook{}
7374 \fi
7375 \@onlypreamble\@begindocumenthook
7376 \def\AtBeginDocument{\g@addto@macro\@begindocumenthook}

```

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

```

7377 \def\AtEndOfPackage#1{\g@addto@macro\@endofldf{#1}}
7378 \@onlypreamble\AtEndOfPackage
7379 \def\@endofldf{}
7380 \@onlypreamble\@endofldf
7381 \let\bbl@afterlang\@empty
7382 \chardef\bbl@opt@hyphenmap\z@

```

$\LaTeX$  needs to be able to switch off writing to its auxiliary files; plain doesn't have them by default. There is a trick to hide some conditional commands from the outer `\ifx`. The same trick is applied below.

```

7383 \catcode`\&=\z@
7384 \ifx&\if@filesw\@undefined
7385 \expandafter\let\csname if@filesw\expandafter\endcsname
7386 \csname iffalse\endcsname
7387 \fi
7388 \catcode`\&=4

```

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

```

7389 \def\newcommand{\@star@or@long\new@command}
7390 \def\new@command#1{%
7391 \@testopt{\@newcommand#1}0}
7392 \def\@newcommand#1[#2]{%
7393 \@ifnextchar [{\@xargdef#1[#2]}%
7394 {\@argdef#1[#2]}}
7395 \long\def\@argdef#1[#2]#3{%
7396 \@yargdef#1\@ne{#2}{#3}}
7397 \long\def\@xargdef#1[#2][#3]#4{%
7398 \expandafter\def\expandafter#1\expandafter{%

```

```

7399 \expandafter\@protected@testopt\expandafter #1%
7400 \csname\string#1\expandafter\endcsname{#3}}%
7401 \expandafter\@yargdef \csname\string#1\endcsname
7402 \tw@{#2}{#4}}
7403 \long\def\@yargdef#1#2#3{%
7404 \@tempcnta#3\relax
7405 \advance \@tempcnta \@ne
7406 \let\@hash@\relax
7407 \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
7408 \@tempcntb #2%
7409 \@whilenum\@tempcntb <\@tempcnta
7410 \do{%
7411 \edef\reserved@a{\reserved@a\@hash@\the\@tempcntb}%
7412 \advance\@tempcntb \@ne}%
7413 \let\@hash@##%
7414 \l@ngrel@x\expandafter\def\expandafter#1\reserved@a}
7415 \def\providecommand{\@star@or@long\provide@command}
7416 \def\provide@command#1{%
7417 \begingroup
7418 \escapechar\m@ne\xdef\@gtempa{\string#1}}%
7419 \endgroup
7420 \expandafter\@ifundefined\@gtempa
7421 {\def\reserved@a{\new@command#1}}%
7422 {\let\reserved@a\relax
7423 \def\reserved@a{\new@command\reserved@a}}%
7424 \reserved@a}%
7425 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
7426 \def\declare@robustcommand#1{%
7427 \edef\reserved@a{\string#1}%
7428 \def\reserved@b{#1}%
7429 \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
7430 \edef#1{%
7431 \ifx\reserved@a\reserved@b
7432 \noexpand\x@protect
7433 \noexpand#1%
7434 \fi
7435 \noexpand\protect
7436 \expandafter\noexpand\csname
7437 \expandafter\@gobble\string#1 \endcsname
7438 }%
7439 \expandafter\new@command\csname
7440 \expandafter\@gobble\string#1 \endcsname
7441 }
7442 \def\x@protect#1{%
7443 \ifx\protect\@typeset@protect\else
7444 \@x@protect#1%
7445 \fi
7446 }
7447 \catcode`\&=\z@ % Trick to hide conditionals
7448 \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`.

```

7449 \def\bbl@tempa{\csname newif\endcsname&ifin@}
7450 \catcode`\&=4
7451 \ifx\in@\undefined
7452 \def\in@#1#2{%
7453 \def\in@@##1##2##3\in@@{%
7454 \ifx\in@##2\in@false\else\in@true\fi}%
7455 \in@@#2#1\in@\in@@}
7456 \else
7457 \let\bbl@tempa\empty

```

```

7458 \fi
7459 \bbl@tempa

```

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

```

7460 \def@ifpackagewith#1#2#3#4{#3}

```

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

```

7461 \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  $\LaTeX 2_{\epsilon}$  versions; just enough to make things work in plain  $\TeX$  environments.

```

7462 \ifx\@tempcnta\@undefined
7463   \csname newcount\endcsname\@tempcnta\relax
7464 \fi
7465 \ifx\@tempcntb\@undefined
7466   \csname newcount\endcsname\@tempcntb\relax
7467 \fi

```

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

```

7468 \ifx\bye\@undefined
7469   \advance\count10 by -2\relax
7470 \fi
7471 \ifx@ifnextchar\@undefined
7472   \def@ifnextchar#1#2#3{%
7473     \let\reserved@d=#1%
7474     \def\reserved@a{#2}\def\reserved@b{#3}%
7475     \futurelet\@let@token\@ifnch}
7476   \def@ifnch{%
7477     \ifx\@let@token\@sptoken
7478       \let\reserved@c\@xifnch
7479     \else
7480       \ifx\@let@token\reserved@d
7481         \let\reserved@c\reserved@a
7482       \else
7483         \let\reserved@c\reserved@b
7484       \fi
7485     \fi
7486     \reserved@c}
7487   \def\:{\let\@sptoken= } \: % this makes \@sptoken a space token
7488   \def\:{\@xifnch} \expandafter\def\:{\futurelet\@let@token\@ifnch}
7489 \fi
7490 \def\@testopt#1#2{%
7491   \@ifnextchar[{\#1}{\#1[#2]}}
7492 \def\@protected@testopt#1{%
7493   \ifx\protect\@typeset@protect
7494     \expandafter\@testopt
7495   \else
7496     \@x@protect#1%
7497   \fi}
7498 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
7499   #2\relax}\fi}
7500 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
7501   \else\expandafter\@gobble\fi{#1}}

```

## 15.4 Encoding related macros

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



```

7502 \def\DeclareTextCommand{%
7503   \@dec@text@cmd\providecommand
7504 }
7505 \def\ProvideTextCommand{%
7506   \@dec@text@cmd\providecommand
7507 }
7508 \def\DeclareTextSymbol#1#2#3{%
7509   \@dec@text@cmd\chardef#1{#2}#3\relax
7510 }
7511 \def\@dec@text@cmd#1#2#3{%
7512   \expandafter\def\expandafter#2%
7513     \expandafter{%
7514       \csname#3-cmd\expandafter\endcsname
7515       \expandafter#2%
7516       \csname#3\string#2\endcsname
7517     }%
7518 %   \let\@ifdefinable\rc@ifdefinable
7519   \expandafter#1\csname#3\string#2\endcsname
7520 }
7521 \def\@current@cmd#1{%
7522   \ifx\protect\@typeset@protect\else
7523     \noexpand#1\expandafter\@gobble
7524   \fi
7525 }
7526 \def\@changed@cmd#1#2{%
7527   \ifx\protect\@typeset@protect
7528     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
7529       \expandafter\ifx\csname ?\string#1\endcsname\relax
7530         \expandafter\def\csname ?\string#1\endcsname{%
7531           \@changed@x@err{#1}%
7532         }%
7533       \fi
7534       \global\expandafter\let
7535         \csname\cf@encoding \string#1\expandafter\endcsname
7536         \csname ?\string#1\endcsname
7537     \fi
7538     \csname\cf@encoding\string#1%
7539     \expandafter\endcsname
7540   \else
7541     \noexpand#1%
7542   \fi
7543 }
7544 \def\@changed@x@err#1{%
7545   \errhelp{Your command will be ignored, type <return> to proceed}%
7546   \errmessage{Command \protect#1 undefined in encoding \cf@encoding}%
7547 \def\DeclareTextCommandDefault#1{%
7548   \DeclareTextCommand#1?%
7549 }
7550 \def\ProvideTextCommandDefault#1{%
7551   \ProvideTextCommand#1?%
7552 }
7553 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
7554 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
7555 \def\DeclareTextAccent#1#2#3{%
7556   \DeclareTextCommand#1{#2}[1]{\accent#3 ##1}
7557 }
7558 \def\DeclareTextCompositeCommand#1#2#3#4{%
7559   \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
7560   \edef\reserved@b{\string##1}%
7561   \edef\reserved@c{%
7562     \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
7563   \ifx\reserved@b\reserved@c
7564     \expandafter\expandafter\expandafter\ifx

```

```

7565     \expandafter\@car\reserved@a\relax\relax\@nil
7566     \@text@composite
7567     \else
7568         \edef\reserved@b##1{%
7569             \def\expandafter\noexpand
7570                 \csname#2\string#1\endcsname####1{%
7571                 \noexpand\@text@composite
7572                     \expandafter\noexpand\csname#2\string#1\endcsname
7573                     #####1\noexpand\@empty\noexpand\@text@composite
7574                     {##1}%
7575                 }%
7576             }%
7577             \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
7578         \fi
7579         \expandafter\def\csname\expandafter\string\csname
7580             #2\endcsname\string#1-\string#3\endcsname{#4}
7581         \else
7582             \errhelp{Your command will be ignored, type <return> to proceed}%
7583             \errmessage{\string\DeclareTextCompositeCommand\space used on
7584                 inappropriate command \protect#1}
7585         \fi
7586     }
7587 \def\@text@composite#1#2#3\@text@composite{%
7588     \expandafter\@text@composite@x
7589     \csname\string#1-\string#2\endcsname
7590 }
7591 \def\@text@composite@x#1#2{%
7592     \ifx#1\relax
7593         #2%
7594     \else
7595         #1%
7596     \fi
7597 }
7598 %
7599 \def\@strip@args#1:#2-#3\@strip@args{#2}
7600 \def\DeclareTextComposite#1#2#3#4{%
7601     \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
7602     \bgroup
7603         \lccode`\@=#4%
7604         \lowercase{%
7605             \egroup
7606             \reserved@a \@%
7607         }%
7608 }
7609 %
7610 \def\UseTextSymbol#1#2{#2}
7611 \def\UseTextAccent#1#2#3{}
7612 \def\@use@text@encoding#1{}
7613 \def\DeclareTextSymbolDefault#1#2{%
7614     \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
7615 }
7616 \def\DeclareTextAccentDefault#1#2{%
7617     \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
7618 }
7619 \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.

7620 \DeclareTextAccent{"}{OT1}{127}
7621 \DeclareTextAccent{'}{OT1}{19}
7622 \DeclareTextAccent{^}{OT1}{94}
7623 \DeclareTextAccent{\`}{OT1}{18}
7624 \DeclareTextAccent{\~}{OT1}{126}

```

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

```
7625 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
7626 \DeclareTextSymbol{\textquotedblright}{OT1}{`"}
7627 \DeclareTextSymbol{\textquoteleft}{OT1}{`'}
7628 \DeclareTextSymbol{\textquoteright}{OT1}{`'}
7629 \DeclareTextSymbol{\i}{OT1}{16}
7630 \DeclareTextSymbol{\ss}{OT1}{25}
```

For a couple of languages we need the  $\LaTeX$ -control sequence `\scriptsize` to be available. Because plain  $\TeX$  doesn't have such a sophisticated font mechanism as  $\LaTeX$  has, we just `\let` it to `\sevenrm`.

```
7631 \ifx\scriptsize\undefined
7632   \let\scriptsize\sevenrm
7633 \fi
```

And a few more “dummy” definitions.

```
7634 \def\language{english}%
7635 \let\bbl@opt@shorthands\@nnil
7636 \def\bbl@ifshorthand#1#2#3{#2}%
7637 \let\bbl@language@opts\@empty
7638 \ifx\babeloptionstrings\undefined
7639   \let\bbl@opt@strings\@nnil
7640 \else
7641   \let\bbl@opt@strings\babeloptionstrings
7642 \fi
7643 \def\BabelStringsDefault{generic}
7644 \def\bbl@tempa{normal}
7645 \ifx\babeloptionmath\bbl@tempa
7646   \def\bbl@mathnormal{\noexpand\textormath}
7647 \fi
7648 \def\AfterBabelLanguage#1#2{}
7649 \ifx\BabelModifiers\undefined\let\BabelModifiers\relax\fi
7650 \let\bbl@afterlang\relax
7651 \def\bbl@opt@safe{BR}
7652 \ifx\@uclclist\undefined\let\@uclclist\@empty\fi
7653 \ifx\bbl@trace\undefined\def\bbl@trace#1{}\fi
7654 \expandafter\newif\csname ifbbl@single\endcsname
7655 \chardef\bbl@bidimode\z@
7656 <</Emulate LaTeX>>
```

A proxy file:

```
7657 <*\plain>
7658 \input babel.def
7659 </\plain>
```

## 16 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  $\LaTeX$  styles*, *TUGboat* 10 (1989) #3, p. 401–406.
- [3] Yannis Haralambous, *Fonts & Encodings*, O'Reilly, 2007.
- [4] Donald E. Knuth, *The  $\TeX$ book*, Addison-Wesley, 1986.

- [5] Jukka K. Korpela, *Unicode Explained*, O'Reilly, 2006.
- [6] Leslie Lamport, *LaTeX, A document preparation System*, Addison-Wesley, 1986.
- [7] Leslie Lamport, in: *T<sub>E</sub>Xhax Digest*, Volume 89, #13, 17 February 1989.
- [8] Ken Lunde, *CJKV Information Processing*, O'Reilly, 2nd ed., 2009.
- [9] Hubert Partl, *German T<sub>E</sub>X*, *TUGboat* 9 (1988) #1, p. 70–72.
- [10] Joachim Schrod, *International L<sup>A</sup>T<sub>E</sub>X is ready to use*, *TUGboat* 11 (1990) #1, p. 87–90.
- [11] Apostolos Syropoulos, Antonis Tsolomitis and Nick Sofroniu, *Digital typography using L<sup>A</sup>T<sub>E</sub>X*, Springer, 2002, p. 301–373.
- [12] K.F. Treebus. *Tekstwijzer, een gids voor het grafisch verwerken van tekst*, SDU Uitgeverij ('s-Gravenhage, 1988).