

# Babel

Version 3.31.1640

2019/05/13

*Original author*

Johannes L. Braams

*Current maintainer*

Javier Bezos

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

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

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

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

# Contents

<b>I</b>	<b>User guide</b>	<b>4</b>
<b>1</b>	<b>The user interface</b>	<b>4</b>
1.1	Monolingual documents . . . . .	4
1.2	Multilingual documents . . . . .	5
1.3	Modifiers . . . . .	6
1.4	xelatex and luatex . . . . .	7
1.5	Troubleshooting . . . . .	7
1.6	Plain . . . . .	8
1.7	Basic language selectors . . . . .	8
1.8	Auxiliary language selectors . . . . .	9
1.9	More on selection . . . . .	10
1.10	Shorthands . . . . .	11
1.11	Package options . . . . .	14
1.12	The base option . . . . .	16
1.13	ini files . . . . .	17
1.14	Selecting fonts . . . . .	24
1.15	Modifying a language . . . . .	26
1.16	Creating a language . . . . .	26
1.17	Digits . . . . .	29
1.18	Getting the current language name . . . . .	29
1.19	Hyphenation and line breaking . . . . .	30
1.20	Selecting scripts . . . . .	31
1.21	Selecting directions . . . . .	32
1.22	Language attributes . . . . .	36
1.23	Hooks . . . . .	36
1.24	Languages supported by babel with ldf files . . . . .	37
1.25	Unicode character properties in luatex . . . . .	39
1.26	Tips, workarounds, know issues and notes . . . . .	39
1.27	Current and future work . . . . .	40
1.28	Tentative and experimental code . . . . .	41
<b>2</b>	<b>Loading languages with language.dat</b>	<b>41</b>
2.1	Format . . . . .	41
<b>3</b>	<b>The interface between the core of babel and the language definition files</b>	<b>42</b>
3.1	Guidelines for contributed languages . . . . .	43
3.2	Basic macros . . . . .	44
3.3	Skeleton . . . . .	45
3.4	Support for active characters . . . . .	46
3.5	Support for saving macro definitions . . . . .	47
3.6	Support for extending macros . . . . .	47
3.7	Macros common to a number of languages . . . . .	47
3.8	Encoding-dependent strings . . . . .	47
<b>4</b>	<b>Changes</b>	<b>51</b>
4.1	Changes in babel version 3.9 . . . . .	51
<b>II</b>	<b>Source code</b>	<b>52</b>
<b>5</b>	<b>Identification and loading of required files</b>	<b>52</b>

<b>6</b>	<b>locale directory</b>	<b>52</b>
<b>7</b>	<b>Tools</b>	<b>53</b>
7.1	Multiple languages . . . . .	56
<b>8</b>	<b>The Package File (<math>\LaTeX</math>, babel.sty)</b>	<b>57</b>
8.1	base . . . . .	57
8.2	key=value options and other general option . . . . .	59
8.3	Conditional loading of shorthands . . . . .	61
8.4	Language options . . . . .	62
<b>9</b>	<b>The kernel of Babel (babel.def, common)</b>	<b>65</b>
9.1	Tools . . . . .	65
9.2	Hooks . . . . .	67
9.3	Setting up language files . . . . .	69
9.4	Shorthands . . . . .	71
9.5	Language attributes . . . . .	80
9.6	Support for saving macro definitions . . . . .	83
9.7	Short tags . . . . .	84
9.8	Hyphens . . . . .	84
9.9	Multiencoding strings . . . . .	86
9.10	Macros common to a number of languages . . . . .	91
9.11	Making glyphs available . . . . .	92
9.11.1	Quotation marks . . . . .	92
9.11.2	Letters . . . . .	93
9.11.3	Shorthands for quotation marks . . . . .	94
9.11.4	Umlauts and tremas . . . . .	95
9.12	Layout . . . . .	96
9.13	Load engine specific macros . . . . .	97
9.14	Creating languages . . . . .	97
<b>10</b>	<b>The kernel of Babel (babel.def, only <math>\LaTeX</math>)</b>	<b>107</b>
10.1	The redefinition of the style commands . . . . .	107
10.2	Cross referencing macros . . . . .	107
10.3	Marks . . . . .	110
10.4	Preventing clashes with other packages . . . . .	111
10.4.1	ifthen . . . . .	111
10.4.2	varioref . . . . .	112
10.4.3	hhline . . . . .	113
10.4.4	hyperref . . . . .	113
10.4.5	fancyhdr . . . . .	113
10.5	Encoding and fonts . . . . .	114
10.6	Basic bidi support . . . . .	116
10.7	Local Language Configuration . . . . .	119
<b>11</b>	<b>Multiple languages (switch.def)</b>	<b>119</b>
11.1	Selecting the language . . . . .	120
11.2	Errors . . . . .	129
<b>12</b>	<b>Loading hyphenation patterns</b>	<b>130</b>
<b>13</b>	<b>Font handling with fontspec</b>	<b>135</b>

<b>14</b>	<b>Hooks for XeTeX and LuaTeX</b>	<b>139</b>
14.1	XeTeX . . . . .	139
14.2	Layout . . . . .	141
14.3	LuaTeX . . . . .	144
14.4	Southeast Asian scripts . . . . .	149
14.5	CJK line breaking . . . . .	152
14.6	Layout . . . . .	153
14.7	Auto bidi with basic and basic-r . . . . .	155
<b>15</b>	<b>Data for CJK</b>	<b>166</b>
<b>16</b>	<b>The ‘nil’ language</b>	<b>166</b>
<b>17</b>	<b>Support for Plain T<sub>E</sub>X (plain.def)</b>	<b>167</b>
17.1	Not renaming hyphen.tex . . . . .	167
17.2	Emulating some L <sup>A</sup> T <sub>E</sub> X features . . . . .	168
17.3	General tools . . . . .	168
17.4	Encoding related macros . . . . .	172
<b>18</b>	<b>Acknowledgements</b>	<b>175</b>

## Troubleshooting

Paragraph ended before \UTFviii@three@octets was complete . . . . .	4
No hyphenation patterns were preloaded for (babel) the language ‘LANG’ into the format . . . . .	5
You are loading directly a language style . . . . .	7
Unknown language ‘LANG’ . . . . .	8
Argument of \language@active@arg” has an extra } . . . . .	11
Package fontspec Warning: ‘Language ‘LANG’ not available for font ‘FONT’ with script ‘SCRIPT’ ‘Default’ language used instead’ . . . . .	25

## Part I

# User guide

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

## 1 The user interface

### 1.1 Monolingual documents

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

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

```
\documentclass{article}

\usepackage[T1]{fontenc}
\usepackage[utf8]{inputenc}

\usepackage[french]{babel}

\begin{document}

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

\end{document}
```

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

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

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

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

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

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

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

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

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

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

## 1.2 Multilingual documents

In multilingual documents, just use several options. The last one is considered the main language, activated by default. Sometimes, the main language changes the document layout (eg, spanish and french).

**EXAMPLE** In L<sup>A</sup>T<sub>E</sub>X, the preamble of the document:

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

would tell L<sup>A</sup>T<sub>E</sub>X 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:

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

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

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

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

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

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

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

**EXAMPLE** A full bilingual document follows. The main language is french, which is activated when the document begins. The package `inputenc` may be omitted with L<sup>A</sup>T<sub>E</sub>X ≥ 2018-04-01 if the encoding is UTF-8.

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

### 1.3 Modifiers

**New 3.9c** The basic behavior of some languages can be modified when loading babel by means of *modifiers*. They are set after the language name, and are prefixed with a dot (only when the language is set as package option – neither global options nor the `main` key accept 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 is a more general mechanism.

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

## 1.4 xelatex and luatex

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

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

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

```
\documentclass{article}

\usepackage[vietnamese,danish]{babel}

\begin{document}

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

\selectlanguage{vietnamese}

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

\end{document}
```

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

```
\documentclass{article}

\usepackage[russian]{babel}

\babelfont{rm}{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}
```

## 1.5 Troubleshooting

- Loading directly `sty` files in  $\text{\LaTeX}$  (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}.
```

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



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

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

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

## 1.6 Plain

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

```
\input estonian.sty
\begindocument
```

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

## 1.7 Basic language selectors

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

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

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.

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

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

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

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

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

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

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

## 1.8 Auxiliary language selectors

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

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

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

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

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

Spaces after the environment are ignored.

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

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

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

**`\begin{hyphenrules}`**    `{\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 discouraged and `otherlanguage*` (the starred version) is preferred, as the former does not take into account possible changes in

encodings of characters like, say, ' done by some languages (eg, italian, french, ukraineb). To set hyphenation exceptions, use `\babelhyphenation` (see below).

## 1.9 More on selection

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

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

It defines `\text<tag1>\{<text>\}` to be `\foreignlanguage<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.

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

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

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

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

```
\babelensure{polish}
```

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

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

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

## 1.10 Shorthands

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

**NOTE** Note the following:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

---

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

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

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

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

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

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

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

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

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

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

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

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

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

(You may also need to activate them with, for example, `\usesshorthands`.)

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

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

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

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

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

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

**Languages with only " as defined shorthand character** Albanian, Bulgarian, Danish, Dutch, Finnish, German (old and new orthography, also Austrian), Icelandic, Italian, Norwegian, Polish, Portuguese (also Brazilian), Russian, Serbian (with Latin script), Slovene, Swedish, Ukrainian, Upper Sorbian

**Basque** " ' ~

**Breton** : ; ? !

**Catalan** " ' `

**Czech** " -

**Esperanto** ^

**Estonian** " ~

**French** (all varieties) : ; ? !

**Galician** " . ' ~ < >

**Greek** ~

**Hungarian** `

**Kurmanji** ^

**Latin** " ^ =

**Slovak** " ^ ' -

**Spanish** " . < > ' ^

**Turkish** : ! =

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

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

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

## 1.11 Package options

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

**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.

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

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

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

**activegrave** Same for `.

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

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

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

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

**safe=** none | ref | bib

Some  $\TeX$  macros are redefined so that using shorthands is safe. With safe=bib only \nocite, \bibcite and \bibitem are redefined. With safe=ref only \newlabel, \ref and \pageref are redefined (as well as a few macros from varioref and ifthen). With safe=none no macro is redefined. This option is strongly recommended, because a good deal of incompatibilities and errors are related to these redefinitions – of course, in such a case you cannot use shorthands in these macros, but this is not a real problem (just use “allowed” characters).

**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 any more.

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

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

**main=**  $\langle language \rangle$

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=**  $\langle language \rangle$

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 spoiled by an unexpected .cfg file. However, if the key config is set, this file is loaded.

**showlanguages** Prints to the log the list of languages loaded when the format was created: number (remember dialects can share it), name, hyphenation file and exceptions file.



**nocase** New 3.9l Language settings for uppercase and lowercase mapping (as set by `\SetCase`) are ignored. Use only if there are incompatibilities with other packages.

**silent** New 3.9l No warnings and no *infos* are written to the log file.<sup>9</sup>

**strings=** generic | unicode | encoded |  $\langle label \rangle$  |  $\langle font\ encoding \rangle$   
Selects the encoding of strings in languages supporting this feature. Predefined labels are generic (for traditional  $\TeX$ , LICR and ASCII strings), unicode (for engines like xetex and luatex) and encoded (for special cases requiring mixed encodings). Other allowed values are font encoding codes (T1, T2A, LGR, L7X...), but only in languages supporting them. Be aware with encoded captions are protected, but they work in `\MakeUppercase` and the like (this feature misuses some internal  $\LaTeX$  tools, so use it only as a last resort).

**hyphenmap=** off | main | select | other | other\*

New 3.9g Sets the behavior of case mapping for hyphenation, provided the language defines it.<sup>10</sup> It can take the following values:

**off** deactivates this feature and no case mapping is applied;

**first** sets it at the first switching commands in the current or parent scope (typically, when the aux file is first read and at `\begin{document}`), but also the first `\selectlanguage` in the preamble), and it's the default if a single language option has been stated;<sup>11</sup>

**select** sets it only at `\selectlanguage`;

**other** also sets it at `otherlanguage`;

**other\*** also sets it at `otherlanguage*` as well as in heads and foots (if the option `headfoot` is used) and in auxiliary files (ie, at `\select@language`), and it's the default if several language options have been stated. The option `first` can be regarded as an optimized version of `other*` for monolingual documents.<sup>12</sup>

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

New 3.14 Selects the bidi algorithm to be used in luatex and xetex. See sec. 1.21.

**layout=**

New 3.16 Selects which layout elements are adapted in bidi documents. See sec. 1.21.

## 1.12 The base option

With this package option `babel` just loads some basic macros (those in `switch.def`), defines `\AfterBabelLanguage` and exits. It also selects the hyphenations 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 hyphenations patterns of a single language, too.

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

---

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

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

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

<sup>12</sup>Providing `foreign` is pointless, because the case mapping applied is that at the end of 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.

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}
```

## 1.13 ini files

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

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

```
\documentclass{book}

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

\babelfont{rm}{DejaVu Sans}

\begin{document}

\tableofcontents

\chapter{სამზარეულო და სუფრის ტრადიციები}

ქართული ტრადიციული სამზარეულო ერთ-ერთი უმდიდრესია მთელ მსოფლიოში.

\end{document}
```

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

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

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

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

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

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

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

```
\babelprovide[import,hyphenrules=+]{lao}
\babelpatterns[lao]{ໂ ນ ື ື ື ື ື ື ື ື ື ື % Random
```

Khmer clusters are rendered wrongly.

**East Asia scripts** Internal inconsistencies in script and language names must be sorted out, so you may need to set them explicitly in \babel font, as well as CJKShape. luatex does basic line breaking, but currently xetex does not (you may load zhspacing). 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...), . Actually, this is what the ldf does in japanese with luatex, because the following piece of code loads luatexja:

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

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

---

af	Afrikaans <sup>ul</sup>	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>	cs	Czech <sup>ul</sup>

cy	Welsh <sup>ul</sup>	hy	Armenian
da	Danish <sup>ul</sup>	ia	Interlingua <sup>ul</sup>
dav	Taita	id	Indonesian <sup>ul</sup>
de-AT	German <sup>ul</sup>	ig	Igbo
de-CH	German <sup>ul</sup>	ii	Sichuan Yi
de	German <sup>ul</sup>	is	Icelandic <sup>ul</sup>
dje	Zarma	it	Italian <sup>ul</sup>
dsb	Lower Sorbian <sup>ul</sup>	ja	Japanese
dua	Duala	jgo	Ngomba
dyo	Jola-Fonyi	jmc	Machame
dz	Dzongkha	ka	Georgian <sup>ul</sup>
ebu	Embu	kab	Kabyle
ee	Ewe	kam	Kamba
el	Greek <sup>ul</sup>	kde	Makonde
en-AU	English <sup>ul</sup>	kea	Kabuverdianu
en-CA	English <sup>ul</sup>	khq	Koyra Chiini
en-GB	English <sup>ul</sup>	ki	Kikuyu
en-NZ	English <sup>ul</sup>	kk	Kazakh
en-US	English <sup>ul</sup>	kkj	Kako
en	English <sup>ul</sup>	kl	Kalaallisut
eo	Esperanto <sup>ul</sup>	kln	Kalenjin
es-MX	Spanish <sup>ul</sup>	km	Khmer
es	Spanish <sup>ul</sup>	kn	Kannada <sup>ul</sup>
et	Estonian <sup>ul</sup>	ko	Korean
eu	Basque <sup>ul</sup>	kok	Konkani
ewo	Ewondo	ks	Kashmiri
fa	Persian <sup>ul</sup>	ksb	Shambala
ff	Fulah	ksf	Bafia
fi	Finnish <sup>ul</sup>	ksh	Colognian
fil	Filipino	kw	Cornish
fo	Faroese	ky	Kyrgyz
fr	French <sup>ul</sup>	lag	Langi
fr-BE	French <sup>ul</sup>	lb	Luxembourgish
fr-CA	French <sup>ul</sup>	lg	Ganda
fr-CH	French <sup>ul</sup>	lkt	Lakota
fr-LU	French <sup>ul</sup>	ln	Lingala
fur	Friulian <sup>ul</sup>	lo	Lao <sup>ul</sup>
fy	Western Frisian	lrc	Northern Luri
ga	Irish <sup>ul</sup>	lt	Lithuanian <sup>ul</sup>
gd	Scottish Gaelic <sup>ul</sup>	lu	Luba-Katanga
gl	Galician <sup>ul</sup>	luo	Luo
gsw	Swiss German	luy	Luyia
gu	Gujarati	lv	Latvian <sup>ul</sup>
guz	Gusii	mas	Masai
gv	Manx	mer	Meru
ha-GH	Hausa	mfe	Morisyen
ha-NE	Hausa <sup>1</sup>	mg	Malagasy
ha	Hausa	mgh	Makhuwa-Meetto
haw	Hawaiian	mgo	Meta'
he	Hebrew <sup>ul</sup>	mk	Macedonian <sup>ul</sup>
hi	Hindi <sup>u</sup>	ml	Malayalam <sup>ul</sup>
hr	Croatian <sup>ul</sup>	mn	Mongolian
hsb	Upper Sorbian <sup>ul</sup>	mr	Marathi <sup>ul</sup>
hu	Hungarian <sup>ul</sup>	ms-BN	Malay <sup>1</sup>

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

---

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

---

aghem	centralatlastamazight
akan	centralkurdish
albanian	chechen
american	cherokee
amharic	chiga
arabic	chinese-hans-hk
arabic-algeria	chinese-hans-mo
arabic-DZ	chinese-hans-sg
arabic-morocco	chinese-hans
arabic-MA	chinese-hant-hk
arabic-syria	chinese-hant-mo
arabic-SY	chinese-hant
armenian	chinese-simplified-hongkongsarchina
assamese	chinese-simplified-macausarchina
asturian	chinese-simplified-singapore
asu	chinese-simplified
australian	chinese-traditional-hongkongsarchina
austrian	chinese-traditional-macausarchina
azerbaijani-cyrillic	chinese-traditional
azerbaijani-cyrl	chinese
azerbaijani-latin	cognian
azerbaijani-latn	cornish
azerbaijani	croatian
bafia	czech
bambara	danish
basaa	duala
basque	dutch
belarusian	dzongkha
bemba	embu
ben	english-au
bengali	english-australia
bodo	english-ca
bosnian-cyrillic	english-canada
bosnian-cyrl	english-gb
bosnian-latin	english-newzealand
bosnian-latn	english-nz
bosnian	english-unitedkingdom
brazilian	english-unitedstates
breton	english-us
british	english
bulgarian	esperanto
burmese	estonian
canadian	ewe
cantonese	ewondo
catalan	faroes

filipino  
finnish  
french-be  
french-belgium  
french-ca  
french-canada  
french-ch  
french-lu  
french-luxembourg  
french-switzerland  
french  
friulian  
fulah  
galician  
ganda  
georgian  
german-at  
german-austria  
german-ch  
german-switzerland  
german  
greek  
gujarati  
gusii  
hausa-gh  
hausa-ghana  
hausa-ne  
hausa-niger  
hausa  
hawaiian  
hebrew  
hindi  
hungarian  
icelandic  
igbo  
inarisami  
indonesian  
interlingua  
irish  
italian  
japanese  
jolafonyi  
kabuverdianu  
kabyle  
kako  
kalaallisut  
kalenjin  
kamba  
kannada  
kashmiri  
kazakh  
khmer  
kikuyu  
kinyarwanda

konkani  
korean  
koyraborosenni  
koyrachiini  
kwasio  
kyrgyz  
lakota  
langi  
lao  
latvian  
lingala  
lithuanian  
lowersorbian  
lsorbian  
lubakatanga  
luo  
luxembourgish  
luyia  
macedonian  
machame  
makhuwameetto  
makonde  
malagasy  
malay-bn  
malay-brunei  
malay-sg  
malay-singapore  
malay  
malayalam  
maltese  
manx  
marathi  
masai  
mazanderani  
meru  
meta  
mexican  
mongolian  
morisyen  
mundang  
nama  
nepali  
newzealand  
ngiemboon  
ngomba  
norsk  
northernluri  
northernsami  
northndebele  
norwegianbokmal  
norwegiannynorsk  
nswissgerman  
nuer  
nyankole

nynorsk	serbian-latin-bosniaherzegovina
occitan	serbian-latin-kosovo
oriya	serbian-latin-montenegro
oromo	serbian-latin
ossetic	serbian-latn-ba
pashto	serbian-latn-me
persian	serbian-latn-xk
piedmontese	serbian-latn
polish	serbian
portuguese-br	shambala
portuguese-brazil	shona
portuguese-portugal	sichuanyi
portuguese-pt	sinhala
portuguese	slovak
punjabi-arab	slovene
punjabi-arabic	slovenian
punjabi-gurmukhi	soga
punjabi-guru	somali
punjabi	spanish-mexico
quechua	spanish-mx
romanian	spanish
romansh	standardmoroccantamazight
rombo	swahili
rundi	swedish
russian	swissgerman
rwa	tachelhit-latin
sakha	tachelhit-latn
samburu	tachelhit-tfng
samin	tachelhit-tifinagh
sango	tachelhit
sangu	taita
sanskrit-beng	tamil
sanskrit-bengali	tasawaq
sanskrit-deva	telugu
sanskrit-devanagari	teso
sanskrit-gujarati	thai
sanskrit-gujr	tibetan
sanskrit-kannada	tigrinya
sanskrit-knda	tongan
sanskrit-malayalam	turkish
sanskrit-mlym	turkmen
sanskrit-telu	ukenglish
sanskrit-telugu	ukrainian
sanskrit	upporsorbian
scottishgaelic	urdu
sena	usenglish
serbian-cyrillic-bosniaherzegovina	usorbian
serbian-cyrillic-kosovo	uyghur
serbian-cyrillic-montenegro	uzbek-arab
serbian-cyrillic	uzbek-arabic
serbian-cyrl-ba	uzbek-cyrillic
serbian-cyrl-me	uzbek-cyrl
serbian-cyrl-xk	uzbek-latin
serbian-cyrl	uzbek-latn



uzbek	walser
vai-latin	welsh
vai-latn	westernfrisian
vai-vai	yangben
vai-vaii	yiddish
vai	yoruba
vietnam	zarma
vietnamese	zulu afrikaans
vunjo	

---

## 1.14 Selecting fonts

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

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

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

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

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

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

```
\documentclass{article}

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

\babelprovide[import]{hebrew}

\babelfont{rm}{FreeSerif}

\begin{document}

Svenska \foreignlanguage{hebrew}{עִבְרִית} svenska.

\end{document}
```

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

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

<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:

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

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

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

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

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

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

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

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

**WARNING** Do not use `\setxxxxfont` and `\babelfont` at the same time. `\babelfont` follows the standard  $\text{\LaTeX}$  conventions to set the basic families – define `\xxdefault`, and activate it with `\xxfamily`. On the other hand, `\setxxxxfont` in `fontspec` takes a different approach, because `\xxfamily` is redefined with the family name hardcoded (so that `\xxdefault` becomes no-op). Of course, both methods are incompatible, and if you use `\setxxxxfont`, font switching with `\babelfont` just does *not* work (nor the standard `\xxdefault`, for that matter).

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

---

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

## 1.15 Modifying a language

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

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

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

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

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

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

- 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.

## 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⟩*}

Defines the internal structure of the language with some defaults: the hyphen rules, if not available, are set to the current ones, left and right hyphen mins are set to 2 and 3, but captions and date are not defined. Conveniently, babel warns you about what to do. Very likely you will find alerts like that in the log file:

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

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

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

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

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

**`import=`** *⟨language-tag⟩*

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

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

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

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

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

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

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

**captions=**  $\langle$ language-tag $\rangle$

Loads only the strings. For example:

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

**hyphenrules=**  $\langle$ language-list $\rangle$

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

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

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

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

```
\babelprovide[hyphenrules=+]{neo}  
\babelpatterns[neo]{a1 e1 i1 o1 u1}
```

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

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

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

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

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

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

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

**mapfont=** direction

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

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

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

**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. Requires `import`.

**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). Requires `import`.

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

## 1.17 Digits

**New 3.20** About thirty ini files define a field named `digits.native`. When it is present, two macros are created: `\<language>digits` and `\<language>counter` (only xetex and luatex). With the first, a string of ‘Latin’ digits are converted to the native digits of that language; the second takes a counter name as argument. With the option `maparabic` in `\babelprovide`, `\arabic` is redefined to produce the native digits (this is done *globally*, to avoid inconsistencies in, for example, page numbering, and note as well dates do not rely on `\arabic`.)

For example:

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

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

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

## 1.18 Getting the current language name

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

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

`\iflanguage`  $\{\langle language \rangle\}\{\langle true \rangle\}\{\langle false \rangle\}$

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

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

## 1.19 Hyphenation and line breaking

`\babelhyphen`  $\ast\{\langle type \rangle\}$

`\babelhyphen`  $\ast\{\langle text \rangle\}$

**New 3.9a** It is customary to classify hyphens in two types: (1) *explicit* or *hard hyphens*, which in  $\text{\TeX}$  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  $\text{\TeX}$  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  $\text{\TeX}$ , `-` 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, Portugese, 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 provide 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, Portugese and Spanish.
- `\babelhyphen{nobreak}` inserts a hard hyphen without a break after it (even if a space follows).
- `\babelhyphen{empty}` inserts a break oportunity without a hyphen at all.
- `\babelhyphen{\langle text \rangle}` is a hard “hyphen” using  $\langle text \rangle$  instead. A typical case is `\babelhyphen{/}`.

With all of them hyphenation in the rest of the word is enabled. If you don’t want enabling 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  $\text{\LaTeX}$ : (1) the character used is that set for the current font, while in  $\text{\LaTeX}$  it is hardwired to `-` (a typical value); (2) the hyphen to be used in fonts with a negative `\hyphenchar` is `-`, like in  $\text{\LaTeX}$ , but it can be changed to another value by redefining `\babeinullhyphen`; (3) a break after the hyphen is forbidden if preceded by a glue  $>0$  pt (at the beginning of a word, provided it is not immediately preceded by, say, a parenthesis).

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

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

```
\babelhyphenation{Wal-hal-la Dar-bhan-ga}
```

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

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

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

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

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

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

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

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

## 1.20 Selecting scripts

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

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

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

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



defined `\textlatin`, but it was somewhat buggy because in some cases it messed up encodings and fonts (for example, if the main Latin encoding was LY1), and therefore it has been deprecated.<sup>19</sup>

`\ensureascii`  $\langle text \rangle$

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

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

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

## 1.21 Selecting directions

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

**WARNING** The current code for **text** in `luatex` should be considered essentially stable, but, of course, it is not bug free and there could be improvements in the future, because setting bidi text has many subtleties (see for example <https://www.w3.org/TR/html-bidi/>). A basic stable version for other engines must wait. This applies to text, but **graphical** elements, including the picture environment and PDF or PS based graphics, are not yet correctly handled (far from trivial). Also, indexes and the like are under study, as well as math (there are progresses in the latter).

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

There are some package options controlling bidi writing.

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

**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 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 now limited). (They are named `basic` mainly because they only consider the intrinsic direction of scripts and weak directionality.)

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

**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-r` is available in luatex only.

```
\documentclass{article}

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

\babelprovide[import, main]{arabic}

\babelfont{rm}{FreeSerif}

\begin{document}

    وقد عرفت شبه جزيرة العرب طيلة العصر الهليني (الاعريقي) بـ
    Arabia أو Aravia (بالاعريقية 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[mapfont=direction]{arabic}

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

\begin{document}

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

\end{document}
```

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

**NOTE** Boxes are “black boxes”. Numbers inside an `\hbox` (as 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 a future a more complete method, reading recursively boxed text, may be added.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**New 3.17** Something like:

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

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

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

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

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

(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}{\footnote}.
```

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

## 1.22 Language attributes

`\languageattribute`

This is a user-level command, to be used in the preamble of a document (after `\usepackage[...]{babel}`), that declares which attributes are to be used for a given language. It takes two arguments: the first is the name of the language; the second, a (list of) attribute(s) to be used. Attributes must be set in the preamble and only once – they cannot be turned on and off. The command checks whether the language is known in this document and whether the attribute(s) are known for this language.

Very often, using a *modifier* in a package option is better.

Several language definition files use their own methods to set options. For example, `french` uses `\frenchsetup`, `magyar` (1.5) uses `\magyarOptions`; modifiers provided by `spanish` have no attribute counterparts. Macros setting options are also used (eg, `\ProsodicMarksOn` in `latin`).

## 1.23 Hooks

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

`\AddBabelHook` `{\name}{\event}{\code}`

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

Current events are the following; in some of them you can use one to three T<sub>E</sub>X parameters (`#1`, `#2`, `#3`), with the meaning given:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## 1.24 Languages supported by babel with ldf files

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

**Afrikaans** afrikaans

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

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

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

```

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

```

```
\end{document}
```

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

## 1.25 Unicode character properties in luatex

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.

**$\backslash$ 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). 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
```

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

## 1.26 Tips, workarounds, know issues and notes

- If you use the document class book *and* you use  $\backslash$ ref inside the argument of  $\backslash$ chapter (or just use  $\backslash$ ref inside  $\backslash$ MakeUppercase),  $\LaTeX$  will keep complaining about an undefined label. To prevent such problems, you could revert to using uppercase labels, you can use  $\backslash$ lowercase $\{\backslash$ ref $\{\text{foo}\}\}$  inside the argument of  $\backslash$ chapter, or, if you will not use shorthands in labels, set the safe option to none or bib.
- Both ltxdoc and babel use  $\backslash$ 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}}
```

(A recent version of inputenc is required.)



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

## 1.27 Current and future work

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  $\LaTeX$  internals.

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

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

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

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

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

## 1.28 Tentative and experimental code

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

### Old stuff

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

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

So, for example:

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

## 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

<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.

exceptions are stored in a separate file this can be indicated by naming that file *after* the file with the hyphenation patterns.

The file can contain empty lines and comments, as well as lines which start with an equals (=) sign. Such a line will instruct  $\text{\LaTeX}$  that the hyphenation patterns just processed have to be known under an alternative name. Here is an example:

```
% 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 could be set in `\extras{lang}`).

A typical error when using babel is the following:

```
No hyphenation patterns were preloaded for
the language '<lang>' into the format.
Please, configure your TeX system to add them and
rebuild the format. Now I will use the patterns
preloaded for english instead}}
```

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

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

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

The following assumptions are made:

- Some of the language-specific definitions might be used by plain  $\text{\TeX}$  users, so the files have to be coded so that they can be read by both  $\text{\LaTeX}$  and plain  $\text{\TeX}$ . The current format can be checked by looking at the value of the macro `\fmtname`.
- The common part of the babel system redefines a number of macros and environments (defined previously in the document style) to put in the names of macros that replace the previously hard-wired texts. These macros have to be defined in the language definition files.

---

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

- The language definition files must define five macros, used to activate and deactivate the language-specific definitions. These macros are `\langle lang \rangle hyphenmins`, `\langle lang \rangle captions`, `\date \langle lang \rangle`, `\extras \langle lang \rangle` and `\noextras \langle lang \rangle` (the last two may be left empty); where `\langle lang \rangle` is either the name of the language definition file or the name of the  $\LaTeX$  option that is to be used. These macros and their functions are discussed below. You must define all or none for a language (or a dialect); defining, say, `\date \langle lang \rangle` but not `\captions \langle lang \rangle` does not raise an error but can lead to unexpected results.
- When a language definition file is loaded, it can define `\l@ \langle lang \rangle` to be a dialect of `\language 0` when `\l@ \langle lang \rangle` is undefined.
- Language names must be all lowercase. If an unknown language is selected, babel will attempt setting it after lowercasing its name.
- The semantics of modifiers is not defined (on purpose). In most cases, they will just be simple separated options (eg, `spanish`), but a language might require, say, a set of options organized as a tree with suboptions (in such a case, the recommended separator is `/`).

Some recommendations:

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

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

### 3.1 Guidelines for contributed languages

Now language files are “outsourced” and are located in a separate directory (`/macros/latex/contrib/babel-contrib`), so that they are contributed directly to CTAN (please, do not send to me language styles just to upload them to CTAN). Of course, placing your style files in this directory is not mandatory, but if you want to do it, here are a few guidelines.

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

- Do not hesitate stating on the file heads you are the author and the maintainer, if you actually are. There is no need to state the babel maintainer(s) as authors if they have not contributed significantly to your language files.
- Fonts are not strictly part of a language, so they are best placed in the corresponding TeX tree. This includes not only ttf, 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: <http://www.texnia.com/incubator.html>. If you need further assistance and technical advice in the development of language styles, I am willing to help you. And of course, you can make any suggestion you like.

## 3.2 Basic macros

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

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

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

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

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

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

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

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

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

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

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

**\bbl@declare@ttribute** This is a command to be used in the language definition files for declaring a language

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

### 3.3 Skeleton

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

```

\ProvidesLanguage{<language>}
    [2016/04/23 v0.0 <Language> support from the babel system]
\LdfInit{<language>}{captions<language>}

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

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

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

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

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

\StartBabelCommands*{<language>}{date}
\SetString\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  $\TeX$ book states: “Plain  $\TeX$  includes a macro called `\dospecials` that is essentially a set macro, representing the set of all characters that have a special category code.” [2, 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,  $\langle csname \rangle$ , 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  $\langle variable \rangle$ .  
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{ $\langle control sequence \rangle$ { $\langle \TeX code \rangle$ }}` can be used to extend the definition of a macro. The macro need not be defined (ie, it can be undefined or `\relax`). This macro can, for instance, be used in adding instructions to a macro like `\extrasenglish`. Be careful when using this macro, because depending on the case the assignment could be either global (usually) or local (sometimes). That does not seem very consistent, but this behavior is preserved for backward compatibility. If you are using `etoolbox`, by Philipp Lehman, consider using the tools provided by this package instead of `\addto`.

### 3.7 Macros common to a number of languages

`\bbl@allowhyphens` In several languages compound words are used. This means that when  $\TeX$  has to hyphenate such a compound word, it only does so at the ‘-’ that is used in such words. To allow hyphenation in the rest of such a compound word, the macro `\bbl@allowhyphens` can be used.

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

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

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

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

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

### 3.8 Encoding-dependent strings

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

Furthermore, it must be activated explicitly, with the package option `strings`. If there is no `strings`, these blocks are ignored, except `\SetCases` (and except if forced as described

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



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

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

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

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

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

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

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

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

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

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

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

\EndBabelCommands
```

A real example is:

---

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

```

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

\StartBabelCommands{german,austrian}{date}
[unicode, fontenc=TU EU1 EU2, charset=utf8]
\SetString\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\monthvname{Juni}
\SetString\monthviiname{Juli}
\SetString\monthviiname{August}
\SetString\monthixname{September}
\SetString\monthxname{Oktober}
\SetString\monthxiname{November}
\SetString\monthxiiname{Dezenber}
\SetString\today{\number\day.~%
\csname month\romannumeral\month name\endcsname\space
\number\year}

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

\EndBabelCommands

```

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

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

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

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

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

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

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

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

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

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

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

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

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

#1 is replaced by the roman numeral.

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

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

```
\StartBabelCommands{turkish}{}[ot1enc, fontenc=OT1]
\SetCase
{\uccode"10=`I\relax}
{\lccode`I="10\relax}

\StartBabelCommands{turkish}{}[unicode, fontenc=TU EU1 EU2, charset=utf8]
\SetCase
{\uccode`i=`İ\relax
 \uccode`ı=`I\relax}
{\lccode`İ=`i\relax
 \lccode`I=`ı\relax}

\StartBabelCommands{turkish}{}
\SetCase
{\uccode`i="9D\relax
 \uccode"19=`I\relax}
{\lccode"9D=`i\relax
 \lccode`I="19\relax}

\EndBabelCommands
```

(Note the mapping for OT1 is not complete.)

**\SetHyphenMap** {<to-lower-macros>}

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

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

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

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

```
\SetHyphenMap{\BabelLowerMM{"100"}{"11F"}{2}{"101"}}
```

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

## 4 Changes

### 4.1 Changes in babel version 3.9

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

- `\select@language` did not set `\language`. This meant the language in force when auxiliary files were loaded was the one used in, for example, shorthands – if the language was `german`, a `\select@language{spanish}` had no effect.
- `\foreignlanguage` and `otherlanguage*` messed up `\extras<language>`. Scripts, encodings and many other things were not switched correctly.
- The `:ENC` mechanism for hyphenation patterns used the encoding of the *previous* language, not that of the language being selected.
- `'` (with `activeacute`) had the original value when writing to an auxiliary file, and things like an infinite loop could happen. It worked incorrectly with `^` (if activated) and also if deactivated.
- Active chars were not reset at the end of language options, and that led to incompatibilities between languages.
- `\textormath` raised an error with a conditional.
- `\aliasshorthand` didn't work (or only in a few and very specific cases).
- `\l@english` was defined incorrectly (using `\let` instead of `\chardef`).
- `ldf` files not bundled with `babel` were not recognized when called as global options.

## Part II

# Source code

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

## 5 Identification and loading of required files

*Code documentation is still under revision.*

The babel package after unpacking consists of the following files:

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

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

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

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

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

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

## 6 locale directory

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

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

This is a preliminary documentation.

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

Most keys are self-explanatory.

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

**version** of the ini file

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

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

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

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

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

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

## 7 Tools

```
1 <<version=3.31.1640>>
2 <<date=2019/05/13>>
```

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

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

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

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

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

<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.

`\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.

```

28 \def\bbl@tempa#1{%
29   \long\def\bbl@trim##1##2{%
30     \futurelet\bbl@trim@a\bbl@trim@c##2\@nil\@nil#1\@nil\relax{##1}}%
31   \def\bbl@trim@c{%
32     \ifx\bbl@trim@a\@sptoken
33       \expandafter\bbl@trim@b
34     \else
35       \expandafter\bbl@trim@b\expandafter#1%
36     \fi}%
37   \long\def\bbl@trim@b#1##1 \@nil{\bbl@trim@i##1}}
38 \bbl@tempa{ }
39 \long\def\bbl@trim@i#1\@nil#2\relax#3{#3{#1}}
40 \long\def\bbl@trim@def#1{\bbl@trim{\def#1}}

```

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

```

41 \def\bbl@ifunset#1{%
42   \expandafter\ifx\csname#1\endcsname\relax
43     \expandafter\@firstoftwo
44   \else
45     \expandafter\@secondoftwo
46   \fi}
47 \bbl@ifunset{ifcsname}%
48 {}%
49 {\def\bbl@ifunset#1{%
50   \ifcsname#1\endcsname
51     \expandafter\ifx\csname#1\endcsname\relax
52       \bbl@afterelse\expandafter\@firstoftwo
53     \else
54       \bbl@afterfi\expandafter\@secondoftwo
55     \fi
56   \else
57     \expandafter\@firstoftwo
58   \fi}}

```

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

```

59 \def\bbl@ifblank#1{%
60   \bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
61 \long\def\bbl@ifblank@i#1#2\@nil#3#4#5\@nil{#4}

```

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

```

62 \def\bbl@forkv#1#2{%
63   \def\bbl@kvcmd##1##2##3{#2}%
64   \bbl@kvnext#1,\@nil,}
65 \def\bbl@kvnext#1,{%
66   \ifx\@nil#1\relax\else
67     \bbl@ifblank{#1}{\bbl@forkv@eq#1=\@empty=\@nil{#1}}%
68   \expandafter\bbl@kvnext

```

```

69 \fi}
70 \def\bbl@forkv@eq#1=#2=#3\@nil#4{%
71 \bbl@trim@def\bbl@forkv@a{#1}%
72 \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).

```

73 \def\bbl@vforeach#1#2{%
74 \def\bbl@forcmd##1{#2}%
75 \bbl@fornext#1,\@nil,}
76 \def\bbl@fornext#1,{%
77 \ifx\@nil#1\relax\else
78 \bbl@ifblank{#1}{\bbl@trim\bbl@forcmd{#1}}%
79 \expandafter\bbl@fornext
80 \fi}
81 \def\bbl@foreach#1{\expandafter\bbl@vforeach\expandafter{#1}}

```

`\bbl@replace`

```

82 \def\bbl@replace#1#2#3{% in #1 -> repl #2 by #3
83 \toks@{}}%
84 \def\bbl@replace@aux##1#2##2#2{%
85 \ifx\bbl@nil##2%
86 \toks@\expandafter{\the\toks@##1}%
87 \else
88 \toks@\expandafter{\the\toks@##1#3}%
89 \bbl@afterfi
90 \bbl@replace@aux##2#2%
91 \fi}%
92 \expandafter\bbl@replace@aux#1#2\bbl@nil#2%
93 \edef#1{\the\toks@}}

```

`\bbl@exp`

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

```

94 \def\bbl@exp#1{%
95 \begingroup
96 \let\ \noexpand
97 \def\<##1>\expandafter\noexpand\csname##1\endcsname}%
98 \edef\bbl@exp@aux{\endgroup#1}%
99 \bbl@exp@aux}

```

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

```

100 \def\bbl@ifsamestring#1#2{%
101 \begingroup
102 \protected@edef\bbl@tempb{#1}%
103 \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
104 \protected@edef\bbl@tempc{#2}%
105 \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
106 \ifx\bbl@tempb\bbl@tempc
107 \aftergroup\@firstoftwo
108 \else
109 \aftergroup\@secondoftwo
110 \fi
111 \endgroup}
112 \chardef\bbl@engine=%

```



```

113 \ifx\directlua\@undefined
114 \ifx\XeTeXinputencoding\@undefined
115 \z@
116 \else
117 \tw@
118 \fi
119 \else
120 \@ne
121 \fi
122 <</Basic macros>>

```

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

```

123 <<*Make sure ProvidesFile is defined>> ≡
124 \ifx\ProvidesFile\@undefined
125 \def\ProvidesFile#1[#2 #3 #4]{%
126 \wlog{File: #1 #4 #3 <#2>}%
127 \let\ProvidesFile\@undefined}
128 \fi
129 <</Make sure ProvidesFile is defined>>

```

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

```

130 <<*Load patterns in luatex>> ≡
131 \ifx\directlua\@undefined\else
132 \ifx\bbl@luapatterns\@undefined
133 \input luababel.def
134 \fi
135 \fi
136 <</Load patterns in luatex>>

```

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

```

137 <<*Load macros for plain if not LaTeX>> ≡
138 \ifx\AtBeginDocument\@undefined
139 \input plain.def\relax
140 \fi
141 <</Load macros for plain if not LaTeX>>

```

## 7.1 Multiple languages

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

```

142 <<*Define core switching macros>> ≡
143 \ifx\language\@undefined
144 \csname newcount\endcsname\language
145 \fi
146 <</Define core switching macros>>

```

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

`\addlanguage` To add languages to  $\TeX$ 's memory plain  $\TeX$  version 3.0 supplies `\newlanguage`, in a pre-3.0 environment a similar macro has to be provided. For both cases a new macro is defined here, because the original `\newlanguage` was defined to be `\outer`.

For a format based on plain version 2.x, the definition of `\newlanguage` can not be copied because `\count 19` is used for other purposes in these formats. Therefore `\addlanguage` is defined using a definition based on the macros used to define `\newlanguage` in plain  $\TeX$  version 3.0.

For formats based on plain version 3.0 the definition of `\newlanguage` can be simply copied, removing `\outer`. Plain  $\TeX$  version 3.0 uses `\count 19` for this purpose.

```

147 <<*Define core switching macros>> ≡
148 \ifx\newlanguage\undefined
149   \csname newcount\endcsname\last@language
150   \def\addlanguage#1{%
151     \global\advance\last@language\@ne
152     \ifnum\last@language<\@ccclvi
153       \else
154         \errmessage{No room for a new \string\language!}%
155       \fi
156     \global\chardef#1\last@language
157     \wlog{\string#1 = \string\language\the\last@language}}
158 \else
159   \countdef\last@language=19
160   \def\addlanguage{\alloc@9\language\chardef\@ccclvi}
161 \fi
162 <</Define core switching macros>>

```

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

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

## 8 The Package File ( $\TeX$ , `babel.sty`)

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

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

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

### 8.1 base

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

```

163 <*package>
164 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
165 \ProvidesPackage{babel}[\<<date>> \<<version>> The Babel package]

```

```

166 \@ifpackagewith{babel}{debug}
167 {\providecommand\bbl@trace[1]{\message{^^J[ #1 ]}}}%
168 \let\bbl@debug\@firstofone}
169 {\providecommand\bbl@trace[1]}%
170 \let\bbl@debug\@gobble}
171 \ifx\bbl@switchflag\@undefined % Prevent double input
172 \let\bbl@switchflag\relax
173 \input switch.def\relax
174 \fi
175 <<Load patterns in luatex>>
176 <<Basic macros>>
177 \def\AfterBabelLanguage#1{%
178 \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.

```

179 \ifx\bbl@languages\@undefined\else
180 \begingroup
181 \catcode`\^^I=12
182 \@ifpackagewith{babel}{showlanguages}{%
183 \begingroup
184 \def\bbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}}%
185 \wlog{<*languages>}%
186 \bbl@languages
187 \wlog{</languages>}%
188 \endgroup}{%
189 \endgroup
190 \def\bbl@elt#1#2#3#4{%
191 \ifnum#2=\z@
192 \gdef\bbl@nulllanguage{#1}%
193 \def\bbl@elt##1##2##3##4{}}%
194 \fi}%
195 \bbl@languages
196 \fi
197 \ifodd\bbl@engine
198 \def\bbl@activate@preotf{%
199 \let\bbl@activate@preotf\relax % only once
200 \directlua{
201 Babel = Babel or {}
202 %
203 function Babel.pre_otfload_v(head)
204 if Babel.numbers and Babel.digits_mapped then
205 head = Babel.numbers(head)
206 end
207 if Babel.bidi_enabled then
208 head = Babel.bidi(head, false, dir)
209 end
210 return head
211 end
212 %
213 function Babel.pre_otfload_h(head, gc, sz, pt, dir)
214 if Babel.numbers and Babel.digits_mapped then
215 head = Babel.numbers(head)
216 end
217 if Babel.fixboxdirs then % Temporary!
218 head = Babel.fixboxdirs(head)
219 end
220 if Babel.bidi_enabled then
221 head = Babel.bidi(head, false, dir)

```

```

222         end
223         return head
224     end
225     %
226     luatexbase.add_to_callback('pre_linebreak_filter',
227         Babel.pre_otfload_v,
228         'Babel.pre_otfload_v',
229         luatexbase.priority_in_callback('pre_linebreak_filter',
230             'luaotfload.node_processor') or nil)
231     %
232     luatexbase.add_to_callback('hpack_filter',
233         Babel.pre_otfload_h,
234         'Babel.pre_otfload_h',
235         luatexbase.priority_in_callback('hpack_filter',
236             'luaotfload.node_processor') or nil)
237     }}
238     \let\bbl@tempa\relax
239     \@ifpackagewith{babel}{bidi=basic}%
240     {\def\bbl@tempa{basic}}%
241     {\@ifpackagewith{babel}{bidi=basic-r}%
242     {\def\bbl@tempa{basic-r}}%
243     {}}
244     \ifx\bbl@tempa\relax\else
245         \let\bbl@beforeforeign\leavevmode
246         \AtEndOfPackage{\EnableBabelHook{babel-bidi}}%
247         \RequirePackage{luatexbase}%
248         \directlua{
249             require('babel-data-bidi.lua')
250             require('babel-bidi-\bbl@tempa.lua')
251         }
252         \bbl@activate@preotf
253     \fi
254 \fi

```

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

```

255 \bbl@trace{Defining option 'base'}
256 \@ifpackagewith{babel}{base}{%
257     \ifx\directlua\@undefined
258         \DeclareOption*{\bbl@patterns{\CurrentOption}}%
259     \else
260         \DeclareOption*{\bbl@patterns@lua{\CurrentOption}}%
261     \fi
262     \DeclareOption{base}{}%
263     \DeclareOption{showlanguages}{}%
264     \ProcessOptions
265     \global\expandafter\let\csname opt@babel.sty\endcsname\relax
266     \global\expandafter\let\csname ver@babel.sty\endcsname\relax
267     \global\let\@ifl@ter@@\@ifl@ter
268     \def\@ifl@ter#1#2#3#4#5{\global\let\@ifl@ter\@ifl@ter@@}%
269     \endinput}{}%

```

## 8.2 key=value options and other general option

The following macros extract language modifiers, and only real package options are kept in the option list. Modifiers are saved and assigned to `\BabelModifiers` at `\bbl@load@language`; when no modifiers have been given, the former is `\relax`. How

modifiers are handled are left to language styles; they can use `\in@`, loop them with `\@for` or `\loadkeyval`, for example.

```

270 \bbl@trace{key=value and another general options}
271 \bbl@csarg\let{tempa\expandafter}\csname opt@babel.sty\endcsname
272 \def\bbl@tempb#1.#2{%
273   #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%
274 \def\bbl@tempd#1.#2@\nnil{%
275   \ifx\@empty#2%
276     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
277   \else
278     \in@{=}{#1}\ifin@
279     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
280   \else
281     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
282     \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
283   \fi
284 \fi}
285 \let\bbl@tempc\@empty
286 \bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty@\nnil}
287 \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.

```

288 \DeclareOption{KeepShorthandsActive}{}
289 \DeclareOption{activeacute}{}
290 \DeclareOption{activegrave}{}
291 \DeclareOption{debug}{}
292 \DeclareOption{noconfigs}{}
293 \DeclareOption{showlanguages}{}
294 \DeclareOption{silent}{}
295 \DeclareOption{mono}{}
296 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
297 <<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.

```

298 \let\bbl@opt@shorthands@\nnil
299 \let\bbl@opt@config@\nnil
300 \let\bbl@opt@main@\nnil
301 \let\bbl@opt@headfoot@\nnil
302 \let\bbl@opt@layout@\nnil

```

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

```

303 \def\bbl@tempa#1=#2\bbl@tempa{%
304   \bbl@csarg\ifx{opt@#1}\@nnil
305     \bbl@csarg\edef{opt@#1}{#2}%
306   \else
307     \bbl@error{%
308       Bad option `#1=#2'. Either you have misspelled the\\%
309       key or there is a previous setting of `#1'}{%
310       Valid keys are `shorthands', `config', `strings', `main',\\%
311       `headfoot', `safe', `math', among others.}
312   \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.

```

313 \let\bbl@language@opts\@empty
314 \DeclareOption*{%
315   \bbl@xin@{\string=}\CurrentOption}%
316   \ifin@
317     \expandafter\bbl@tempa\CurrentOption\bbl@tempa
318   \else
319     \bbl@add@list\bbl@language@opts{\CurrentOption}%
320   \fi}

```

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

```

321 \ProcessOptions*

```

### 8.3 Conditional loading of shorthands

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

```

322 \bbl@trace{Conditional loading of shorthands}
323 \def\bbl@sh@string#1{%
324   \ifx#1\@empty\else
325     \ifx#1t\string~%
326     \else\ifx#1c\string,%
327     \else\string#1%
328   \fi\fi
329   \expandafter\bbl@sh@string
330 \fi}
331 \ifx\bbl@opt@shorthands\@nnil
332   \def\bbl@ifshorthand#1#2#3{#2}%
333 \else\ifx\bbl@opt@shorthands\@empty
334   \def\bbl@ifshorthand#1#2#3{#3}%
335 \else

```

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

```

336 \def\bbl@ifshorthand#1{%
337   \bbl@xin@{\string#1}\bbl@opt@shorthands}%
338   \ifin@
339     \expandafter\@firstoftwo
340   \else
341     \expandafter\@secondoftwo
342   \fi}

```

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

```

343 \edef\bbl@opt@shorthands{%
344   \expandafter\bbl@sh@string\bbl@opt@shorthands\@empty}%

```

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

```

345 \bbl@ifshorthand{'}%
346   {\PassOptionsToPackage{activeacute}{babel}}{}
347 \bbl@ifshorthand{`}%

```

```

348     {\PassOptionsToPackage{activegrave}{babel}}{}
349 \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.

```

350 \ifx\bbl@opt@headfoot\@nnil\else
351   \g@addto@macro{\@resetactivechars}%
352   \set@typeset@protect
353   \expandafter\select@language@x\expandafter{\bbl@opt@headfoot}%
354   \let\protect\noexpand}
355 \fi

```

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

```

356 \ifx\bbl@opt@safe\@undefined
357   \def\bbl@opt@safe{BR}
358 \fi
359 \ifx\bbl@opt@main\@nnil\else
360   \edef\bbl@language@opts{%
361     \ifx\bbl@language@opts\@empty\else\bbl@language@opts,\fi
362     \bbl@opt@main}
363 \fi

```

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

```

364 \bbl@trace{Defining IfBabelLayout}
365 \ifx\bbl@opt@layout\@nnil
366   \newcommand\IfBabelLayout[3]{#3}%
367 \else
368   \newcommand\IfBabelLayout[1]{%
369     \@expandtwoargs\in@{.#1.}{.\bbl@opt@layout.}%
370     \ifin@
371       \expandafter\@firstoftwo
372     \else
373       \expandafter\@secondoftwo
374     \fi}
375 \fi

```

## 8.4 Language options

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

```

376 \bbl@trace{Language options}
377 \let\bbl@afterlang\relax
378 \let\BabelModifiers\relax
379 \let\bbl@loaded\@empty
380 \def\bbl@load@language#1{%
381   \InputIfFileExists{#1.ldf}%
382   {\edef\bbl@loaded{\CurrentOption
383     \ifx\bbl@loaded\@empty\else,\bbl@loaded\fi}%
384     \expandafter\let\expandafter\bbl@afterlang
385     \csname\CurrentOption.ldf-h@@k\endcsname
386     \expandafter\let\expandafter\BabelModifiers
387     \csname bbl@mod@\CurrentOption\endcsname}%
388   {\bbl@error{%
389     Unknown option '\CurrentOption'. Either you misspelled it\\%

```

```

390 or the language definition file \CurrentOption.ldf was not found}{%
391 Valid options are: shorthands=, KeepShorthandsActive,\%
392 activeacute, activegrave, noconfigs, safe=, main=, math=\%
393 headfoot=, strings=, config=, hyphenmap=, or a language name.}}

```

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

```

394 \def\bbl@try@load@lang#1#2#3{%
395   \IfFileExists{\CurrentOption.ldf}%
396     {\bbl@load@language{\CurrentOption}}%
397     {\bbl@load@language{#2}#3}}
398 \DeclareOption{afrikaans}{\bbl@try@load@lang{}{dutch}}
399 \DeclareOption{brazil}{\bbl@try@load@lang{}{portuges}}
400 \DeclareOption{brazilian}{\bbl@try@load@lang{}{portuges}}
401 \DeclareOption{hebrew}{%
402   \input{rlbabel.def}%
403   \bbl@load@language{hebrew}}
404 \DeclareOption{hungarian}{\bbl@try@load@lang{}{magyar}}
405 \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}}
406 \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}}
407 \DeclareOption{polutonikogreek}{%
408   \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}}
409 \DeclareOption{portuguese}{\bbl@try@load@lang{}{portuges}}
410 \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}}
411 \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}}
412 \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.

```

413 \ifx\bbl@opt@config\@nnil
414   \@ifpackagewith{babel}{noconfigs}}}%
415   {\InputIfFileExists{bblopts.cfg}%
416     {\typeout{*****^J%
417               * Local config file bblopts.cfg used^^J%
418               *}}%
419     {}}}%
420 \else
421   \InputIfFileExists{\bbl@opt@config.cfg}%
422     {\typeout{*****^J%
423               * Local config file \bbl@opt@config.cfg used^^J%
424               *}}%
425     {\bbl@error{%
426       Local config file '\bbl@opt@config.cfg' not found}{%
427       Perhaps you misspelled it.}}}%
428 \fi

```

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

```

429 \bbl@for\bbl@tempa\bbl@language@opts{%
430   \bbl@ifunset{ds@\bbl@tempa}%
431     {\edef\bbl@tempb{%
432       \noexpand\DeclareOption
433       {\bbl@tempa}%

```



```

434      {\noexpand\bbload@language{\bb@tempa}}}%
435      \bb@tempb}%
436      \@empty}

```

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

```

437 \bb@foreach\@classoptionslist{%
438   \bb@ifunset{ds@#1}%
439     {\IfFileExists{#1.ldf}%
440      {\DeclareOption{#1}{\bbload@language{#1}}}%
441      {}}%
442   {}}

```

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

```

443 \ifx\bbload@opt@main\@nnil\else
444   \expandafter
445   \let\expandafter\bbloadmain\csname ds@\bbload@opt@main\endcsname
446   \DeclareOption{\bbload@opt@main}{}
447 \fi

```

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

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

```

448 \def\AfterBabelLanguage#1{%
449   \bb@ifsamestring\CurrentOption{#1}{\global\bbadd\bb@afterlang}{}}
450 \DeclareOption*{}
451 \ProcessOptions*

```

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

```

452 \ifx\bbload@opt@main\@nnil
453   \edef\bbload@tempa{\@classoptionslist,\bbload@language@opts}
454   \let\bbload@tempc\@empty
455   \bb@for\bbload@tempb\bbload@tempa{%
456     \bbloadxin@{\bbload@tempb},{\bbload@loaded},}%
457   \ifin\edef\bbload@tempc{\bbload@tempb}\fi}
458 \def\bbload@tempa#1,#2\@nnil{\def\bbload@tempb{#1}}
459 \expandafter\bbload@tempa\bbload@loaded,\@nnil
460 \ifx\bbload@tempb\bbload@tempc\else
461   \bbload@warning{%
462     Last declared language option is '\bbload@tempc',\%
463     but the last processed one was '\bbload@tempb'.\%
464     The main language cannot be set as both a global\%
465     and a package option. Use 'main=\bbload@tempc' as\%
466     option. Reported}%
467   \fi
468 \else
469   \DeclareOption{\bbload@opt@main}{\bbload@loadmain}
470   \ExecuteOptions{\bbload@opt@main}
471   \DeclareOption*{}
472   \ProcessOptions*
473 \fi
474 \def\AfterBabelLanguage{%

```

```

475 \bbl@error
476 {Too late for \string\AfterBabelLanguage}%
477 {Languages have been loaded, so I can do nothing}}

```

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

```

478 \ifx\bbl@main@language\@undefined
479 \bbl@info{%
480   You haven't specified a language. I'll use 'nil'\%
481   as the main language. Reported}
482 \bbl@load@language{nil}
483 \fi
484 \</package>
485 \*core>

```

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

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

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

### 9.1 Tools

```

486 \ifx\ldf@quit\@undefined
487 \else
488   \expandafter\endinput
489 \fi
490 <<Make sure ProvidesFile is defined>>
491 \ProvidesFile{babel.def}[\<<date>>] <<version>> Babel common definitions]
492 <<Load macros for plain if not LaTeX>>

```

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

`\BabelModifiers` can be set too (but not sure it works).

```

493 \ifx\bbl@ifshorthand\@undefined
494   \let\bbl@opt@shorthands\@nnil
495   \def\bbl@ifshorthand#1#2#3{#2}%
496   \let\bbl@language@opts\@empty
497   \ifx\babeloptionstrings\@undefined
498     \let\bbl@opt@strings\@nnil
499   \else

```

```

500 \let\bbl@opt@strings\babeloptionstrings
501 \fi
502 \def\BabelStringsDefault{generic}
503 \def\bbl@tempa{normal}
504 \ifx\babeloptionmath\bbl@tempa
505 \def\bbl@mathnormal{\noexpand\textormath}
506 \fi
507 \def\AfterBabelLanguage#1#2{}
508 \ifx\BabelModifiers\@undefined\let\BabelModifiers\relax\fi
509 \let\bbl@afterlang\relax
510 \def\bbl@opt@safe{BR}
511 \ifx\@uclclist\@undefined\let\@uclclist\@empty\fi
512 \ifx\bbl@trace\@undefined\def\bbl@trace#1{}\fi
513 \fi

And continue.
514 \ifx\bbl@switchflag\@undefined % Prevent double input
515 \let\bbl@switchflag\relax
516 \input switch.def\relax
517 \fi
518 \bbl@trace{Compatibility with language.def}
519 \ifx\bbl@languages\@undefined
520 \ifx\directlua\@undefined
521 \openin1 = language.def
522 \ifeof1
523 \closein1
524 \message{I couldn't find the file language.def}
525 \else
526 \closein1
527 \begingroup
528 \def\addlanguage#1#2#3#4#5{%
529 \expandafter\ifx\csname lang@#1\endcsname\relax\else
530 \global\expandafter\let\csname l@#1\endcsname\expandafter\endcsname
531 \csname lang@#1\endcsname
532 \fi}%
533 \def\uselanguage#1{}\%
534 \input language.def
535 \endgroup
536 \fi
537 \fi
538 \chardef\l@english\z@
539 \fi
540 <<Load patterns in luatex>>
541 <<Basic macros>>

```

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

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

```

542 \def\addto#1#2{%
543 \ifx#1\@undefined
544 \def#1{#2}%
545 \else
546 \ifx#1\relax

```

```

547     \def#1{#2}%
548     \else
549     {\toks@\expandafter{#1#2}%
550      \xdef#1{\the\toks@}}%
551     \fi
552 \fi}

```

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

```

553 \def\bbl@withactive#1#2{%
554   \begingroup
555   \lccode`~=`#2\relax
556   \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).

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

```

557 \def\bbl@redefine#1{%
558   \edef\bbl@tempa{\bbl@stripslash#1}%
559   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
560   \expandafter\def\csname\bbl@tempa\endcsname}

```

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

```

561 \@onlypreamble\bbl@redefine

```

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

```

562 \def\bbl@redefine@long#1{%
563   \edef\bbl@tempa{\bbl@stripslash#1}%
564   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
565   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
566 \@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`.

```

567 \def\bbl@redefineroobust#1{%
568   \edef\bbl@tempa{\bbl@stripslash#1}%
569   \bbl@ifunset{\bbl@tempa\space}%
570   {\expandafter\let\csname org@\bbl@tempa\endcsname#1%
571    \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
572   {\bbl@exp{\let\<org@\bbl@tempa>\<\bbl@tempa\space>}}}%
573   \@namedef{\bbl@tempa\space}}

```

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

```

574 \@onlypreamble\bbl@redefineroobust

```

## 9.2 Hooks

Note they are loaded in `babel.def`. `switch.def` only provides a “hook” for hooks (with a default value which is a no-op, below). Admittedly, the current implementation is a somewhat simplistic and does very little to catch errors, but it is intended for developers,

after all. `\bbl@usehooks` is the commands used by babel to execute hooks defined for an event.

```

575 \bbl@trace{Hooks}
576 \def\AddBabelHook#1#2{%
577   \bbl@ifunset{bbl@hk@#1}{\EnableBabelHook{#1}}{}%
578   \def\bbl@tempa##1,#2=##2,##3\@empty{\def\bbl@tempb{##2}}%
579   \expandafter\bbl@tempa\bbl@evargs,#2=,\@empty
580   \bbl@ifunset{bbl@ev@#1@#2}%
581     {\bbl@csarg\bbl@add{ev@#2}{\bbl@elt{#1}}}%
582     \bbl@csarg\newcommand}%
583     {\bbl@csarg\let{ev@#1@#2}\relax
584     \bbl@csarg\newcommand}%
585     {ev@#1@#2}[\bbl@tempb]}
586 \def\EnableBabelHook#1{\bbl@csarg\let{hk@#1}\@firstofone}
587 \def\DisableBabelHook#1{\bbl@csarg\let{hk@#1}\@gobble}
588 \def\bbl@usehooks#1#2{%
589   \def\bbl@elt##1{%
590     \@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1}#2}}%
591   \@nameuse{bbl@ev@#1}}

```

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).

```

592 \def\bbl@evargs{,% <- don't delete this comma
593   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
594   adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
595   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
596   hyphenation=2,initiateactive=3,afterreset=0,foreign=0,foreign*=0}

```

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

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

```

597 \bbl@trace{Defining babelensure}
598 \newcommand\babelensure[2][{}% TODO - revise test files
599   \AddBabelHook{babel-ensure}{afterextras}{%
600     \ifcase\bbl@select@type
601       \@nameuse{bbl@e@\languagename}%
602     \fi}%
603   \begingroup
604     \let\bbl@ens@include\@empty
605     \let\bbl@ens@exclude\@empty
606     \def\bbl@ens@fontenc{\relax}%
607     \def\bbl@tempb##1{%
608       \ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}%
609     \edef\bbl@tempa{\bbl@tempb#1\@empty}%
610     \def\bbl@tempb##1=##2\@{\@namedef{bbl@ens@##1}{##2}}%
611     \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
612     \def\bbl@tempc{\bbl@ensure}%
613     \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%

```

```

614 \expandafter{\bbl@ens@include}}%
615 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
616 \expandafter{\bbl@ens@exclude}}%
617 \toks@\expandafter{\bbl@tempc}%
618 \bbl@exp{%
619 \endgroup
620 \def\<bbl@e@#2>{\the\toks@\bbl@ens@fontenc}}%
621 \def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
622 \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
623 \ifx##1\@undefined
624 \edef##1{\noexpand\bbl@nocaption
625 {\bbl@stripslash##1}{\language\name\bbl@stripslash##1}}%
626 \fi
627 \ifx##1\@empty\else
628 \in@{##1}{#2}%
629 \ifin@else
630 \bbl@ifunset{\bbl@ensure@\language\name}%
631 {\bbl@exp{%
632 \\\DeclareRobustCommand\<bbl@ensure@\language\name>[1]{%
633 \\\foreignlanguage{\language\name}%
634 {\ifx\relax#3\else
635 \\\fontencoding{#3}\selectfont
636 \fi
637 #####1}}}%
638 }%
639 \toks@\expandafter{##1}%
640 \edef##1{%
641 \bbl@csarg\noexpand{\ensure@\language\name}%
642 {\the\toks@}}%
643 \fi
644 \expandafter\bbl@tempb
645 \fi}%
646 \expandafter\bbl@tempb\bbl@captionslist\today\@empty
647 \def\bbl@tempa##1{% elt for include list
648 \ifx##1\@empty\else
649 \bbl@csarg\in@{\ensure@\language\name\expandafter}\expandafter{##1}%
650 \ifin@else
651 \bbl@tempb##1\@empty
652 \fi
653 \expandafter\bbl@tempa
654 \fi}%
655 \bbl@tempa#1\@empty}
656 \def\bbl@captionslist{%
657 \prefacename\refname\abstractname\bibname\chaptername\appendixname
658 \contentsname\listfigurename\listtablename\indexname\figurename
659 \tablename\partname\encname\ccname\headtoname\pagename\seename
660 \alsoname\proofname\glossaryname}

```

### 9.3 Setting up language files

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

At the start of processing a language definition file we always check the category code of the at-sign. We make sure that it is a ‘letter’ during the processing of the file. We also save its name as the last called option, even if not loaded.

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

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

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

Finally we check `\originalTeX`.

```

661 \bbl@trace{Macros for setting language files up}
662 \def\bbl@ldfinit{%
663   \let\bbl@screset\@empty
664   \let\BabelStrings\bbl@opt@string
665   \let\BabelOptions\@empty
666   \let\BabelLanguages\relax
667   \ifx\originalTeX\@undefined
668     \let\originalTeX\@empty
669   \else
670     \originalTeX
671   \fi}
672 \def\LdfInit#1#2{%
673   \chardef\atcatcode=\catcode`\@
674   \catcode`\@=11\relax
675   \chardef\eqcatcode=\catcode`\=
676   \catcode`\==12\relax
677   \expandafter\if\expandafter\@backslashchar
678     \expandafter\@car\string#2\@nil
679   \ifx#2\@undefined\else
680     \ldf@quit{#1}%
681   \fi
682 \else
683   \expandafter\ifx\csname#2\endcsname\relax\else
684     \ldf@quit{#1}%
685   \fi
686 \fi
687 \bbl@ldfinit}

```

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

```

688 \def\ldf@quit#1{%
689   \expandafter\main@language\expandafter{#1}%
690   \catcode`\@=\atcatcode \let\atcatcode\relax
691   \catcode`\==\eqcatcode \let\eqcatcode\relax
692   \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.

```

693 \def\bbl@afterldf#1{%
694   \bbl@afterlang
695   \let\bbl@afterlang\relax
696   \let\BabelModifiers\relax

```

```

697 \let\bbl@screset\relax}%
698 \def\ldf@finish#1{%
699 \loadlocalcfg{#1}%
700 \bbl@afterldf{#1}%
701 \expandafter\main@language\expandafter{#1}%
702 \catcode\@=\atcatcode \let\atcatcode\relax
703 \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  $\LaTeX$ .

```

704 \@onlypreamble\LdfInit
705 \@onlypreamble\ldf@quit
706 \@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.

```

707 \def\main@language#1{%
708 \def\bbl@main@language{#1}%
709 \let\language\main@language
710 \bbl@id@assign
711 \chardef\localeid\@nameuse{bbl@id@\language}%
712 \bbl@patterns{\language}}

```

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

```

713 \AtBeginDocument{%
714 \expandafter\selectlanguage\expandafter{\bbl@main@language}%
715 \ifcase\bbl@engine\or\pagedir\bodydir\fi} % TODO - a better place

```

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

```

716 \def\select@language#1{%
717 \ifcase\bbl@select@type
718 \bbl@ifsamestring\language{#1}{\select@language{#1}}%
719 \else
720 \select@language{#1}%
721 \fi}

```

## 9.4 Shorthands

`\bbl@add@special` The macro `\bbl@add@special` is used to add a new character (or single character control sequence) to the macro `\dospecials` (and `\@sanitize` if  $\LaTeX$  is used). It is used only at one place, namely when `\initiate@active@char` is called (which is ignored if the char has been made active before). Because `\@sanitize` can be undefined, we put the definition inside a conditional.

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

```

722 \bbl@trace{Shorhands}
723 \def\bbl@add@special#1{% 1:a macro like \", \?, etc.
724 \bbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
725 \bbl@ifunset{@sanitize}{\bbl@add\@sanitize{\@makeother#1}}%
726 \ifx\nfss@catcodes\undefined\else % TODO - same for above
727 \begingroup
728 \catcode`#1\active
729 \nfss@catcodes

```



```

730     \ifnum\catcode`#1=\active
731     \endgroup
732     \bbl@add\nfss@catcodes{\@makeother#1}%
733     \else
734     \endgroup
735     \fi
736 \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.

```

737 \def\bbl@remove@special#1{%
738   \begingroup
739   \def\x##1##2{\ifnum`#1= `##2\noexpand\@empty
740     \else\noexpand##1\noexpand##2\fi}%
741   \def\do{\x\do}%
742   \def\@makeother{\x\@makeother}%
743   \edef\x{\endgroup
744     \def\noexpand\dospecials{\dospecials}%
745     \expandafter\ifx\csname @sanitize\endcsname\relax\else
746       \def\noexpand\@sanitize{\@sanitize}%
747     \fi}%
748   \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).

```

749 \def\bbl@active@def#1#2#3#4{%
750   \@namedef{#3#1}{%
751     \expandafter\ifx\csname#2@sh@#1\endcsname\relax
752       \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
753     \else
754       \bbl@afterfi\csname#2@sh@#1\endcsname
755     \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.

```

756   \long\@namedef{#3@arg#1}##1{%
757     \expandafter\ifx\csname#2@sh@#1\string##1\endcsname\relax
758       \bbl@afterelse\csname#4#1\endcsname##1%

```

```

759 \else
760 \bbl@afterfi\csname#2@sh@#1@\string##1@\endcsname
761 \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.

```

762 \def\initiate@active@char#1{%
763 \bbl@ifunset{active@char\string#1}%
764 {\bbl@withactive
765 {\expandafter\@initiate@active@char\expandafter}#1\string#1}%
766 {}}

```

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).

```

767 \def\@initiate@active@char#1#2#3{%
768 \bbl@csarg\edef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%
769 \ifx#1\@undefined
770 \bbl@csarg\edef{oridef@#2}{\let\noexpand#1\noexpand\@undefined}%
771 \else
772 \bbl@csarg\let{oridef@#2}#1%
773 \bbl@csarg\edef{oridef@#2}{%
774 \let\noexpand#1%
775 \expandafter\noexpand\csname bbl@oridef@@#2\endcsname}%
776 \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*").

```

777 \ifx#1#3\relax
778 \expandafter\let\csname normal@char#2\endcsname#3%
779 \else
780 \bbl@info{Making #2 an active character}%
781 \ifnum\mathcode`#2="8000
782 \@namedef{normal@char#2}{%
783 \textormath{#3}{\csname bbl@oridef@@#2\endcsname}}%
784 \else
785 \@namedef{normal@char#2}{#3}%
786 \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).

```

787 \bbl@restoreactive{#2}%
788 \AtBeginDocument{%
789 \catcode`#2\active
790 \if@filesw
791 \immediate\write\@mainaux{\catcode`\string#2\active}%
792 \fi}%
793 \expandafter\bbl@add@special\csname#2\endcsname

```

```

794 \catcode`#2\active
795 \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⟩`).

```

796 \let\bbl@tempa\@firstoftwo
797 \if\string`#2%
798   \def\bbl@tempa{\noexpand\textormath}%
799 \else
800   \ifx\bbl@mathnormal\@undefined\else
801     \let\bbl@tempa\bbl@mathnormal
802   \fi
803 \fi
804 \expandafter\edef\csname active@char#2\endcsname{%
805   \bbl@tempa
806   {\noexpand\if@safe@actives
807     \noexpand\expandafter
808     \expandafter\noexpand\csname normal@char#2\endcsname
809     \noexpand\else
810       \noexpand\expandafter
811       \expandafter\noexpand\csname bbl@doactive#2\endcsname
812       \noexpand\fi}%
813   {\expandafter\noexpand\csname normal@char#2\endcsname}}}%
814 \bbl@csarg\edef{doactive#2}{%
815   \expandafter\noexpand\csname user@active#2\endcsname}%

```

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

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

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

```

816 \bbl@csarg\edef{active@#2}{%
817   \noexpand\active@prefix\noexpand#1%
818   \expandafter\noexpand\csname active@char#2\endcsname}%
819 \bbl@csarg\edef{normal@#2}{%
820   \noexpand\active@prefix\noexpand#1%
821   \expandafter\noexpand\csname normal@char#2\endcsname}%
822 \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.

```

823 \bbl@active@def#2\user@group{user@active}{language@active}%
824 \bbl@active@def#2\language@group{language@active}{system@active}%
825 \bbl@active@def#2\system@group{system@active}{normal@char}%

```

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

```

826 \expandafter\edef\csname\user@group @sh#2@@\endcsname
827   {\expandafter\noexpand\csname normal@char#2\endcsname}%
828 \expandafter\edef\csname\user@group @sh#2@\string\protect\endcsname
829   {\expandafter\noexpand\csname user@active#2\endcsname}%

```

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

```
830 \if\string'#2%
831   \let\prim@s\bbl@prim@s
832   \let\active@math@prime#1%
833 \fi
834 \bbl@usehooks{initiateactive}{\#1}{\#2}{\#3}}
```

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

```
835 <<More package options>> ≡
836 \DeclareOption{math=active}{}
837 \DeclareOption{math=normal}{\def\bbl@mathnormal{\noexpand\textormath}}
838 <</More package options>>
```

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

```
839 \@ifpackagewith{babel}{KeepShorthandsActive}%
840   {\let\bbl@restoreactive\@gobble}%
841   {\def\bbl@restoreactive#1{%
842     \bbl@exp{%
843       \\\AfterBabelLanguage\\CurrentOption
844       {\catcode`#1=\the\catcode`#1\relax}%
845       \\\AtEndOfPackage
846       {\catcode`#1=\the\catcode`#1\relax}}}%
847   \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.

```
848 \def\bbl@sh@select#1#2{%
849   \expandafter\ifx\curname#1@sh@#2@sel\endcurname\relax
850   \bbl@afterelse\bbl@scndcs
851   \else
852   \bbl@afterfi\curname#1@sh@#2@sel\endcurname
853   \fi}
```

`\active@prefix` The command `\active@prefix` which is used in the expansion of active characters has a function similar to `\OT1-cmd` in that it `\protects` the active character whenever `\protect` is *not* `\@typeset@protect`.

```
854 \def\active@prefix#1{%
855   \ifx\protect\@typeset@protect
856   \else
```

When `\protect` is set to `\@unexpandable@protect` we make sure that the active character is *not* expanded by inserting `\noexpand` in front of it. The `\@gobble` is needed to remove a token such as `\activechar:` (when the double colon was the active character to be dealt with).

```
857   \ifx\protect\@unexpandable@protect
858     \noexpand#1%
```

```

859 \else
860 \protect#1%
861 \fi
862 \expandafter\@gobble
863 \fi}

\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>.

864 \newif\if@safe@actives
865 \@safe@activesfalse

\bbbl@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.

866 \def\bbbl@restore@actives{\if@safe@actives\@safe@activesfalse\fi}

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

867 \def\bbbl@activate#1{%
868 \bbbl@withactive{\expandafter\let\expandafter}#1%
869 \csname bbl@active@\string#1\endcsname}
870 \def\bbbl@deactivate#1{%
871 \bbbl@withactive{\expandafter\let\expandafter}#1%
872 \csname bbl@normal@\string#1\endcsname}

\bbbl@firstcs \bbbl@scndcs These macros have two arguments. They use one of their arguments to build a control
sequence from.

873 \def\bbbl@firstcs#1#2{\csname#1\endcsname}
874 \def\bbbl@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.

875 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil}
876 \def\@decl@short#1#2#3\@nil#4{%
877 \def\bbbl@tempa{#3}%
878 \ifx\bbbl@tempa\@empty
879 \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbbl@scndcs
880 \bbbl@ifunset{#1@sh@\string#2@}{}%
881 {\def\bbbl@tempa{#4}%
882 \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbbl@tempa
883 \else
884 \bbbl@info
885 {Redefining #1 shorthand \string#2\\%
886 in language \CurrentOption}%
887 \fi}%
888 \@namedef{#1@sh@\string#2@}{#4}%
889 \else
890 \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbbl@firstcs
891 \bbbl@ifunset{#1@sh@\string#2@\string#3@}{}%

```

```

892      {\def\bbl@tempa{#4}%
893       \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
894       \else
895         \bbl@info
896         {Redefining #1 shorthand \string#2\string#3\\%
897          in language \CurrentOption}%
898       \fi}%
899       \@namedef{#1@sh@\string#2@\string#3@}{#4}%
900     \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.

```

901 \def\textormath{%
902   \ifmmode
903     \expandafter\@secondoftwo
904   \else
905     \expandafter\@firstoftwo
906   \fi}

```

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

```

907 \def\user@group{user}
908 \def\language@group{english}
909 \def\system@group{system}

```

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

```

910 \def\useshorthands{%
911   \@ifstar\bbl@usesh@s{\bbl@usesh@x{}}
912 \def\bbl@usesh@s#1{%
913   \bbl@usesh@x
914   {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bbl@activate{#1}}}%
915   {#1}}
916 \def\bbl@usesh@x#1#2{%
917   \bbl@ifshorthand{#2}%
918   {\def\user@group{user}%
919    \initiate@active@char{#2}%
920    #1%
921    \bbl@activate{#2}}%
922   {\bbl@error
923    {Cannot declare a shorthand turned off (\string#2)}
924    {Sorry, but you cannot use shorthands which have been\\%
925     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.

```

926 \def\user@language@group{user@\language@group}
927 \def\bbl@set@user@generic#1#2{%
928   \bbl@ifunset{user@generic@active#1}%

```

```

929 {\bbl@active@def#1\user@language@group{user@active}{user@generic@active}%
930 \bbl@active@def#1\user@group{user@generic@active}{language@active}%
931 \expandafter\edef\csname#2@sh@#1@@\endcsname{%
932 \expandafter\noexpand\csname normal@char#1\endcsname}%
933 \expandafter\edef\csname#2@sh@#1@\string\protect@\endcsname{%
934 \expandafter\noexpand\csname user@active#1\endcsname}}%
935 \@empty}
936 \newcommand\defineshorthand[3][user]{%
937 \edef\bbl@tempa{\zap@space#1 \@empty}%
938 \bbl@for\bbl@tempb\bbl@tempa{%
939 \if*\expandafter\@car\bbl@tempb\@nil
940 \edef\bbl@tempb{user@\expandafter\@gobble\bbl@tempb}%
941 \@expandtwoargs
942 \bbl@set@user@generic{\expandafter\string\@car#2\@nil}\bbl@tempb
943 \fi
944 \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.

```

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

```

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

```

946 \def\aliasshorthand#1#2{%
947 \bbl@ifshorthand{#2}%
948 {\expandafter\ifx\csname active@char\string#2\endcsname\relax
949 \ifx\document\@notprerr
950 \@notshorthand{#2}%
951 \else
952 \initiate@active@char{#2}%

```

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

```

953 \expandafter\let\csname active@char\string#2\expandafter\endcsname
954 \csname active@char\string#1\endcsname
955 \expandafter\let\csname normal@char\string#2\expandafter\endcsname
956 \csname normal@char\string#1\endcsname
957 \bbl@activate{#2}%
958 \fi
959 \fi}%
960 {\bbl@error
961 {Cannot declare a shorthand turned off (\string#2)}
962 {Sorry, but you cannot use shorthands which have been\\%
963 turned off in the package options}}}

```

`\@notshorthand`

```

964 \def\@notshorthand#1{%
965 \bbl@error{%
966 The character '\string #1' should be made a shorthand character;\\%
967 add the command \string\usesshorthands\string{#1\string} to
968 the preamble.\\%
969 I will ignore your instruction}%
970 {You may proceed, but expect unexpected results}}

```

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

```

971 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
972 \DeclareRobustCommand*\shorthandoff{%
973   \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
974 \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.

```

975 \def\bbl@switch@sh#1#2{%
976   \ifx#2\@nnil\else
977     \bbl@ifunset{\bbl@active@\string#2}%
978     {\bbl@error
979       {I cannot switch '\string#2' on or off--not a shorthand}%
980       {This character is not a shorthand. Maybe you made\\%
981         a typing mistake? I will ignore your instruction}}}%
982     {\ifcase#1%
983       \catcode`#212\relax
984       \or
985       \catcode`#2\active
986       \or
987       \csname bbl@oricat@\string#2\endcsname
988       \csname bbl@oridef@\string#2\endcsname
989       \fi}%
990     \bbl@afterfi\bbl@switch@sh#1%
991   \fi}

```

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

```

992 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
993 \def\bbl@putsh#1{%
994   \bbl@ifunset{\bbl@active@\string#1}%
995   {\bbl@putsh@i#1\@empty\@nnil}%
996   {\csname bbl@active@\string#1\endcsname}}
997 \def\bbl@putsh@i#1#2\@nnil{%
998   \csname\languagename @sh@\string#1@%
999   \ifx\@empty#2\else\string#2@fi\endcsname}
1000 \ifx\bbl@opt@shorthands\@nnil\else
1001   \let\bbl@s@initiate@active@char\initiate@active@char
1002   \def\initiate@active@char#1{%
1003     \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
1004   \let\bbl@s@switch@sh\bbl@switch@sh
1005   \def\bbl@switch@sh#1#2{%
1006     \ifx#2\@nnil\else
1007       \bbl@afterfi
1008       \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
1009       \fi}
1010   \let\bbl@s@activate\bbl@activate
1011   \def\bbl@activate#1{%
1012     \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
1013   \let\bbl@s@deactivate\bbl@deactivate
1014   \def\bbl@deactivate#1{%

```



```

1015 \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
1016 \fi

```

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

```

1017 \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.

```

1018 \def\bbl@prim@s{%
1019 \prime\futurelet\@let@token\bbl@pr@m@s}
1020 \def\bbl@if@primes#1#2{%
1021 \ifx#1\@let@token
1022 \expandafter\@firstoftwo
1023 \else\ifx#2\@let@token
1024 \bbl@afterelse\expandafter\@firstoftwo
1025 \else
1026 \bbl@afterfi\expandafter\@secondoftwo
1027 \fi\fi}
1028 \begingroup
1029 \catcode`\^=7 \catcode`\*= \active \lccode`\*=`\^
1030 \catcode`\'=12 \catcode`\`= \active \lccode`\`=``
1031 \lowercase{%
1032 \gdef\bbl@pr@m@s{%
1033 \bbl@if@primes" "%
1034 \pr@@@s
1035 {\bbl@if@primes*^ \pr@@@t\egroup}}}
1036 \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).

```

1037 \initiate@active@char{~}
1038 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
1039 \bbl@activate{~}

```

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

```

1040 \expandafter\def\csname OT1dqpos\endcsname{127}
1041 \expandafter\def\csname T1dqpos\endcsname{4}

```

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

```

1042 \ifx\f@encoding\@undefined
1043 \def\f@encoding{OT1}
1044 \fi

```

## 9.5 Language attributes

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

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

```

1045 \bbl@trace{Language attributes}
1046 \newcommand\languageattribute[2]{%
1047   \def\bbl@tempc{#1}%
1048   \bbl@fixname\bbl@tempc
1049   \bbl@iflanguage\bbl@tempc{%
1050     \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.

```

1051     \ifx\bbl@known@attribs\@undefined
1052       \in@false
1053     \else

```

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

```

1054       \bbl@xin@{,\bbl@tempc-##1,},{,\bbl@known@attribs,}%
1055     \fi

```

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

```

1056     \ifin@
1057       \bbl@warning{%
1058         You have more than once selected the attribute '##1'\%
1059         for language #1. Reported}%
1060     \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.

```

1061       \bbl@exp{%
1062         \\bbl@add@list\\bbl@known@attribs{\bbl@tempc-##1}}%
1063       \edef\bbl@tempa{\bbl@tempc-##1}%
1064       \expandafter\bbl@ifknown@trib\expandafter{\bbl@tempa}\bbl@attributes%
1065       {\csname\bbl@tempc @attr##1\endcsname}%
1066       {\@attrerr{\bbl@tempc}{##1}}%
1067     \fi}}

```

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

```

1068 \@onlypreamble\languageattribute

```

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

```

1069 \newcommand*{\@attrerr}[2]{%
1070   \bbl@error
1071   {The attribute #2 is unknown for language #1.}%
1072   {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}`.

```

1073 \def\bbl@declare@ttribute#1#2#3{%
1074   \bbl@xin@{,#2,},{,\BabelModifiers,}%
1075   \ifin@
1076     \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%

```

```

1077 \fi
1078 \bbl@add@list\bbl@attributes{#1-#2}%
1079 \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.

```

1080 \def\bbl@ifattributeset#1#2#3#4{%

```

First we need to find out if any attributes were set; if not we're done.

```

1081 \ifx\bbl@known@attribs\@undefined
1082 \in@false
1083 \else

```

The we need to check the list of known attributes.

```

1084 \bbl@xin@{, #1-#2, }{, \bbl@known@attribs,}%
1085 \fi

```

When we're this far `\ifin@` has a value indicating if the attribute in question was set or not. Just to be safe the code to be executed is 'thrown over the `\fi`'.

```

1086 \ifin@
1087 \bbl@afterelse#3%
1088 \else
1089 \bbl@afterfi#4%
1090 \fi
1091 }

```

`\bbl@ifknown@ttrib` An internal macro to check whether a given language/attribute is known. The macro takes 4 arguments, the language/attribute, the attribute list, the  $\TeX$ -code to be executed when the attribute is known and the  $\TeX$ -code to be executed otherwise.

```

1092 \def\bbl@ifknown@ttrib#1#2{%

```

We first assume the attribute is unknown.

```

1093 \let\bbl@tempa\@secondoftwo

```

Then we loop over the list of known attributes, trying to find a match.

```

1094 \bbl@loopx\bbl@tempb{#2}%
1095 \expandafter\in@\expandafter{\expandafter,\bbl@tempb,}{, #1,}%
1096 \ifin@

```

When a match is found the definition of `\bbl@tempa` is changed.

```

1097 \let\bbl@tempa\@firstoftwo
1098 \else
1099 \fi}%

```

Finally we execute `\bbl@tempa`.

```

1100 \bbl@tempa
1101 }

```

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

```

1102 \def\bbl@clear@ttribs{%
1103 \ifx\bbl@attributes\@undefined\else
1104 \bbl@loopx\bbl@tempa{\bbl@attributes}{%

```

```

1105     \expandafter\bb1@clear@ttrib\bb1@tempa.
1106     }%
1107     \let\bb1@attributes\@undefined
1108     \fi}
1109 \def\bb1@clear@ttrib#1-#2.{%
1110     \expandafter\let\csname#1@attr@#2\endcsname\@undefined}
1111 \AtBeginDocument{\bb1@clear@ttribs}

```

## 9.6 Support for saving macro definitions

To save the meaning of control sequences using `\babel@save`, we use temporary control sequences. To save hash table entries for these control sequences, we don't use the name of the control sequence to be saved to construct the temporary name. Instead we simply use the value of a counter, which is reset to zero each time we begin to save new values. This works well because we release the saved meanings before we begin to save a new set of control sequence meanings (see `\selectlanguage` and `\originalTeX`). Note undefined macros are not undefined any more when saved – they are `\relax`'ed.

`\babel@savecnt` The initialization of a new save cycle: reset the counter to zero.

`\babel@beginsave` 1112 `\bb1@trace{Macros for saving definitions}`  
 1113 `\def\babel@beginsave{\babel@savecnt\z@}`

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

```

1114 \newcount\babel@savecnt
1115 \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.

```

1116 \def\babel@save#1{%
1117     \expandafter\let\csname babel@\number\babel@savecnt\endcsname#1\relax
1118     \toks@\expandafter{\originalTeX\let#1=}%
1119     \bb1@exp{%
1120         \def\\originalTeX{\the\toks@<\babel@\number\babel@savecnt>\relax}}%
1121     \advance\babel@savecnt\@ne}

```

`\babel@savevariable` The macro `\babel@savevariable<variable>` saves the value of the variable. `<variable>` can be anything allowed after the `\the` primitive.

```

1122 \def\babel@savevariable#1{%
1123     \toks@\expandafter{\originalTeX #1=}%
1124     \bb1@exp{\def\\originalTeX{\the\toks@\the#1\relax}}}

```

`\bb1@frenchspacing` Some languages need to have `\frenchspacing` in effect. Others don't want that. The  
`\bb1@nonfrenchspacing` command `\bb1@frenchspacing` switches it on when it isn't already in effect and  
`\bb1@nonfrenchspacing` switches it off if necessary.

```

1125 \def\bb1@frenchspacing{%
1126     \ifnum\the\sfcodes\@m
1127     \let\bb1@nonfrenchspacing\relax
1128     \else
1129     \frenchspacing
1130     \let\bb1@nonfrenchspacing\nonfrenchspacing
1131     \fi}
1132 \let\bb1@nonfrenchspacing\nonfrenchspacing

```

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

## 9.7 Short tags

`\babeltags` This macro is straightforward. After zapping spaces, we loop over the list and define the macros `\text{<tag>}` and `\<tag>`. Definitions are first expanded so that they don't contain `\csname` but the actual macro.

```
1133 \bbl@trace{Short tags}
1134 \def\babeltags#1{%
1135   \edef\bbl@tempa{\zap@space#1 \@empty}%
1136   \def\bbl@tempb##1=##2\@{#1}%
1137   \edef\bbl@tempc{%
1138     \noexpand\newcommand
1139     \expandafter\noexpand\csname ##1\endcsname{%
1140       \noexpand\protect
1141       \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
1142     \noexpand\newcommand
1143     \expandafter\noexpand\csname text##1\endcsname{%
1144       \noexpand\foreignlanguage{##2}}
1145     \bbl@tempc}%
1146   \bbl@for\bbl@tempa\bbl@tempa{%
1147     \expandafter\bbl@tempb\bbl@tempa\@{#1}}
```

## 9.8 Hyphens

`\babelhyphenation` This macro saves hyphenation exceptions. Two macros are used to store them: `\bbl@hyphenation@` for the global ones and `\bbl@hyphenation<lang>` for language ones. See `\bbl@patterns` above for further details. We make sure there is a space between words when multiple commands are used.

```
1148 \bbl@trace{Hyphens}
1149 \@onlypreamble\babelhyphenation
1150 \AtEndOfPackage{%
1151   \newcommand\babelhyphenation[2][\@empty]{%
1152     \ifx\bbl@hyphenation@ relax
1153     \let\bbl@hyphenation@\@empty
1154     \fi
1155     \ifx\bbl@hyphlist\@empty\else
1156       \bbl@warning{%
1157         You must not intermingle \string\selectlanguage\space and\%
1158         \string\babelhyphenation\space or some exceptions will not\%
1159         be taken into account. Reported}%
1160       \fi
1161       \ifx\@empty#1%
1162         \protected@edef\bbl@hyphenation@{\bbl@hyphenation@\space#2}%
1163       \else
1164         \bbl@vforeach{#1}{%
1165           \def\bbl@tempa{##1}%
1166           \bbl@fixname\bbl@tempa
1167           \bbl@iflanguage\bbl@tempa{%
1168             \bbl@csarg\protected@edef{hyphenation@\bbl@tempa}{%
1169               \bbl@ifunset{bbl@hyphenation@\bbl@tempa}%
1170               \@empty
1171               {\csname bbl@hyphenation@\bbl@tempa\endcsname\space}%
1172               #2}}}%
1173         \fi}}
```

`\bbl@allowhyphens` This macro makes hyphenation possible. Basically its definition is nothing more than `\nobreak \hskip 0pt plus 0pt33`.

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

```

1174 \def\bbl@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}
1175 \def\bbl@t@one{T1}
1176 \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.

```

1177 \newcommand\babellnullhyphen{\char\hyphenchar\font}
1178 \def\babelhyphen{\active@prefix\babelhyphen\bbl@hyphen}
1179 \def\bbl@hyphen{%
1180   \@ifstar{\bbl@hyphen@i @}{\bbl@hyphen@i \@empty}}
1181 \def\bbl@hyphen@i#1#2{%
1182   \bbl@ifunset{\bbl@hy@#1#2\@empty}%
1183   {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%
1184   {\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 hyphen 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 discretionaty 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.

```

1185 \def\bbl@usehyphen#1{%
1186   \leavevmode
1187   \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
1188   \nobreak\hskip\z@skip}
1189 \def\bbl@usehyphen#1{%
1190   \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

The following macro inserts the hyphen char.

```

1191 \def\bbl@hyphenchar{%
1192   \ifnum\hyphenchar\font=\m@ne
1193     \babellnullhyphen
1194   \else
1195     \char\hyphenchar\font
1196   \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.

```

1197 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1198 \def\bbl@hy@@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1199 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
1200 \def\bbl@hy@@hard{\bbl@usehyphen\bbl@hyphenchar}
1201 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}
1202 \def\bbl@hy@@nobreak{\mbox{\bbl@hyphenchar}}
1203 \def\bbl@hy@repeat{%
1204   \bbl@usehyphen{%
1205     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}
1206 \def\bbl@hy@@repeat{%
1207   \bbl@usehyphen{%
1208     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}
1209 \def\bbl@hy@empty{\hskip\z@skip}
1210 \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.

```

1211 \def\bbl@disc#1#2{\nobreak\discretionary{#2-}{}{#1}\bbl@allowhyphens}

```

## 9.9 Multiencoding strings

The aim following commands is to provide a common interface for strings in several encodings. They also contains several hooks which can be used by luatex and xetex. The code is organized here with pseudo-guards, so we start with the basic commands.

**Tools** But first, a couple of tools. The first one makes global a local variable. This is not the best solution, but it works.

```
1212 \bbl@trace{Multiencoding strings}
1213 \def\bbl@tglobal#1{\global\let#1#1}
1214 \def\bbl@recatcode#1{%
1215   \@tempcnta="7F
1216   \def\bbl@tempa{%
1217     \ifnum\@tempcnta>"FF\else
1218       \catcode\@tempcnta=#1\relax
1219       \advance\@tempcnta\@ne
1220       \expandafter\bbl@tempa
1221     \fi}%
1222   \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).

```
1223 \@ifpackagewith{babel}{nocase}%
1224   {\let\bbl@patchuclc\relax}%
1225   {\def\bbl@patchuclc{%
1226     \global\let\bbl@patchuclc\relax
1227     \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}%
1228     \gdef\bbl@uclc##1{%
1229       \let\bbl@encoded\bbl@encoded@uclc
1230       \bbl@ifunset{\language @bbl@uclc}% and resumes it
1231       {##1}%
1232       {\let\bbl@tempa##1\relax % Used by LANG@bbl@uclc
1233         \csname\language @bbl@uclc\endcsname}%
1234       {\bbl@tolower\@empty}{\bbl@toupper\@empty}}%
1235     \gdef\bbl@tolower{\csname\language @bbl@lc\endcsname}%
1236     \gdef\bbl@toupper{\csname\language @bbl@uc\endcsname}}%
1237 <<More package options>> ≡
1238 \DeclareOption{nocase}{}
1239 <</More package options>>
```

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

```
1240 <<More package options>> ≡
1241 \let\bbl@opt@strings\@nnil % accept strings=value
1242 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1243 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1244 \def\BabelStringsDefault{generic}
1245 <</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.

```

1246 \@onlypreamble\StartBabelCommands
1247 \def\StartBabelCommands{%
1248   \begingroup
1249   \bbl@recatcode{11}%
1250   <⟨Macros local to BabelCommands⟩
1251   \def\bbl@provstring##1##2{%
1252     \providecommand##1{##2}%
1253     \bbl@toglobal##1}%
1254   \global\let\bbl@scafter\@empty
1255   \let\StartBabelCommands\bbl@startcmds
1256   \ifx\BabelLanguages\relax
1257     \let\BabelLanguages\CurrentOption
1258   \fi
1259   \begingroup
1260   \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
1261   \StartBabelCommands}
1262 \def\bbl@startcmds{%
1263   \ifx\bbl@screset\@nnil\else
1264     \bbl@usehooks{stopcommands}{}%
1265   \fi
1266   \endgroup
1267   \begingroup
1268   \@ifstar
1269     {\ifx\bbl@opt@strings\@nnil
1270       \let\bbl@opt@strings\BabelStringsDefault
1271     \fi
1272     \bbl@startcmds@i}%
1273   \bbl@startcmds@i}
1274 \def\bbl@startcmds@i#1#2{%
1275   \edef\bbl@L{\zap@space#1 \@empty}%
1276   \edef\bbl@G{\zap@space#2 \@empty}%
1277   \bbl@startcmds@ii}

```

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

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

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

```

1278 \newcommand\bbl@startcmds@ii[1][\@empty]{%
1279   \let\SetString\@gobbletwo
1280   \let\bbl@stringdef\@gobbletwo
1281   \let\AfterBabelCommands\@gobble
1282   \ifx\@empty#1%
1283     \def\bbl@sc@label{generic}%
1284     \def\bbl@encstring##1##2{%
1285       \ProvideTextCommandDefault##1{##2}%
1286       \bbl@toglobal##1%
1287       \expandafter\bbl@toglobal\csname\string?\string##1\endcsname}%
1288     \let\bbl@sctest\in@true
1289   \else

```



```

1290 \let\bbl@sc@charset\space % <- zapped below
1291 \let\bbl@sc@fontenc\space % <- " "
1292 \def\bbl@tempa##1=##2\@nil{%
1293 \bbl@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
1294 \bbl@vforeach{label=#1}{\bbl@tempa##1\@nil}%
1295 \def\bbl@tempa##1 ##2{% space -> comma
1296 ##1%
1297 \ifx\@empty##2\else\ifx,##1,\else,\fi\bbl@afterfi\bbl@tempa##2\fi}%
1298 \edef\bbl@sc@fontenc{\expandafter\bbl@tempa\bbl@sc@fontenc\@empty}%
1299 \edef\bbl@sc@label{\expandafter\zap@space\bbl@sc@label\@empty}%
1300 \edef\bbl@sc@charset{\expandafter\zap@space\bbl@sc@charset\@empty}%
1301 \def\bbl@encstring##1##2{%
1302 \bbl@foreach\bbl@sc@fontenc{%
1303 \bbl@ifunset{T@###1}%
1304 {}%
1305 {\ProvideTextCommand##1{####1}{##2}%
1306 \bbl@tglobal##1%
1307 \expandafter
1308 \bbl@tglobal\csname####1\string##1\endcsname}}}%
1309 \def\bbl@sctest{%
1310 \bbl@xin@{\, \bbl@opt@strings,}{, \bbl@sc@label, \bbl@sc@fontenc,}%
1311 \fi
1312 \ifx\bbl@opt@strings\@nnil % ie, no strings key -> defaults
1313 \else\ifx\bbl@opt@strings\relax % ie, strings=encoded
1314 \let\AfterBabelCommands\bbl@aftercmds
1315 \let\SetString\bbl@setstring
1316 \let\bbl@stringdef\bbl@encstring
1317 \else % ie, strings=value
1318 \bbl@sctest
1319 \ifin@
1320 \let\AfterBabelCommands\bbl@aftercmds
1321 \let\SetString\bbl@setstring
1322 \let\bbl@stringdef\bbl@provstring
1323 \fi\fi\fi
1324 \bbl@scswitch
1325 \ifx\bbl@G\@empty
1326 \def\SetString##1##2{%
1327 \bbl@error{Missing group for string \string##1}%
1328 {You must assign strings to some category, typically\\
1329 captions or extras, but you set none}}%
1330 \fi
1331 \ifx\@empty#1%
1332 \bbl@usehooks{defaultcommands}{}%
1333 \else
1334 \@expandtwoargs
1335 \bbl@usehooks{encodedcommands}{\bbl@sc@charset}{\bbl@sc@fontenc}}%
1336 \fi}

```

There are two versions of `\bbl@scswitch`. The first version is used when `ldfs` are read, and it makes sure `\group\language` 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\language` 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).

```

1337 \def\bbl@forlang#1#2{%
1338 \bbl@for#1\bbl@L{%

```

```

1339 \bbl@xin@{, #1, }{, \BabelLanguages,}%
1340 \ifin@#2\relax\fi}}
1341 \def\bbl@scswitch{%
1342 \bbl@forlang\bbl@tempa{%
1343 \ifx\bbl@G\empty\else
1344 \ifx\SetString\@gobbletwo\else
1345 \edef\bbl@GL{\bbl@G\bbl@tempa}%
1346 \bbl@xin@{, \bbl@GL, }{, \bbl@screset,}%
1347 \ifin@\else
1348 \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
1349 \xdef\bbl@screset{\bbl@screset, \bbl@GL}%
1350 \fi
1351 \fi
1352 \fi}}
1353 \AtEndOfPackage{%
1354 \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{}{#2}}}%
1355 \let\bbl@scswitch\relax}
1356 \@onlypreamble\EndBabelCommands
1357 \def\EndBabelCommands{%
1358 \bbl@usehooks{stopcommands}{}}%
1359 \endgroup
1360 \endgroup
1361 \bbl@scafter}

```

Now we define commands to be used inside \StartBabelCommands.

**Strings** The following macro is the actual definition of \SetString when it is “active”. First save the “switcher”. Create it if undefined. Strings are defined only if undefined (ie, like \providescommand). With the event stringprocess you can preprocess the string by manipulating the value of \BabelString. If there are several hooks assigned to this event, preprocessing is done in the same order as defined. Finally, the string is set.

```

1362 \def\bbl@setstring#1#2{%
1363 \bbl@forlang\bbl@tempa{%
1364 \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
1365 \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
1366 {\global\expandafter % TODO - con \bbl@exp ?
1367 \bbl@add\csname\bbl@G\bbl@tempa\expandafter\endcsname\expandafter
1368 {\expandafter\bbl@scset\expandafter#1\csname\bbl@LC\endcsname}}}%
1369 }%
1370 \def\BabelString{#2}%
1371 \bbl@usehooks{stringprocess}{}}%
1372 \expandafter\bbl@stringdef
1373 \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.

```

1374 \ifx\bbl@opt@strings\relax
1375 \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
1376 \bbl@patchuclc
1377 \let\bbl@encoded\relax
1378 \def\bbl@encoded@uclc#1{%
1379 \@inmathwarn#1%
1380 \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
1381 \expandafter\ifx\csname ?\string#1\endcsname\relax
1382 \TextSymbolUnavailable#1%
1383 \else

```

```

1384     \csname ?\string#1\endcsname
1385     \fi
1386   \else
1387     \csname\cf@encoding\string#1\endcsname
1388     \fi}
1389 \else
1390   \def\bbl@scset#1#2{\def#1{#2}}
1391 \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.

```

1392 <<*Macros local to BabelCommands>> ≡
1393 \def\SetStringLoop##1##2{%
1394   \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
1395   \count@\z@
1396   \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
1397     \advance\count@\@ne
1398     \toks@\expandafter{\bbl@tempa}%
1399     \bbl@exp{%
1400       \\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
1401       \count@=\the\count@\relax}}}%
1402 <</Macros local to BabelCommands>>

```

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

```

1403 \def\bbl@aftercmds#1{%
1404   \toks@\expandafter{\bbl@scafter#1}%
1405   \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.

```

1406 <<*Macros local to BabelCommands>> ≡
1407 \newcommand\SetCase[3][{}%
1408   \bbl@patchuclc
1409   \bbl@forlang\bbl@tempa{%
1410     \expandafter\bbl@encstring
1411     \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
1412     \expandafter\bbl@encstring
1413     \csname\bbl@tempa @bbl@uc\endcsname{##2}%
1414     \expandafter\bbl@encstring
1415     \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
1416 <</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.

```

1417 <<*Macros local to BabelCommands>> ≡
1418 \newcommand\SetHyphenMap[1]{%
1419   \bbl@forlang\bbl@tempa{%
1420     \expandafter\bbl@stringdef
1421     \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
1422 <</Macros local to BabelCommands>>

```

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

```

1423 \newcommand\BabelLower[2]{% one to one.
1424   \ifnum\lccode#1=#2\else
1425     \babel@savevariable{\lccode#1}%
1426     \lccode#1=#2\relax
1427   \fi}
1428 \newcommand\BabelLowerMM[4]{% many-to-many
1429   \@tempcnta=#1\relax
1430   \@tempcntb=#4\relax
1431   \def\bbl@tempa{%
1432     \ifnum\@tempcnta>#2\else
1433       \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
1434       \advance\@tempcnta#3\relax
1435       \advance\@tempcntb#3\relax
1436       \expandafter\bbl@tempa
1437     \fi}%
1438   \bbl@tempa}
1439 \newcommand\BabelLowerMO[4]{% many-to-one
1440   \@tempcnta=#1\relax
1441   \def\bbl@tempa{%
1442     \ifnum\@tempcnta>#2\else
1443       \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
1444       \advance\@tempcnta#3
1445       \expandafter\bbl@tempa
1446     \fi}%
1447   \bbl@tempa}

```

The following package options control the behavior of hyphenation mapping.

```

1448 << *More package options >> ≡
1449 \DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
1450 \DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\@ne}
1451 \DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
1452 \DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@@}
1453 \DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}
1454 << /More package options >>

```

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

```

1455 \AtEndOfPackage{%
1456   \ifx\bbl@opt@hyphenmap\undefined
1457     \bbl@xin@{,}{\bbl@language@opts}%
1458     \chardef\bbl@opt@hyphenmap\ifin4\else\@ne\fi
1459   \fi}

```

## 9.10 Macros common to a number of languages

`\set@low@box` The following macro is used to lower quotes to the same level as the comma. It prepares its argument in box register 0.

```

1460 \bbl@trace{Macros related to glyphs}
1461 \def\set@low@box#1{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}%
1462   \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@%
1463   \setbox\z@\hbox{\lower\dimen\z@ \box\z@}\ht\z@\ht\tw@ \dp\z@\dp\tw@}

```

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

```

1464 \def\save@sf@q#1{\leavevmode
1465   \begingroup
1466   \edef\@SF{\spacefactor\the\spacefactor}#1\@SF
1467   \endgroup}

```

## 9.11 Making glyphs available

This section makes a number of glyphs available that either do not exist in the OT1 encoding and have to be ‘faked’, or that are not accessible through T1enc.def.

### 9.11.1 Quotation marks

`\quotedblbase` In the T1 encoding the opening double quote at the baseline is available as a separate character, accessible via `\quotedblbase`. In the OT1 encoding it is not available, therefore we make it available by lowering the normal open quote character to the baseline.

```
1468 \ProvideTextCommand{\quotedblbase}{OT1}{%
1469   \save@sf@q{\set@low@box{\textquotedblright\}}%
1470   \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.

```
1471 \ProvideTextCommandDefault{\quotedblbase}{%
1472   \UseTextSymbol{OT1}{\quotedblbase}}
```

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

```
1473 \ProvideTextCommand{\quotesinglbase}{OT1}{%
1474   \save@sf@q{\set@low@box{\textquoteright\}}%
1475   \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.

```
1476 \ProvideTextCommandDefault{\quotesinglbase}{%
1477   \UseTextSymbol{OT1}{\quotesinglbase}}
```

`\guillemotleft` The guillemet characters are not available in OT1 encoding. They are faked.  
`\guillemotright`

```
1478 \ProvideTextCommand{\guillemotleft}{OT1}{%
1479   \ifmmode
1480     \ll
1481   \else
1482     \save@sf@q{\nobreak
1483       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bb1@allowhyphens}%
1484   \fi}
1485 \ProvideTextCommand{\guillemotright}{OT1}{%
1486   \ifmmode
1487     \gg
1488   \else
1489     \save@sf@q{\nobreak
1490       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bb1@allowhyphens}%
1491   \fi}
```

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

```
1492 \ProvideTextCommandDefault{\guillemotleft}{%
1493   \UseTextSymbol{OT1}{\guillemotleft}}
1494 \ProvideTextCommandDefault{\guillemotright}{%
1495   \UseTextSymbol{OT1}{\guillemotright}}
```

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

```
1496 \ProvideTextCommand{\guilsinglleft}{OT1}{%
1497   \ifmmode
1498     <%
1499   \else
```

```

1500 \save@sf@q{\nobreak
1501 \raise.2ex\hbox{$\scriptscriptstyle<$}\bbl@allowhyphens}%
1502 \fi}
1503 \ProvideTextCommand{\guilsinglright}{OT1}{%
1504 \ifmode
1505 >%
1506 \else
1507 \save@sf@q{\nobreak
1508 \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
1509 \fi}

```

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

```

1510 \ProvideTextCommandDefault{\guilsinglleft}{%
1511 \UseTextSymbol{OT1}{\guilsinglleft}}
1512 \ProvideTextCommandDefault{\guilsinglright}{%
1513 \UseTextSymbol{OT1}{\guilsinglright}}

```

### 9.11.2 Letters

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

```

1514 \DeclareTextCommand{\ij}{OT1}{%
1515 i\kern-0.02em\bbl@allowhyphens j}
1516 \DeclareTextCommand{\IJ}{OT1}{%
1517 I\kern-0.02em\bbl@allowhyphens J}
1518 \DeclareTextCommand{\ij}{T1}{\char188}
1519 \DeclareTextCommand{\IJ}{T1}{\char156}

```

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

```

1520 \ProvideTextCommandDefault{\ij}{%
1521 \UseTextSymbol{OT1}{\ij}}
1522 \ProvideTextCommandDefault{\IJ}{%
1523 \UseTextSymbol{OT1}{\IJ}}

```

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

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

```

1524 \def\crrtic@{\hrule height0.1ex width0.3em}
1525 \def\crrtic@{\hrule height0.1ex width0.33em}
1526 \def\ddj@{%
1527 \setbox0\hbox{d}\dimen@=\ht0
1528 \advance\dimen@1ex
1529 \dimen@.45\dimen@
1530 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1531 \advance\dimen@ii.5ex
1532 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
1533 \def\DDJ@{%
1534 \setbox0\hbox{D}\dimen@=.55\ht0
1535 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1536 \advance\dimen@ii.15ex % correction for the dash position
1537 \advance\dimen@ii-.15\fontdimen7\font % correction for cmmt font
1538 \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
1539 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
1540 %

```

```

1541 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
1542 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}

```

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

```

1543 \ProvideTextCommandDefault{\dj}{%
1544   \UseTextSymbol{OT1}{\dj}}
1545 \ProvideTextCommandDefault{\DJ}{%
1546   \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.

```

1547 \DeclareTextCommand{\SS}{OT1}{\SS}
1548 \ProvideTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}

```

### 9.11.3 Shorthands for quotation marks

Shorthands are provided for a number of different quotation marks, which make them usable both outside and inside mathmode. They are defined with `\ProvideTextCommandDefault`, but this is very likely not required because their definitions are based on encoding dependent macros.

`\glq` The ‘german’ single quotes.

```

\grq 1549 \ProvideTextCommandDefault{\glq}{%
1550   \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}

```

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

```

1551 \ProvideTextCommand{\grq}{T1}{%
1552   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
1553 \ProvideTextCommand{\grq}{TU}{%
1554   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
1555 \ProvideTextCommand{\grq}{OT1}{%
1556   \save@sf@q{\kern-.0125em
1557     \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
1558     \kern.07em\relax}}
1559 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}{\grq}}

```

`\glqq` The ‘german’ double quotes.

```

\grqq 1560 \ProvideTextCommandDefault{\glqq}{%
1561   \textormath{\quotedblbase}{\mbox{\quotedblbase}}}

```

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

```

1562 \ProvideTextCommand{\grqq}{T1}{%
1563   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1564 \ProvideTextCommand{\grqq}{TU}{%
1565   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1566 \ProvideTextCommand{\grqq}{OT1}{%
1567   \save@sf@q{\kern-.07em
1568     \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
1569     \kern.07em\relax}}
1570 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}{\grqq}}

```

`\flq` The ‘french’ single guillemets.

```

\frq 1571 \ProvideTextCommandDefault{\flq}{%
1572   \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
1573 \ProvideTextCommandDefault{\frq}{%
1574   \textormath{\guilsinglright}{\mbox{\guilsinglright}}}

```

`\flqq` The ‘french’ double guillemets.

```
\frqq 1575 \ProvideTextCommandDefault{\flqq}{%
1576   \textormath{\guillemotleft}{\mbox{\guillemotleft}}}
1577 \ProvideTextCommandDefault{\frqq}{%
1578   \textormath{\guillemotright}{\mbox{\guillemotright}}}
```

#### 9.11.4 Umlauts and tremas

The command `\` needs to have a different effect for different languages. For German for instance, the ‘umlaut’ should be positioned lower than the default position for placing it over the letters a, o, u, A, O and U. When placed over an e, i, E or I it can retain its normal position. For Dutch the same glyph is always placed in the lower position.

`\umlauthigh` To be able to provide both positions of `\` we provide two commands to switch the  
`\umlautlow` positioning, the default will be `\umlauthigh` (the normal positioning).

```
1579 \def\umlauthigh{%
1580   \def\bbl@umlauta##1{\leavevmode\bgroup%
1581     \expandafter\accent\csname\fontencoding dqpos\endcsname
1582     ##1\bbl@allowhyphens\egroup}%
1583   \let\bbl@umlaute\bbl@umlauta}
1584 \def\umlautlow{%
1585   \def\bbl@umlauta{\protect\lower@umlaut}}
1586 \def\umlautelower{%
1587   \def\bbl@umlaute{\protect\lower@umlaut}}
1588 \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.

```
1589 \expandafter\ifx\csname U@D\endcsname\relax
1590   \csname newdimen\endcsname\U@D
1591 \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.

```
1592 \def\lower@umlaut#1{%
1593   \leavevmode\bgroup
1594     \U@D 1ex%
1595     {\setbox\z@\hbox{%
1596       \expandafter\char\csname\fontencoding dqpos\endcsname}%
1597       \dimen@ -.45ex\advance\dimen@\ht\z@
1598       \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
1599     \expandafter\accent\csname\fontencoding dqpos\endcsname
1600     \fontdimen5\font\U@D #1%
1601   \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).

```

1602 \AtBeginDocument{%
1603   \DeclareTextCompositeCommand{"}{OT1}{a}{\bbl@umlauta{a}}%
1604   \DeclareTextCompositeCommand{"}{OT1}{e}{\bbl@umlaute{e}}%
1605   \DeclareTextCompositeCommand{"}{OT1}{i}{\bbl@umlaute{i}}%
1606   \DeclareTextCompositeCommand{"}{OT1}{\i}{\bbl@umlaute{i}}%
1607   \DeclareTextCompositeCommand{"}{OT1}{o}{\bbl@umlauta{o}}%
1608   \DeclareTextCompositeCommand{"}{OT1}{u}{\bbl@umlauta{u}}%
1609   \DeclareTextCompositeCommand{"}{OT1}{A}{\bbl@umlauta{A}}%
1610   \DeclareTextCompositeCommand{"}{OT1}{E}{\bbl@umlaute{E}}%
1611   \DeclareTextCompositeCommand{"}{OT1}{I}{\bbl@umlaute{I}}%
1612   \DeclareTextCompositeCommand{"}{OT1}{O}{\bbl@umlauta{O}}%
1613   \DeclareTextCompositeCommand{"}{OT1}{U}{\bbl@umlauta{U}}%
1614 }

```

Finally, the default is to use English as the main language.

```

1615 \ifx\l@english\@undefined
1616   \chardef\l@english\z@
1617 \fi
1618 \main@language{english}

```

## 9.12 Layout

### Work in progress.

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

```

1619 \bbl@trace{Bidi layout}
1620 \providecommand\IfBabelLayout[3]{#3}%
1621 \newcommand\BabelPatchSection[1]{%
1622   \ifundefined{#1}{%
1623     \bbl@exp{\let\<bbl@ss@#1>\<#1>}%
1624     \namedef{#1}{%
1625       \ifstar{\bbl@presec{s{#1}}%
1626         {\@dblarg{\bbl@presec{x{#1}}}}}%
1627     \def\bbl@presec{x#1[#2]#3}%
1628     \bbl@exp{%
1629       \\select@language{x{\bbl@main@language}}%
1630       \\@nameuse{bbl@sspre@#1}%
1631       \\@nameuse{bbl@ss@#1}%
1632       [\\foreignlanguage{\languagename}{\unexpanded{#2}}}%
1633       {\\foreignlanguage{\languagename}{\unexpanded{#3}}}%
1634       \\select@language{x{\languagename}}}%
1635     \def\bbl@presec{s#1#2{%
1636       \bbl@exp{%
1637         \\select@language{x{\bbl@main@language}}%
1638         \\@nameuse{bbl@sspre@#1}%
1639         \\@nameuse{bbl@ss@#1}%
1640         {\\foreignlanguage{\languagename}{\unexpanded{#2}}}%
1641         \\select@language{x{\languagename}}}%
1642     \IfBabelLayout{sectioning}%
1643     {\BabelPatchSection{part}}%
1644     \BabelPatchSection{chapter}%
1645     \BabelPatchSection{section}%
1646     \BabelPatchSection{subsection}%

```

```

1647 \BabelPatchSection{subsubsection}%
1648 \BabelPatchSection{paragraph}%
1649 \BabelPatchSection{subparagraph}%
1650 \def\babel@toc#1{%
1651 \select@language@x{\bbl@main@language}}{}
1652 \IfBabelLayout{captions}%
1653 {\BabelPatchSection{caption}}{}

```

### 9.13 Load engine specific macros

```

1654 \bbl@trace{Input engine specific macros}
1655 \ifcase\bbl@engine
1656 \input txtbabel.def
1657 \or
1658 \input luababel.def
1659 \or
1660 \input xebabel.def
1661 \fi

```

### 9.14 Creating languages

`\babelprovide` is a general purpose tool for creating and modifying languages. It creates the language infrastructure, and loads, if requested, an ini file. It may be used in conjunction to previously loaded ldf files.

```

1662 \bbl@trace{Creating languages and reading ini files}
1663 \newcommand\babelprovide[2][{}]{%
1664 \let\bbl@savelangname\language
1665 \edef\bbl@savelocaleid{\the\localeid}%
1666 % Set name and locale id
1667 \def\language{#2}%
1668 \bbl@id@assign
1669 \chardef\localeid\@nameuse{\bbl@id@\language}%
1670 \let\bbl@KVP@captions\@nil
1671 \let\bbl@KVP@import\@nil
1672 \let\bbl@KVP@main\@nil
1673 \let\bbl@KVP@script\@nil
1674 \let\bbl@KVP@language\@nil
1675 \let\bbl@KVP@dir\@nil
1676 \let\bbl@KVP@hyphenrules\@nil
1677 \let\bbl@KVP@mapfont\@nil
1678 \let\bbl@KVP@maparabic\@nil
1679 \let\bbl@KVP@mapdigits\@nil
1680 \let\bbl@KVP@intraspace\@nil
1681 \let\bbl@KVP@intrapenalty\@nil
1682 \bbl@forkv{#1}{\bbl@csarg\def{KVP@##1}{##2}}% TODO - error handling
1683 \ifx\bbl@KVP@import\@nil\else
1684 \bbl@exp{\bbl@ifblank{\bbl@KVP@import}}%
1685 {\begingroup
1686 \def\BabelBeforeIni##1##2{\gdef\bbl@KVP@import{##1}\endinput}%
1687 \InputIfFileExists{babel-#2.tex}{}{}%
1688 \endgroup}%
1689 {}%
1690 \fi
1691 \ifx\bbl@KVP@captions\@nil
1692 \let\bbl@KVP@captions\bbl@KVP@import
1693 \fi
1694 % Load ini
1695 \bbl@ifunset{date#2}%

```

```

1696 {\bbl@provide@new{#2}}%
1697 {\bbl@ifblank{#1}}%
1698 {\bbl@error
1699   {If you want to modify `#2' you must tell how in\\%
1700     the optional argument. See the manual for the\\%
1701     available options.}%
1702   {Use this macro as documented}}%
1703 {\bbl@provide@renew{#2}}}%
1704 % Post tasks
1705 \bbl@exp{\\babelensure[exclude=\\today]{#2}}%
1706 \bbl@ifunset{bbl@ensure@\\languagename}%
1707   {\bbl@exp{%
1708     \\DeclareRobustCommand<bbl@ensure@\\languagename>[1]{%
1709       \\foreignlanguage{\\languagename}%
1710       {###1}}}%
1711   }%
1712 % At this point all parameters are defined if 'import'. Now we
1713 % execute some code depending on them. But what about if nothing was
1714 % imported? We just load the very basic parameters: ids and a few
1715 % more.
1716 \bbl@ifunset{bbl@lname@#2}%
1717   {\def\BabelBeforeIni##1##2{%
1718     \begingroup
1719       \catcode`\[=12 \catcode`\]=12 \catcode`\==12 %
1720       \let\bbl@ini@captions@aux\@gobbletwo
1721       \def\bbl@inidate ###1.###2.###3.###4\relax ###5###6{%
1722         \bbl@read@ini{##1}%
1723         \bbl@exportkey{chrng}{characters.ranges}{}%
1724         \bbl@exportkey{dgnat}{numbers.digits.native}{}%
1725         \endgroup%
1726         {\setbox\z@\hbox{\InputIfFileExists{babel-#2.tex}{}}}%
1727       }%
1728   } -
1729 % Override script and language names with script= and language=
1730 \ifx\bbl@KVP@script\@nil\else
1731   \bbl@csarg\edef{sname@#2}{\bbl@KVP@script}%
1732 \fi
1733 \ifx\bbl@KVP@language\@nil\else
1734   \bbl@csarg\edef{lname@#2}{\bbl@KVP@language}%
1735 \fi
1736 % For bidi texts, to switch the language based on direction
1737 \ifx\bbl@KVP@mapfont\@nil\else
1738   \bbl@ifsamestring{\bbl@KVP@mapfont}{direction}{}%
1739   {\bbl@error{Option `bbl@KVP@mapfont' unknown for\\%
1740     mapfont. Use `direction'.%
1741     {See the manual for details.}}}%
1742 \bbl@ifunset{bbl@lsys@\\languagename}{\bbl@provide@lsys{\\languagename}}}%
1743 \bbl@ifunset{bbl@wdir@\\languagename}{\bbl@provide@dirs{\\languagename}}}%
1744 \ifx\bbl@mapselect\@undefined
1745   \AtBeginDocument{%
1746     \expandafter\bbl@add\csname selectfont \endcsname{{\bbl@mapselect}}%
1747     {\selectfont}}%
1748   \def\bbl@mapselect{%
1749     \let\bbl@mapselect\relax
1750     \edef\bbl@prefontid{\fontid\font}%
1751     \def\bbl@mapdir##1{%
1752       {\def\\languagename{##1}%
1753         \let\bbl@ifrestoring\@firstoftwo % avoid font warning
1754         \bbl@switchfont

```

```

1755 \directlua{Babel.fontmap
1756 [\the\csname bbl@wdir@##1\endcsname]%
1757 [\bbl@prefontid]=\fontid\font}}}%
1758 \fi
1759 \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\language\language}}}%
1760 \fi
1761 % For East Asian, Southeast Asian, if interspace in ini - TODO: as hook?
1762 \ifx\bbl@KVP@intraspace\@nil\else % We may override the ini
1763 \bbl@csarg\edef{intsp@#2}{\bbl@KVP@intraspace}%
1764 \fi
1765 \ifcase\bbl@engine\or
1766 \bbl@ifunset{\bbl@intsp@\language\language}{}%
1767 {\expandafter\ifx\csname bbl@intsp@\language\language\endcsname\@empty\else
1768 \bbl@xin@{\bbl@cs{sbcp@\language\language}}{Hant,Hans,Jpan,Kore,Kana}%
1769 \ifin@
1770 \bbl@cjkintspace
1771 \directlua{
1772 Babel = Babel or {}
1773 Babel.locale_props = Babel.locale_props or {}
1774 Babel.locale_props[\the\localeid].linebreak = 'c'
1775 }%
1776 \bbl@exp{\bbl@intraspace\bbl@cs{intsp@\language\language}}\@{}%
1777 \ifx\bbl@KVP@intrapenalty\@nil
1778 \bbl@intrapenalty0\@@
1779 \fi
1780 \else
1781 \bbl@seaintraspace
1782 \bbl@exp{\bbl@intraspace\bbl@cs{intsp@\language\language}}\@{}%
1783 \directlua{
1784 Babel = Babel or {}
1785 Babel.sea_ranges = Babel.sea_ranges or {}
1786 Babel.set_chranges('\bbl@cs{sbcp@\language\language}',
1787 '\bbl@cs{chrng@\language\language}')
1788 }%
1789 \ifx\bbl@KVP@intrapenalty\@nil
1790 \bbl@intrapenalty0\@@
1791 \fi
1792 \fi
1793 \fi
1794 \ifx\bbl@KVP@intrapenalty\@nil\else
1795 \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@@
1796 \fi}%
1797 \or
1798 \bbl@xin@{\bbl@cs{sbcp@\language\language}}{Thai,Lao,Khmr}%
1799 \ifin@
1800 \bbl@ifunset{\bbl@intsp@\language\language}{}%
1801 {\expandafter\ifx\csname bbl@intsp@\language\language\endcsname\@empty\else
1802 \ifx\bbl@KVP@intraspace\@nil
1803 \bbl@exp{%
1804 \bbl@intraspace\bbl@cs{intsp@\language\language}}\@{}%
1805 \fi
1806 \ifx\bbl@KVP@intrapenalty\@nil
1807 \bbl@intrapenalty0\@@
1808 \fi
1809 \fi
1810 \ifx\bbl@KVP@intraspace\@nil\else % We may override the ini
1811 \expandafter\bbl@intraspace\bbl@KVP@intraspace\@@
1812 \fi
1813 \ifx\bbl@KVP@intrapenalty\@nil\else

```

```

1814     \expandafter\bb1@intrapenalty\bb1@KVP@intrapenalty\@@
1815     \fi
1816     \ifx\bb1@ispace\@undefined
1817         \AtBeginDocument{%
1818             \expandafter\bb1@add
1819             \csname selectfont \endcsname{\bb1@ispace}%
1820             \def\bb1@ispace{\bb1@cs{xeisp@\bb1@cs{sbcp@language}}}%
1821         \fi}%
1822     \fi
1823 \fi
1824 % Native digits, if provided in ini (TeX level, xe and lua)
1825 \ifcase\bb1@engine\else
1826     \bb1@ifunset{\bb1@dgnat@language}{}%
1827     {\expandafter\ifx\csname \bb1@dgnat@language\endcsname\@empty\else
1828         \expandafter\expandafter\expandafter
1829         \bb1@setdigits\csname \bb1@dgnat@language\endcsname
1830         \ifx\bb1@KVP@maparabic\@nil\else
1831             \ifx\bb1@latinarabic\@undefined
1832                 \expandafter\let\expandafter\@arabic
1833                 \csname \bb1@counter@language\endcsname
1834             \else % ie, if layout=counters, which redefines \@arabic
1835                 \expandafter\let\expandafter\bb1@latinarabic
1836                 \csname \bb1@counter@language\endcsname
1837             \fi
1838         \fi
1839     \fi}%
1840 \fi
1841 % Native digits (lua level).
1842 \ifodd\bb1@engine
1843     \ifx\bb1@KVP@mapdigits\@nil\else
1844         \bb1@ifunset{\bb1@dgnat@language}{}%
1845         {\RequirePackage{luatexbase}%
1846         \bb1@activate@preotf
1847         \directlua{
1848             Babel = Babel or {} %%% -> presets in luababel
1849             Babel.digits_mapped = true
1850             Babel.digits = Babel.digits or {}
1851             Babel.digits[\the\localeid] =
1852                 table.pack(string.utfvalue('\bb1@cs{dgnat@language}'))
1853             if not Babel.numbers then
1854                 function Babel.numbers(head)
1855                     local LOCALE = luatexbase.registernumber'\bb1@attr@locale'
1856                     local GLYPH = node.id'glyph'
1857                     local inmath = false
1858                     for item in node.traverse(head) do
1859                         if not inmath and item.id == GLYPH then
1860                             local temp = node.get_attribute(item, LOCALE)
1861                             if Babel.digits[temp] then
1862                                 local chr = item.char
1863                                 if chr > 47 and chr < 58 then
1864                                     item.char = Babel.digits[temp][chr-47]
1865                                 end
1866                             end
1867                         elseif item.id == node.id'math' then
1868                             inmath = (item.subtype == 0)
1869                         end
1870                     end
1871                     return head
1872                 end

```

```

1873         end
1874     }}
1875 \fi
1876 \fi
1877 % To load or reload the babel-*.tex, if require.babel in ini
1878 \bbl@ifunset{bbl@rqtex@language}{%
1879     {\expandafter\ifx\csname bbl@rqtex@language\endcsname\@empty\else
1880         \let\BabelBeforeIni\@gobbletwo
1881         \chardef\atcatcode=\catcode`\@
1882         \catcode`\@=11\relax
1883         \InputIfFileExists{babel-\bbl@cs{rqtex@language}.tex}{\}%
1884         \catcode`\@=\atcatcode
1885         \let\atcatcode\relax
1886     \fi}%
1887 \let\language\bbbl@savelangname
1888 \chardef\localeid\bbbl@savelocaleid\relax}

```

[illegible]

```

1919 \def\bbl@provide@new#1{%
1920   \namedef{date#1}{}% marks lang exists - required by \StartBabelCommands
1921   \namedef{extras#1}{}%
1922   \namedef{noextras#1}{}%
1923   \StartBabelCommands*{#1}{captions}%
1924   \ifx\bbl@KVP@captions\@nil % and also if import, implicit
1925     \def\bbl@tempb##1{% elt for \bbl@captionslist

```

```

1926     \ifx##1\@empty\else
1927     \bbl@exp{%
1928         \\\SetString\\##1{%
1929             \\\bbl@nocaption{\bbl@stripslash##1}{#1\bbl@stripslash##1}}}%
1930     \expandafter\bbl@tempb
1931     \fi}%
1932     \expandafter\bbl@tempb\bbl@captionslist\@empty
1933 \else
1934     \bbl@read@ini{\bbl@KVP@captions}% Here all letters cat = 11
1935     \bbl@after@ini
1936     \bbl@savestrings
1937     \fi
1938 \StartBabelCommands*{#1}{date}%
1939 \ifx\bbl@KVP@import\@nil
1940     \bbl@exp{%
1941         \\\SetString\\today{\\\bbl@nocaption{today}{#1today}}}%
1942     \else
1943         \bbl@savetoday
1944         \bbl@savedate
1945     \fi
1946 \EndBabelCommands
1947 \bbl@exp{%
1948     \def\<#1hyphenmins>{%
1949         {\bbl@ifunset{\bbl@ifthm@#1}{2}{\@nameuse{\bbl@ifthm@#1}}}%
1950         {\bbl@ifunset{\bbl@rgthm@#1}{3}{\@nameuse{\bbl@rgthm@#1}}}}}%
1951 \bbl@provide@hyphens{#1}%
1952 \ifx\bbl@KVP@main\@nil\else
1953     \expandafter\main@language\expandafter{#1}%
1954 \fi}
1955 \def\bbl@provide@renew#1{%
1956     \ifx\bbl@KVP@captions\@nil\else
1957         \StartBabelCommands*{#1}{captions}%
1958         \bbl@read@ini{\bbl@KVP@captions}% Here all letters cat = 11
1959         \bbl@after@ini
1960         \bbl@savestrings
1961         \EndBabelCommands
1962     \fi
1963     \ifx\bbl@KVP@import\@nil\else
1964         \StartBabelCommands*{#1}{date}%
1965         \bbl@savetoday
1966         \bbl@savedate
1967         \EndBabelCommands
1968     \fi
1969     \bbl@provide@hyphens{#1}}

```

The hyphenrules option is handled with an auxiliary macro.

```

1970 \def\bbl@provide@hyphens#1{%
1971     \let\bbl@tempa\relax
1972     \ifx\bbl@KVP@hyphenrules\@nil\else
1973         \bbl@replace\bbl@KVP@hyphenrules{ }{,}%
1974         \bbl@foreach\bbl@KVP@hyphenrules{%
1975             \ifx\bbl@tempa\relax % if not yet found
1976                 \bbl@ifsamestring{##1}{+}%
1977                 {\bbl@exp{\addlanguage\<l@##1>}}}%
1978                 {}%
1979             \bbl@ifunset{l@##1}%
1980                 {}%
1981                 {\bbl@exp{\let\bbl@tempa\<l@##1>}}}%
1982         \fi}%

```

```

1983 \fi
1984 \ifx\bbbl@tempa\relax % if no opt or no language in opt found
1985 \ifx\bbbl@KVP@import\@nil\else % if importing
1986 \bbbl@exp{% and hyphenrules is not empty
1987 \\\bbbl@ifblank{\@nameuse{bbbl@hyphr@#1}}}%
1988 }}%
1989 {\let\\bbbl@tempa\<l@ \@nameuse{bbbl@hyphr@ \language name}>}}}%
1990 \fi
1991 \fi
1992 \bbbl@ifunset{bbbl@tempa}% ie, relax or undefined
1993 {\bbbl@ifunset{l@#1}% no hyphenrules found - fallback
1994 {\bbbl@exp{\\adddialect\<l@#1>\language}}}%
1995 }}% so, l@<lang> is ok - nothing to do
1996 {\bbbl@exp{\\adddialect\<l@#1>\bbbl@tempa}}}% found in opt list or ini

```

The reader of ini files. There are 3 possible cases: a section name (in the form [ . . . ]), a comment (starting with ;) and a key/value pair. *TODO - Work in progress.*

```

1997 \def\bbbl@read@ini#1{%
1998 \openin1=babel-#1.ini % FIXME - number must not be hardcoded
1999 \ifeof1
2000 \bbbl@error
2001 {There is no ini file for the requested language\\%
2002 (#1). Perhaps you misspelled it or your installation\\%
2003 is not complete.}%
2004 {Fix the name or reinstall babel.}%
2005 \else
2006 \let\bbbl@section\@empty
2007 \let\bbbl@savestrings\@empty
2008 \let\bbbl@savetoday\@empty
2009 \let\bbbl@savestate\@empty
2010 \let\bbbl@inireader\bbbl@iniskip
2011 \bbbl@info{Importing data from babel-#1.ini for \language name}%
2012 \loop
2013 \if T\ifeof1F\fi T\relax % Trick, because inside \loop
2014 \endlinechar\m@ne
2015 \read1 to \bbbl@line
2016 \endlinechar\^^M
2017 \ifx\bbbl@line\@empty\else
2018 \expandafter\bbbl@inline\bbbl@line\bbbl@inline
2019 \fi
2020 \repeat
2021 \fi}
2022 \def\bbbl@inline#1\bbbl@inline{%
2023 \@ifnextchar[\bbbl@inisec{\@ifnextchar;\bbbl@iniskip\bbbl@inireader}#1\@{}% ]

```

The special cases for comment lines and sections are handled by the two following commands. In sections, we provide the possibility to take extra actions at the end or at the start (TODO - but note the last section is not ended). By default, key=val pairs are ignored.

```

2024 \def\bbbl@iniskip#1\@{}% if starts with ;
2025 \def\bbbl@inisec[#1]#2\@{}% if starts with opening bracket
2026 \@nameuse{bbbl@secpost@bbbl@section}% ends previous section
2027 \def\bbbl@section#1}%
2028 \@nameuse{bbbl@secpre@bbbl@section}% starts current section
2029 \bbbl@ifunset{bbbl@inikv@#1}%
2030 {\let\bbbl@inireader\bbbl@iniskip}%
2031 {\bbbl@exp{\let\\bbbl@inireader\<bbbl@inikv@#1>}}%

```

Reads a key=val line and stores the trimmed val in \bbbl@kv@<section>.<key>.

```

2032 \def\bbbl@inikv#1=#2\@{}% key=value

```



```

2033 \bbl@trim@def\bbl@tempa{#1}%
2034 \bbl@trim\toks@{#2}%
2035 \bbl@csarg\edef{kv@\bbl@section.\bbl@tempa}{\the\toks@}}

```

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

```

2036 \def\bbl@exportkey#1#2#3{%
2037 \bbl@ifunset{bbl@kv@#2}%
2038 {\bbl@csarg\gdef{#1@\language\name}{#3}}%
2039 {\expandafter\ifx\csname bbl@kv@#2\endcsname\empty
2040 \bbl@csarg\gdef{#1@\language\name}{#3}}%
2041 \else
2042 \bbl@exp{\global\let\<bbl@#1@\language\name>\<bbl@kv@#2>}%
2043 \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.

```

2044 \let\bbl@inikv@identification\bbl@inikv
2045 \def\bbl@secpost@identification{%
2046 \bbl@exportkey{lname}{identification.name.english}{}}%
2047 \bbl@exportkey{lbcpr}{identification.tag.bcp47}{}}%
2048 \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
2049 \bbl@exportkey{sname}{identification.script.name}{}}%
2050 \bbl@exportkey{sbcpr}{identification.script.tag.bcp47}{}}%
2051 \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}}
2052 \let\bbl@inikv@typography\bbl@inikv
2053 \let\bbl@inikv@characters\bbl@inikv
2054 \let\bbl@inikv@numbers\bbl@inikv
2055 \def\bbl@after@ini{%
2056 \bbl@exportkey{lftm}{typography.lefthyphenmin}{2}}%
2057 \bbl@exportkey{rgtm}{typography.righthyphenmin}{3}}%
2058 \bbl@exportkey{hyphr}{typography.hyphenrules}{}}%
2059 \bbl@exportkey{intsp}{typography.intraspace}{}}%
2060 \bbl@exportkey{jstfy}{typography.justify}{w}}%
2061 \bbl@exportkey{chrng}{characters.ranges}{}}%
2062 \bbl@exportkey{dgnat}{numbers.digits.native}{}}%
2063 \bbl@exportkey{rqtex}{identification.require.babel}{}}%
2064 \bbl@xin@{0.5}{\@nameuse{bbl@kv@identification.version}}%
2065 \ifin@
2066 \bbl@warning{%
2067 There are neither captions nor date in '\language\name'.\%
2068 It may not be suitable for proper typesetting, and it\%
2069 could change. Reported}%
2070 \fi
2071 \bbl@xin@{0.9}{\@nameuse{bbl@kv@identification.version}}%
2072 \ifin@
2073 \bbl@warning{%
2074 The '\language\name' date format may not be suitable\%
2075 for proper typesetting, and therefore it very likely will\%
2076 change in a future release. Reported}%
2077 \fi
2078 \bbl@toglobal\bbl@savetoday
2079 \bbl@toglobal\bbl@savestate}

```

Now captions and captions.licr, depending on the engine. And below also for dates. They rely on a few auxiliary macros. It is expected the ini file provides the complete set in Unicode and LICR, in that order.

```

2080 \ifcase\bbl@engine
2081 \bbl@csarg\def{inikv@captions.licr}#1=#2\@@{%

```

```

2082 \bbl@ini@captions@aux{#1}{#2}}
2083 \else
2084 \def\bbl@inikv@captions#1=#2\@@{%
2085 \bbl@ini@captions@aux{#1}{#2}}
2086 \fi

```

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

```

2087 \def\bbl@ini@captions@aux#1#2{%
2088 \bbl@trim@def\bbl@tempa{#1}%
2089 \bbl@ifblank{#2}%
2090 {\bbl@exp{%
2091 \toks@{\bbl@nocaption{\bbl@tempa}{\language\language\bbl@tempa name}}}%
2092 {\bbl@trim\toks@{#2}}}%
2093 \bbl@exp{%
2094 \bbl@add\bbl@savestrings{%
2095 \SetString\<\bbl@tempa name>{\the\toks@}}}%

```

But dates are more complex. The full date format is stores in date.gregorian, so we must read it in non-Unicode engines, too (saved months are just discarded when the LICR section is reached).

TODO. Remove copypaste pattern.

```

2096 \bbl@csarg\def{inikv@date.gregorian}#1=#2\@@{% for defaults
2097 \bbl@inidate#1...\relax{#2}{}}
2098 \bbl@csarg\def{inikv@date.islamic}#1=#2\@@{%
2099 \bbl@inidate#1...\relax{#2}{islamic}}
2100 \bbl@csarg\def{inikv@date.hebrew}#1=#2\@@{%
2101 \bbl@inidate#1...\relax{#2}{hebrew}}
2102 \bbl@csarg\def{inikv@date.persian}#1=#2\@@{%
2103 \bbl@inidate#1...\relax{#2}{persian}}
2104 \bbl@csarg\def{inikv@date.indian}#1=#2\@@{%
2105 \bbl@inidate#1...\relax{#2}{indian}}
2106 \ifcase\bbl@engine
2107 \bbl@csarg\def{inikv@date.gregorian.licr}#1=#2\@@{% override
2108 \bbl@inidate#1...\relax{#2}{}}
2109 \bbl@csarg\def{secpre@date.gregorian.licr}{% discard uni
2110 \ifcase\bbl@engine\let\bbl@savestate\empty\fi}
2111 \fi
2112 % eg: 1=months, 2=wide, 3=1, 4=dummy
2113 \def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
2114 \bbl@trim@def\bbl@tempa{#1.#2}%
2115 \bbl@ifsamestring{\bbl@tempa}{months.wide}% to savedate
2116 {\bbl@trim@def\bbl@tempa{#3}%
2117 \bbl@trim\toks@{#5}%
2118 \bbl@exp{%
2119 \bbl@add\bbl@savestate{%
2120 \SetString\<month\romannumeral\bbl@tempa#6name>{\the\toks@}}}%
2121 {\bbl@ifsamestring{\bbl@tempa}{date.long}% defined now
2122 {\bbl@trim@def\bbl@toreplace{#5}%
2123 \bbl@TG@date
2124 \global\bbl@csarg\let{date@\language}\bbl@toreplace
2125 \bbl@exp{%
2126 \gdef\<\language date>{\protect\<\language date >}%
2127 \gdef\<\language date >####1####2####3{%
2128 \bbl@usedategroupttrue
2129 \<\bbl@ensure@\language>{%
2130 \<\bbl@date@\language>{####1}{####2}{####3}}}%
2131 \bbl@add\bbl@savetoday{%
2132 \SetString\<\today{%
2133 \<\language date>{\the\year}{\the\month}{\the\day}}}%

```

2134       {}}

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.

```
2135 \let\bbl@calendar\@empty
2136 \newcommand\BabelDateSpace{\nobreakspace}
2137 \newcommand\BabelDateDot{.\@}
2138 \newcommand\BabelDated[1]{\number#1}
2139 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
2140 \newcommand\BabelDateM[1]{\number#1}
2141 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}
2142 \newcommand\BabelDateMMMM[1]{\%
2143   \csname month\romannumeral#1\bbl@calendar name\endcsname}%
2144 \newcommand\BabelDatey[1]{\number#1}%
2145 \newcommand\BabelDateyy[1]{\%
2146   \ifnum#1<10 0\number#1 %
2147   \else\ifnum#1<100 \number#1 %
2148   \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
2149   \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
2150   \else
2151     \bbl@error
2152     {Currently two-digit years are restricted to the\
2153     range 0-9999.}%
2154     {There is little you can do. Sorry.}%
2155     \fi\fi\fi\fi}}
2156 \newcommand\BabelDateyyyy[1]{\number#1} % FIXME - add leading 0
2157 \def\bbl@replace@finish@iii#1{%
2158   \bbl@exp{\def\#1####1####2####3{\the\toks@}}
2159 \def\bbl@TG@date{%
2160   \bbl@replace\bbl@toreplace{[ ]}{\BabelDateSpace{}}%
2161   \bbl@replace\bbl@toreplace{[. ]}{\BabelDateDot{}}%
2162   \bbl@replace\bbl@toreplace{[d]}{\BabelDated{####3}}%
2163   \bbl@replace\bbl@toreplace{[dd]}{\BabelDatedd{####3}}%
2164   \bbl@replace\bbl@toreplace{[M]}{\BabelDateM{####2}}%
2165   \bbl@replace\bbl@toreplace{[MM]}{\BabelDateMM{####2}}%
2166   \bbl@replace\bbl@toreplace{[MMMM]}{\BabelDateMMMM{####2}}%
2167   \bbl@replace\bbl@toreplace{[y]}{\BabelDatey{####1}}%
2168   \bbl@replace\bbl@toreplace{[yy]}{\BabelDateyy{####1}}%
2169   \bbl@replace\bbl@toreplace{[yyyy]}{\BabelDateyyyy{####1}}%
2170 % Note after \bbl@replace \toks@ contains the resulting string.
2171 % TODO - Using this implicit behavior doesn't seem a good idea.
2172   \bbl@replace@finish@iii\bbl@toreplace}
```

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

```
2173 \def\bbl@provide@lsys#1{%
2174   \bbl@ifunset{\bbl@lname@#1}%
2175     {\bbl@ini@ids{#1}}%
2176     {}%
2177   \bbl@csarg\let{lsys@#1}\@empty
2178   \bbl@ifunset{\bbl@sname@#1}{\bbl@csarg\gdef{sname@#1}{Default}}{}%
2179   \bbl@ifunset{\bbl@sotf@#1}{\bbl@csarg\gdef{sotf@#1}{DFLT}}{}%
2180   \bbl@csarg\bbl@add@list{lsys@#1}{Script=\bbl@cs{sname@#1}}%
2181   \bbl@ifunset{\bbl@lname@#1}{%
2182     {\bbl@csarg\bbl@add@list{lsys@#1}{Language=\bbl@cs{lname@#1}}}%
2183     \bbl@csarg\bbl@together{lsys@#1}}
```

The following ini reader ignores everything but the identification section. It is called when a font is defined (ie, when the language is first selected) to know which

script/language must be enabled. This means we must make sure a few characters are not active. The ini is not read directly, but with a proxy tex file named as the language.

```

2184 \def\bbl@ini@ids#1{%
2185   \def\BabelBeforeIni##1##2{%
2186     \begingroup
2187       \bbl@add\bbl@secpost@identification{\closein1 }%
2188       \catcode`\[=12 \catcode`\]=12 \catcode`\==12 %
2189       \bbl@read@ini{##1}%
2190       \endgroup}%
2191   \setbox\z@\hbox{\InputIfFileExists{babel-#1.tex}{}}}%

```

## 10 The kernel of Babel (babel.def, only $\LaTeX$ )

### 10.1 The redefinition of the style commands

The rest of the code in this file can only be processed by  $\LaTeX$ , so we check the current format. If it is plain  $\TeX$ , processing should stop here. But, because of the need to limit the scope of the definition of `\format`, a macro that is used locally in the following `\if` statement, this comparison is done inside a group. To prevent  $\TeX$  from complaining about an unclosed group, the processing of the command `\endinput` is deferred until after the group is closed. This is accomplished by the command `\aftergroup`.

```

2192 {\def\format{lplain}
2193 \ifx\fmtname\format
2194 \else
2195   \def\format{LaTeX2e}
2196   \ifx\fmtname\format
2197   \else
2198     \aftergroup\endinput
2199   \fi
2200 \fi}

```

### 10.2 Cross referencing macros

The  $\LaTeX$  book states:

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

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

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

The only way to accomplish this in most cases is to use the trick described in the  $\TeX$ book [2] (Appendix D, page 382). The primitive `\meaning` applied to a token expands to the current meaning of this token. For example, ‘`\meaning\A`’ with `\A` defined as ‘`\def\A#1{\B}`’ expands to the characters ‘`macro:#1->\B`’ with all category codes set to ‘other’ or ‘space’.

`\newlabel` The macro `\label` writes a line with a `\newlabel` command into the `.aux` file to define labels.

```

2201 %\bbl@redefine\newlabel#1#2{%
2202 % \@safe@activestrue\org@newlabel{#1}{#2}\@safe@activesfalse}

```

`\@newl@bel` We need to change the definition of the  $\LaTeX$ -internal macro `\@newl@bel`. This is needed because we need to make sure that shorthand characters expand to their non-active version.

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

```
2203 <<*More package options>> ≡
2204 \DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
2205 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
2206 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
2207 <</More package options>>
```

First we open a new group to keep the changed setting of `\protect` local and then we set the `@safe@actives` switch to true to make sure that any shorthand that appears in any of the arguments immediately expands to its non-active self.

```
2208 \bbl@trace{Cross referencing macros}
2209 \ifx\bbl@opt@safe\@empty\else
2210   \def\@newl@bel#1#2#3{%
2211     {\@safe@activestrue
2212       \bbl@ifunset{#1@#2}%
2213       \relax
2214       {\gdef\@multiplelabels{%
2215         \@latex@warning@no@line{There were multiply-defined labels}}%
2216         \@latex@warning@no@line{Label `#2' multiply defined}}%
2217       \global\@namedef{#1@#2}{#3}}}
```

`\@testdef` An internal  $\LaTeX$  macro used to test if the labels that have been written on the `.aux` file have changed. It is called by the `\enddocument` macro. This macro needs to be completely rewritten, using `\meaning`. The reason for this is that in some cases the expansion of `\#1@#2` contains the same characters as the `#3`; but the character codes differ. Therefore  $\LaTeX$  keeps reporting that the labels may have changed.

```
2218 \CheckCommand*\@testdef[3]{%
2219   \def\reserved@a{#3}%
2220   \expandafter\ifx\cname#1@#2\endcsname\reserved@a
2221   \else
2222     \@tempwattrue
2223   \fi}
```

Now that we made sure that `\@testdef` still has the same definition we can rewrite it. First we make the shorthands ‘safe’.

```
2224 \def\@testdef#1#2#3{%
2225   \@safe@activestrue
```

Then we use `\bbl@tempa` as an ‘alias’ for the macro that contains the label which is being checked.

```
2226   \expandafter\let\expandafter\bbl@tempa\cname #1@#2\endcsname
```

Then we define `\bbl@tempb` just as `\@newl@bel` does it.

```
2227   \def\bbl@tempb{#3}%
2228   \@safe@activesfalse
```

When the label is defined we replace the definition of `\bbl@tempa` by its meaning.

```
2229   \ifx\bbl@tempa\relax
2230   \else
2231     \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
2232   \fi
```

We do the same for `\bbl@tempb`.

```
2233   \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
```

If the label didn't change, `\bbl@tempa` and `\bbl@tempb` should be identical macros.

```
2234 \ifx\bbl@tempa\bbl@tempb
2235 \else
2236 \@tempswatru
2237 \fi}
2238 \fi
```

`\ref`    The same holds for the macro `\ref` that references a label and `\pageref` to reference a  
`\pageref` page. So we redefine `\ref` and `\pageref`. While we change these macros, we make them  
robust as well (if they weren't already) to prevent problems if they should become  
expanded at the wrong moment.

```
2239 \bbl@xin@{R}\bbl@opt@safe
2240 \ifin@
2241 \bbl@redefineroobust\ref#1{%
2242 \@safe@activetrue\org@ref{#1}\@safe@activfalse}
2243 \bbl@redefineroobust\pageref#1{%
2244 \@safe@activetrue\org@pageref{#1}\@safe@activfalse}
2245 \else
2246 \let\org@ref\ref
2247 \let\org@pageref\pageref
2248 \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.

```
2249 \bbl@xin@{B}\bbl@opt@safe
2250 \ifin@
2251 \bbl@redefine\@citex[#1]#2{%
2252 \@safe@activetrue\edef\@tempa{#2}\@safe@activfalse
2253 \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.

```
2254 \AtBeginDocument{%
2255 \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.)

```
2256 \def\@citex[#1][#2]#3{%
2257 \@safe@activetrue\edef\@tempa{#3}\@safe@activfalse
2258 \org@@citex[#1][#2]{\@tempa}}%
2259 }{}}
```

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

```
2260 \AtBeginDocument{%
2261 \ifpackageloaded{cite}{%
2262 \def\@citex[#1]#2{%
2263 \@safe@activetrue\org@@citex[#1][#2]\@safe@activfalse}%
2264 }{}}
```

`\nocite` The macro `\nocite` which is used to instruct BiB<sub>T</sub><sub>E</sub>X to extract uncited references from the database.

```

2265 \bbl@redefine\nocite#1{%
2266   \@safe@activestrue\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@activestrue` 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.

```

2267 \bbl@redefine\bibcite{%
2268   \bbl@cite@choice
2269   \bibcite}

```

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

```

2270 \def\bbl@bibcite#1#2{%
2271   \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.

```

2272 \def\bbl@cite@choice{%
2273   \global\let\bibcite\bbl@bibcite

```

Then, when `natbib` is loaded we restore the original definition of `\bibcite`. For `cite` we do the same.

```

2274   \@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{}%
2275   \@ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{}%

```

Make sure this only happens once.

```

2276   \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.

```

2277   \AtBeginDocument{\bbl@cite@choice}

```

`\@bibitem` One of the two internal L<sub>A</sub>T<sub>E</sub>X macros called by `\bibitem` that write the citation label on the `.aux` file.

```

2278 \bbl@redefine\@bibitem#1{%
2279   \@safe@activestrue\org@bibitem{#1}\@safe@activesfalse}
2280 \else
2281   \let\org@nocite\nocite
2282   \let\org@@citex\citex
2283   \let\org@bibcite\bibcite
2284   \let\org@@bibitem\@bibitem
2285 \fi

```

### 10.3 Marks

`\markright` Because the output routine is asynchronous, we must pass the current language attribute to the head lines, together with the text that is put into them. To achieve this we need to adapt the definition of `\markright` and `\markboth` somewhat.

We check whether the argument is empty; if it is, we just make sure the scratch token register is empty. Next, we store the argument to `\markright` in the scratch token register. This way these commands will not be expanded later, and we make sure that the text is typeset using the correct language settings. While doing so, we make sure that active characters that may end up in the mark are not disabled by the output routine kicking in while `\@safe@activetrue` is in effect.

```

2286 \bbl@trace{Marks}
2287 \IfBabelLayout{sectioning}
2288   {\ifx\bbl@opt@headfoot\@nnil
2289     \g@addto@macro\@resetactivechars{%
2290       \set@typeset@protect
2291       \expandafter\select@language@x\expandafter{\bbl@main@language}%
2292       \let\protect\noexpand
2293       \edef\thepage{%
2294         \noexpand\babelsublr{\unexpanded\expandafter{\thepage}}}%
2295     \fi}
2296   {\bbl@redefine\markright#1{%
2297     \bbl@ifblank{#1}%
2298     {\org@markright{}}%
2299     {\toks@{#1}%
2300      \bbl@exp{%
2301        \org@markright{\protect\foreignlanguage{\language}%
2302          {\protect\bbl@restore@actives\the\toks@}}}%

```

`\markboth`    The definition of `\markboth` is equivalent to that of `\markright`, except that we need two token registers. The documentclasses `report` and `book` define and set the headings for the page. While doing so they also store a copy of `\markboth` in `\@mkboth`. Therefore we need to check whether `\@mkboth` has already been set. If so we need to do that again with the new definition of `\markboth`.

```

2303   \ifx\@mkboth\markboth
2304     \def\bbl@tempc{\let\@mkboth\markboth}
2305   \else
2306     \def\bbl@tempc{}
2307   \fi

```

Now we can start the new definition of `\markboth`

```

2308   \bbl@redefine\markboth#1#2{%
2309     \protected@edef\bbl@tempb##1{%
2310       \protect\foreignlanguage
2311       {\language}{\protect\bbl@restore@actives##1}}%
2312     \bbl@ifblank{#1}%
2313     {\toks@{}}%
2314     {\toks@\expandafter{\bbl@tempb{#1}}}%
2315     \bbl@ifblank{#2}%
2316     {\@temptokena{}}%
2317     {\@temptokena\expandafter{\bbl@tempb{#2}}}%
2318     \bbl@exp{\org@markboth{\the\toks@}{\the\@temptokena}}%

```

and copy it to `\@mkboth` if necessary.

```

2319   \bbl@tempc} % end \IfBabelLayout

```

## 10.4 Preventing clashes with other packages

### 10.4.1 `ifthen`

`\ifthenelse`    Sometimes a document writer wants to create a special effect depending on the page a certain fragment of text appears on. This can be achieved by the following piece of code:



```

\ifthenelse{\isodd{\pageref{some:label}}}{
  {code for odd pages}
  {code for even pages}
}

```

In order for this to work the argument of `\isodd` needs to be fully expandable. With the above redefinition of `\pageref` it is not in the case of this example. To overcome that, we add some code to the definition of `\ifthenelse` to make things work.

The first thing we need to do is check if the package `ifthen` is loaded. This should be done at `\begin{document}` time.

```

2320 \bbl@trace{Preventing clashes with other packages}
2321 \bbl@xin@{R}\bbl@opt@safe
2322 \ifin@
2323   \AtBeginDocument{%
2324     \ifpackageloaded{ifthen}{%

```

Then we can redefine `\ifthenelse`:

```

2325     \bbl@redefine@long\ifthenelse#1#2#3{%

```

We want to revert the definition of `\pageref` and `\ref` to their original definition for the first argument of `\ifthenelse`, so we first need to store their current meanings.

```

2326       \let\bbl@temp@pref\pageref
2327       \let\pageref\org@pageref
2328       \let\bbl@temp@ref\ref
2329       \let\ref\org@ref

```

Then we can set the `\@safe@activetrue` switch and call the original `\ifthenelse`. In order to be able to use shorthands in the second and third arguments of `\ifthenelse` the resetting of the switch *and* the definition of `\pageref` happens inside those arguments. When the package wasn't loaded we do nothing.

```

2330       \@safe@activetrue
2331       \org@ifthenelse{#1}%
2332       {\let\pageref\bbl@temp@pref
2333        \let\ref\bbl@temp@ref
2334        \@safe@activetrue
2335        #2}%
2336       {\let\pageref\bbl@temp@pref
2337        \let\ref\bbl@temp@ref
2338        \@safe@activetrue
2339        #3}%
2340     }%
2341   }{}%
2342 }

```

#### 10.4.2 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`.

```

\Ref 2343 \AtBeginDocument{%
2344   \ifpackageloaded{varioref}{%
2345     \bbl@redefine\@@vpageref#1[#2]#3{%
2346       \@safe@activetrue
2347       \org@@@vpageref{#1}[#2]#3}%
2348       \@safe@activetrue}%

```

The same needs to happen for `\vrefpagemum`.

```

2349     \bbl@redefine\vrefpagemum#1#2{%

```

```

2350      \@safe@activetrue
2351      \org@vrefpagenum{#1}{#2}%
2352      \@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.

```

2353      \expandafter\def\csname Ref \endcsname#1{%
2354          \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
2355      }{}%
2356  }
2357 \fi

```

#### 10.4.3 `hhline`

`\hhline` Delaying the activation of the shorthand characters has introduced a problem with the `hhline` package. The reason is that it uses the “`:`” character which is made active by the french support in `babel`. Therefore we need to *reload* the package when the “`:`” is an active character.

So at `\begin{document}` we check whether `hhline` is loaded.

```

2358 \AtEndOfPackage{%
2359     \AtBeginDocument{%
2360         \ifpackageloaded{hhline}%

```

Then we check whether the expansion of `\normal@char:` is not equal to `\relax`.

```

2361         {\expandafter\ifx\csname normal@char:string\endcsname\relax
2362             \else

```

In that case we simply reload the package. Note that this happens *after* the category code of the `@`-sign has been changed to other, so we need to temporarily change it to letter again.

```

2363         \makeatletter
2364         \def\@currname{hhline}\input{hhline.sty}\makeatother
2365     \fi}%
2366 }{}%

```

#### 10.4.4 `hyperref`

`\pdfstringdefDisableCommands` A number of interworking problems between `babel` and `hyperref` are tackled by `hyperref` itself. The following code was introduced to prevent some annoying warnings but it broke bookmarks. This was quickly fixed in `hyperref`, which essentially made it no-op. However, it will not be removed for the moment because `hyperref` is expecting it.

```

2367 \AtBeginDocument{%
2368     \ifx\pdfstringdefDisableCommands\@undefined\else
2369         \pdfstringdefDisableCommands{\languageshorthands{system}}%
2370     \fi}

```

#### 10.4.5 `fancyhdr`

`\FOREIGNLANGUAGE` The package `fancyhdr` treats the running head and foot lines somewhat differently as the standard classes. A symptom of this is that the command `\foreignlanguage` which `babel` adds to the marks can end up inside the argument of `\MakeUppercase`. To prevent unexpected results we need to define `\FOREIGNLANGUAGE` here.

```

2371 \DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
2372     \lowercase{\foreignlanguage{#1}}%

```

`\substitutefontfamily` The command `\substitutefontfamily` creates an `.fd` file on the fly. The first argument is an encoding mnemonic, the second and third arguments are font family names.

```

2373 \def\substitutefontfamily#1#2#3{%
2374   \lowercase{\immediate\openout15=#1#2.fd\relax}%
2375   \immediate\write15{%
2376     \string\ProvidesFile{#1#2.fd}%
2377     [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
2378     \space generated font description file]^^J
2379     \string\DeclareFontFamily{#1}{#2}{^^J
2380     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{^^J
2381     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{^^J
2382     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{^^J
2383     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{^^J
2384     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{^^J
2385     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{^^J
2386     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{^^J
2387     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{^^J
2388     }%
2389     \closeout15
2390   }

```

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

```

2391 \@onlypreamble\substitutefontfamily

```

## 10.5 Encoding and fonts

Because documents may use non-ASCII font encodings, we make sure that the logos of  $\text{\TeX}$  and  $\text{\LaTeX}$  always come out in the right encoding. There is a list of non-ASCII encodings. Unfortunately, `fontenc` deletes its package options, so we must guess which encodings has been loaded by traversing `\@filelist` to search for `\enc.def`. If a non-ASCII has been loaded, we define versions of  $\text{\TeX}$  and  $\text{\LaTeX}$  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`

```

2392 \bbl@trace{Encoding and fonts}
2393 \newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU}
2394 \newcommand\BabelNonText{TS1,T3,TS3}
2395 \let\org@TeX\TeX
2396 \let\org@LaTeX\LaTeX
2397 \let\ensureascii\@firstofone
2398 \AtBeginDocument{%
2399   \in@false
2400   \bbl@foreach\BabelNonASCII{% is there a text non-ascii enc?
2401     \ifin@false
2402       \lowercase{\bbl@xin@{,#1enc.def},{,\@filelist,}}%
2403       \fi}%
2404   \ifin@ % if a text non-ascii has been loaded
2405     \def\ensureascii#1{{\fontencoding{OT1}\selectfont#1}}%
2406     \DeclareTextCommandDefault{\TeX}{\org@TeX}%
2407     \DeclareTextCommandDefault{\LaTeX}{\org@LaTeX}%
2408     \def\bbl@tempb#1\@{\uppercase{\bbl@tempc#1}ENC.DEF\empty\@}%
2409     \def\bbl@tempc#1ENC.DEF#2\@{\%
2410       \ifx\@empty#2\else
2411         \bbl@ifunset{T@#1}%
2412         {}%
2413         {\bbl@xin@{,#1,},{,\BabelNonASCII,\BabelNonText,}}%

```

```

2414         \ifin@
2415         \DeclareTextCommand{\TeX}{#1}{\ensureascii{\org@TeX}}%
2416         \DeclareTextCommand{\LaTeX}{#1}{\ensureascii{\org@LaTeX}}%
2417         \else
2418         \def\ensureascii##1{{\fontencoding{#1}\selectfont##1}}%
2419         \fi}%
2420     \fi}%
2421     \bbl@foreach\@filelist{\bbl@tempb#1\@}% TODO - \@ de mas??
2422     \bbl@xin@{\cf@encoding,}{,\BabelNonASCII,\BabelNonText,}%
2423     \ifin@\else
2424     \edef\ensureascii#1{%
2425         \noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}}%
2426     \fi
2427     \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.

```

2428 \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.

```

2429 \AtBeginDocument{%
2430     \@ifpackageloaded{fontspec}%
2431     {\xdef\latinencoding{%
2432         \ifx\UTFencname\@undefined
2433         EU\ifcase\bbl@engine\or2\or1\fi
2434         \else
2435         \UTFencname
2436         \fi}}%
2437     {\gdef\latinencoding{OT1}%
2438         \ifx\cf@encoding\bbl@t@one
2439         \xdef\latinencoding{\bbl@t@one}%
2440         \else
2441         \@ifl@aded{def}{t1enc}{\xdef\latinencoding{\bbl@t@one}}}%
2442     \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.

```

2443 \DeclareRobustCommand{\latintext}{%
2444     \fontencoding{\latinencoding}\selectfont
2445     \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.

```

2446 \ifx\@undefined\DeclareTextFontCommand
2447     \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
2448 \else
2449     \DeclareTextFontCommand{\textlatin}{\latintext}
2450 \fi

```

## 10.6 Basic bidi support

**Work in progress.** This code is currently placed here for practical reasons.

It is loosely based on `rlbabel.def`, but most of it has been developed from scratch. This babel module (by Johannes Braams and Boris Lavva) has served the purpose of typesetting R documents for two decades, and despite its flaws I think it is still a good starting point (some parts have been copied here almost verbatim), partly thanks to its simplicity. I've also looked at ARABI (by Youssef Jabri), which is compatible with babel.

There are two ways of modifying macros to make them “bidi”, namely, by patching the internal low level macros (which is what I have done with lists, columns, counters, tocs, much like `rlbabel` did), and by introducing a “middle layer” just below the user interface (sectioning, footnotes).

- `pdfTeX` provides a minimal support for bidi text, and it must be done by hand. Vertical typesetting is not possible.
- `xetex` is somewhat better, thanks to its font engine (even if not always reliable) and a few additional tools. However, very little is done at the paragraph level. Another challenging problem is text direction does not honour  $\TeX$  grouping.
- `luatex` can provide the most complete solution, as we can manipulate almost freely the node list, the generated lines, and so on, but bidi text does not work out of the box and some development is necessary. It also provides tools to properly set left-to-right and right-to-left page layouts. As `Lua $\TeX$ -ja` shows, vertical typesetting is possible, too. Its main drawback is font handling is often considered to be less mature than `xetex`, mainly in Indic scripts (but there are steps to make HarfBuzz, the `xetex` font engine, available in `luatex`; see <<https://github.com/tatzetwerk/luatex-harfbuzz>>).

```
2451 \bbl@trace{Basic (internal) bidi support}
2452 \def\bbl@alscripts{,Arabic,Syriac,Thaana,}
2453 \def\bbl@rscripts{%
2454   ,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
2455   Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaean,%
2456   Manichaean,Meroitic Cursive,Meroitic,Old North Arabian,%
2457   Nabataean,N'Ko,Orkhon,Palmyrene,Inscriptional Pahlavi,%
2458   Psalter Pahlavi,Phoenician,Inscriptional Parthian,Samaritan,%
2459   Old South Arabian,}%
2460 \def\bbl@provide@dirs#1{%
2461   \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts\bbl@rscripts}%
2462   \ifin@
2463     \global\bbl@csarg\chardef{wdir@#1}\@ne
2464     \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts}%
2465     \ifin@
2466       \global\bbl@csarg\chardef{wdir@#1}\tw@ % useless in xetex
2467     \fi
2468   \else
2469     \global\bbl@csarg\chardef{wdir@#1}\z@
2470   \fi
2471   \ifodd\bbl@engine
2472     \bbl@csarg\ifcase{wdir@#1}%
2473       \directlua{ Babel.locale_props[\the\localeid].textdir = 'l' }%
2474     \or
2475       \directlua{ Babel.locale_props[\the\localeid].textdir = 'r' }%
2476     \or
2477       \directlua{ Babel.locale_props[\the\localeid].textdir = 'al' }%
2478     \fi
2479   \fi}
2480 \def\bbl@switchdir{%
```

```

2481 \bbl@ifunset{bbl@lsys@\language}\bbl@provide@lsys{\language}}{}%
2482 \bbl@ifunset{bbl@wdir@\language}\bbl@provide@dirs{\language}}{}%
2483 \bbl@exp{\bbl@setdirs\bbl@cs{wdir@\language}}
2484 \def\bbl@setdirs#1{% TODO - math
2485 \ifcase\bbl@select@type % TODO - strictly, not the right test
2486 \bbl@bodydir{#1}%
2487 \bbl@pardir{#1}%
2488 \fi
2489 \bbl@textdir{#1}}
2490 \ifodd\bbl@engine % luatex=1
2491 \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
2492 \DisableBabelHook{babel-bidi}
2493 \chardef\bbl@thetextdir\z@
2494 \chardef\bbl@thepardir\z@
2495 \def\bbl@getluadir#1{%
2496 \directlua{
2497   if tex.#1dir == 'TLT' then
2498     tex.sprint('0')
2499   elseif tex.#1dir == 'TRT' then
2500     tex.sprint('1')
2501   end}}
2502 \def\bbl@setluadir#1#2#3{% 1=text/par.. 2=\textdir.. 3=0 lr/1 rl
2503 \ifcase#3\relax
2504 \ifcase\bbl@getluadir{#1}\relax\else
2505 #2 TLT\relax
2506 \fi
2507 \else
2508 \ifcase\bbl@getluadir{#1}\relax
2509 #2 TRT\relax
2510 \fi
2511 \fi}
2512 \def\bbl@textdir#1{%
2513 \bbl@setluadir{text}\textdir{#1}%
2514 \chardef\bbl@thetextdir#1\relax
2515 \setattribute\bbl@attr@dir{\numexpr\bbl@thepardir*3+#1}}
2516 \def\bbl@pardir#1{%
2517 \bbl@setluadir{par}\pardir{#1}%
2518 \chardef\bbl@thepardir#1\relax}
2519 \def\bbl@bodydir{\bbl@setluadir{body}\bodydir}
2520 \def\bbl@pagedir{\bbl@setluadir{page}\pagedir}
2521 \def\bbl@dirparastext{\pardir\the\textdir\relax}% %%%
2522 % Sadly, we have to deal with boxes in math with basic.
2523 % Activated every math with the package option bidi=:
2524 \def\bbl@mathboxdir{%
2525 \ifcase\bbl@thetextdir\relax
2526 \everyhbox{\textdir TLT\relax}%
2527 \else
2528 \everyhbox{\textdir TRT\relax}%
2529 \fi}
2530 \else % pdftex=0, xetex=2
2531 \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
2532 \DisableBabelHook{babel-bidi}
2533 \newcount\bbl@dirlevel
2534 \chardef\bbl@thetextdir\z@
2535 \chardef\bbl@thepardir\z@
2536 \def\bbl@textdir#1{%
2537 \ifcase#1\relax
2538 \chardef\bbl@thetextdir\z@
2539 \bbl@textdir@i\beginL\endL

```

```

2540 \else
2541 \chardef\bb1@thetextdir\@ne
2542 \bb1@textdir@i\beginR\endR
2543 \fi}
2544 \def\bb1@textdir@i#1#2{%
2545 \ifhmode
2546 \ifnum\currentgrouplevel>\z@
2547 \ifnum\currentgrouplevel=\bb1@dirlevel
2548 \bb1@error{Multiple bidi settings inside a group}%
2549 {I'll insert a new group, but expect wrong results.}%
2550 \bgroup\aftergroup#2\aftergroup\egroup
2551 \else
2552 \ifcase\currentgrouptype\or % 0 bottom
2553 \aftergroup#2% 1 simple {}
2554 \or
2555 \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
2556 \or
2557 \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
2558 \or\or\or % vbox vtop align
2559 \or
2560 \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
2561 \or\or\or\or\or\or % output math disc insert vcent mathchoice
2562 \or
2563 \aftergroup#2% 14 \begingroup
2564 \else
2565 \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
2566 \fi
2567 \fi
2568 \bb1@dirlevel\currentgrouplevel
2569 \fi
2570 #1%
2571 \fi}
2572 \def\bb1@pardir#1{\chardef\bb1@thepardir#1\relax}
2573 \let\bb1@bodydir\@gobble
2574 \let\bb1@pagedir\@gobble
2575 \def\bb1@dirparastext{\chardef\bb1@thepardir\bb1@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).

```

2576 \def\bb1@xebidipar{%
2577 \let\bb1@xebidipar\relax
2578 \TeXeTstate\@ne
2579 \def\bb1@xeverypar{%
2580 \ifcase\bb1@thepardir
2581 \ifcase\bb1@thetextdir\else\beginR\fi
2582 \else
2583 {\setbox\z@\lastbox\beginR\box\z@}%
2584 \fi}%
2585 \let\bb1@severypar\everypar
2586 \newtoks\everypar
2587 \everypar=\bb1@severypar
2588 \bb1@severypar{\bb1@xeverypar\the\everypar}}
2589 \@ifpackagewith{babel}{bidi=bidi}%
2590 {\let\bb1@textdir@i\@gobbletwo
2591 \let\bb1@xebidipar\@empty
2592 \AddBabelHook{bidi}{foreign}{%
2593 \def\bb1@tempa{\def\BabelText###1}%
2594 \ifcase\bb1@thetextdir

```

```

2595         \expandafter\babel@tempa\expandafter{\BabelText{\LR{##1}}}%
2596     \else
2597         \expandafter\babel@tempa\expandafter{\BabelText{\RL{##1}}}%
2598     \fi}
2599     \def\babel@pardir#1{\ifcase#1\relax\setLR\else\setRL\fi}}
2600 {}%
2601 \fi

```

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

```

2602 \DeclareRobustCommand\babelsublr[1]{\leavevmode{\babel@textdir\z@#1}}
2603 \AtBeginDocument{%
2604     \ifx\pdfstringdefDisableCommands\@undefined\else
2605         \ifx\pdfstringdefDisableCommands\relax\else
2606             \pdfstringdefDisableCommands{\let\babelsublr\@firstofone}%
2607         \fi
2608     \fi}

```

## 10.7 Local Language Configuration

`\loadlocalcfg` At some sites it may be necessary to add site-specific actions to a language definition file. This can be done by creating a file with the same name as the language definition file, but with the extension `.cfg`. For instance the file `norsk.cfg` will be loaded when the language definition file `norsk.ldf` is loaded.

For plain-based formats we don't want to override the definition of `\loadlocalcfg` from `plain.def`.

```

2609 \babel@trace{Local Language Configuration}
2610 \ifx\loadlocalcfg\@undefined
2611     \ifpackagewith{babel}{noconfigs}%
2612     {\let\loadlocalcfg\@gobble}%
2613     {\def\loadlocalcfg#1{%
2614         \InputIfFileExists{#1.cfg}%
2615         {\typeout{*****^J%
2616             * Local config file #1.cfg used^^J%
2617             *}}}%
2618     \@empty}}
2619 \fi

```

Just to be compatible with L<sup>A</sup>T<sub>E</sub>X 2.09 we add a few more lines of code:

```

2620 \ifx\@unexpandable@protect\@undefined
2621     \def\@unexpandable@protect{\noexpand\protect\noexpand}
2622     \long\def\protected@write#1#2#3{%
2623         \begingroup
2624             \let\thepage\relax
2625             #2%
2626             \let\protect\@unexpandable@protect
2627             \edef\reserved@a{\write#1{#3}}%
2628             \reserved@a
2629         \endgroup
2630         \if@nobreak\ifvmode\nobreak\fi\fi}
2631 \fi
2632 </core>
2633 <*kernel>

```

## 11 Multiple languages (switch.def)

Plain T<sub>E</sub>X version 3.0 provides the primitive `\language` that is used to store the current language. When used with a pre-3.0 version this function has to be implemented by



allocating a counter.

```
2634 <<Make sure ProvidesFile is defined>>
2635 \ProvidesFile{switch.def}[\<date>] \<version> Babel switching mechanism]
2636 <<Load macros for plain if not LaTeX>>
2637 <<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.

```
2638 \def\bbl@version{\<version>}
2639 \def\bbl@date{\<date>}
2640 \def\adddialect#1#2{%
2641   \global\chardef#1#2\relax
2642   \bbl@usehooks{adddialect}{\#1}{\#2}}%
2643   \wlog{\string#1 = a dialect from \string\language#2}}
```

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

```
2644 \def\bbl@fixname#1{%
2645   \begingroup
2646   \def\bbl@tempe{l@}%
2647   \edef\bbl@tempd{\noexpand\@ifundefined{\noexpand\bbl@tempe#1}}%
2648   \bbl@tempd
2649   {\lowercase\expandafter{\bbl@tempd}%
2650    {\uppercase\expandafter{\bbl@tempd}%
2651     \@empty
2652      {\edef\bbl@tempd{\def\noexpand#1{\#1}}%
2653       \uppercase\expandafter{\bbl@tempd}}}%
2654   {\edef\bbl@tempd{\def\noexpand#1{\#1}}%
2655    \lowercase\expandafter{\bbl@tempd}}}%
2656   \@empty
2657   \edef\bbl@tempd{\endgroup\def\noexpand#1{\#1}}%
2658   \bbl@tempd}
2659 \def\bbl@iflanguage#1{%
2660   \@ifundefined{l@#1}{\@nolanerr{#1}\@gobble}\@firstofone}
```

**\iflanguage** Users might want to test (in a private package for instance) which language is currently active. For this we provide a test macro, `\iflanguage`, that has three arguments. It checks whether the first argument is a known language. If so, it compares the first argument with the value of `\language`. Then, depending on the result of the comparison, it executes either the second or the third argument.

```
2661 \def\iflanguage#1{%
2662   \bbl@iflanguage{#1}{%
2663     \ifnum\csname l@#1\endcsname=\language
2664       \expandafter\@firstoftwo
2665     \else
2666       \expandafter\@secondoftwo
2667     \fi}}
```

## 11.1 Selecting the language

**\selectlanguage** The macro `\selectlanguage` checks whether the language is already defined before it performs its actual task, which is to update `\language` and activate language-specific

definitions.

To allow the call of `\selectlanguage` either with a control sequence name or with a simple string as argument, we have to use a trick to delete the optional escape character. To convert a control sequence to a string, we use the `\string` primitive. Next we have to look at the first character of this string and compare it with the escape character. Because this escape character can be changed by setting the internal integer `\escapechar` to a character number, we have to compare this number with the character of the string. To do this we have to use  $\TeX$ 's backquote notation to specify the character as a number. If the first character of the `\string`'ed argument is the current escape character, the comparison has stripped this character and the rest in the 'then' part consists of the rest of the control sequence name. Otherwise we know that either the argument is not a control sequence or `\escapechar` is set to a value outside of the character range 0–255. If the user gives an empty argument, we provide a default argument for `\string`. This argument should expand to nothing.

```
2668 \let\bb1@select@type\z@
2669 \edef\selectlanguage{%
2670   \noexpand\protect
2671   \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`.

```
2672 \ifx\@undefined\protect\let\protect\relax\fi
```

As  $\LaTeX$  2.09 writes to files *expanded* whereas  $\LaTeX$  2<sub>ε</sub> takes care *not* to expand the arguments of `\write` statements we need to be a bit clever about the way we add information to .aux files. Therefore we introduce the macro `\xstring` which should expand to the right amount of `\string`'s.

```
2673 \ifx\documentclass\@undefined
2674   \def\xstring{\string\string\string}
2675 \else
2676   \let\xstring\string
2677 \fi
```

Since version 3.5 babel writes entries to the auxiliary files in order to typeset table of contents etc. in the correct language environment.

`\bb1@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 `\bb1@pop@language` to be executed at the end of the group. It calls `\bb1@set@language` with the name of the current language as its argument.

`\bb1@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 `\bb1@language@stack` and initially empty.

```
2678 \def\bb1@language@stack{}
```

When using a stack we need a mechanism to push an element on the stack and to retrieve the information afterwards.

`\bb1@push@language` The stack is simply a list of languagenames, separated with a '+' sign; the push function can  
`\bb1@pop@language` be simple:

```
2679 \def\bb1@push@language{%
2680   \xdef\bb1@language@stack{\languagename+\bb1@language@stack}}
```

Retrieving information from the stack is a little bit less simple, as we need to remove the element from the stack while storing it in the macro `\language`. For this we first define a helper function.

`\bbl@pop@lang` This macro stores its first element (which is delimited by the ‘+’-sign) in `\language` and stores the rest of the string (delimited by ‘-’) in its third argument.

```
2681 \def\bbl@pop@lang#1+#2-#3{%
2682   \edef\language{#1}\xdef#3{#2}}
```

The reason for the somewhat weird arrangement of arguments to the helper function is the fact it is called in the following way. This means that before `\bbl@pop@lang` is executed  $\TeX$  first *expands* the stack, stored in `\bbl@language@stack`. The result of that is that the argument string of `\bbl@pop@lang` contains one or more language names, each followed by a ‘+’-sign (zero language names won’t occur as this macro will only be called after something has been pushed on the stack) followed by the ‘-’-sign and finally the reference to the stack.

```
2683 \let\bbl@ifrestoring\@secondoftwo
2684 \def\bbl@pop@language{%
2685   \expandafter\bbl@pop@lang\bbl@language@stack-\bbl@language@stack
2686   \let\bbl@ifrestoring\@firstoftwo
2687   \expandafter\bbl@set@language\expandafter{\language}%
2688   \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.

```
2689 \chardef\localeid\z@
2690 \def\bbl@id@last{0}    % No real need for a new counter
2691 \def\bbl@id@assign{%
2692   \bbl@ifunset{bbl@id@@\language}%
2693   {\count@\bbl@id@last\relax
2694    \advance\count@\@ne
2695    \bbl@csarg\chardef{id@@\language}\count@
2696    \edef\bbl@id@last{\the\count@}%
2697    \ifcase\bbl@engine\or
2698      \directlua{
2699        Babel = Babel or {}
2700        Babel.locale_props = Babel.locale_props or {}
2701        Babel.locale_props[\bbl@id@last] = {}
2702      }%
2703    \fi}%
2704   {}}
```

The unprotected part of `\selectlanguage`.

```
2705 \expandafter\def\csname selectlanguage \endcsname#1{%
2706   \ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel\tw@ \fi
2707   \bbl@push@language
2708   \aftergroup\bbl@pop@language
2709   \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.

```

2710 \def\BabelContentsFiles{toc,lof,lot}
2711 \def\bbl@set@language#1{% from selectlanguage, pop@
2712   \edef\language{%
2713     \ifnum\escapechar=\expandafter`\string#1\@empty
2714     \else\string#1\@empty\fi}%
2715   \select@language{\language}%
2716   % write to auxs
2717   \expandafter\ifx\csname date\language\endcsname\relax\else
2718     \if@filesw
2719       \protected@write\@auxout{}\string\babel@aux{\language}{}%
2720       \bbl@usehooks{write}{}%
2721     \fi
2722   \fi}
2723 \def\select@language#1{% from set@, babel@aux
2724   % set hymap
2725   \ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi
2726   % set name
2727   \edef\language{#1}%
2728   \bbl@fixname\language
2729   \bbl@iflanguage\language{%
2730     \expandafter\ifx\csname date\language\endcsname\relax
2731       \bbl@error
2732       {Unknown language `#1'. Either you have\\%
2733        misspelled its name, it has not been installed,\\%
2734        or you requested it in a previous run. Fix its name,\\%
2735        install it or just rerun the file, respectively. In\\%
2736        some cases, you may need to remove the aux file}%
2737       {You may proceed, but expect wrong results}%
2738     \else
2739       % set type
2740       \let\bbl@select@type\z@
2741       \expandafter\bbl@switch\expandafter{\language}%
2742     \fi}}
2743 \def\babel@aux#1#2{%
2744   \expandafter\ifx\csname date#1\endcsname\relax
2745     \expandafter\ifx\csname bbl@auxwarn@#1\endcsname\relax
2746       \@namedef{bbl@auxwarn@#1}{}%
2747       \bbl@warning
2748       {Unknown language `#1'. Very likely you\\%
2749        requested it in a previous run. Expect some\\%
2750        wrong results in this run, which should vanish\\%
2751        in the next one. Reported}%
2752     \fi
2753   \else
2754     \select@language{#1}%
2755     \bbl@foreach\BabelContentsFiles{%
2756       \@writefile{##1}{\babel@toc{#1}{#2}}}% % TODO - ok in plain?
2757   \fi}
2758 \def\babel@toc#1#2{%
2759   \select@language{#1}}

```

A bit of optimization. Select in heads/foots the language only if necessary. The real thing is in `babel.def`.

```

2760 \let\select@language@x\select@language

```

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

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

Then we have to *redefine* `\originalTeX` to compensate for the things that have been activated. To save memory space for the macro definition of `\originalTeX`, we construct the control sequence name for the `\noextras<lang>` command at definition time by expanding the `\csname` primitive.

Now activate the language-specific definitions. This is done by constructing the names of three macros by concatenating three words with the argument of `\selectlanguage`, and calling these macros.

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

```

2761 \newif\ifbbl@usedategroup
2762 \def\bbl@switch#1{% from select@, foreign@
2763   % restore
2764   \originalTeX
2765   \expandafter\def\expandafter\originalTeX\expandafter{%
2766     \csname noextras#1\endcsname
2767     \let\originalTeX\@empty
2768     \babel@beginsave}%
2769   \bbl@usehooks{afterreset}{}%
2770   \languageshorthands{none}%
2771   % set the locale id
2772   \bbl@id@assign
2773   \chardef\localeid\@nameuse{bbl@id@\@language}%
2774   % switch captions, date
2775   \ifcase\bbl@select@type
2776     \ifhmode
2777       \hskip\z@skip % trick to ignore spaces
2778       \csname captions#1\endcsname\relax
2779       \csname date#1\endcsname\relax
2780       \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
2781     \else
2782       \csname captions#1\endcsname\relax
2783       \csname date#1\endcsname\relax
2784     \fi
2785   \else
2786     \ifbbl@usedategroup % if \foreign... within \<lang>date
2787       \bbl@usedategroupfalse
2788       \ifhmode
2789         \hskip\z@skip % trick to ignore spaces
2790         \csname date#1\endcsname\relax
2791         \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
2792       \else
2793         \csname date#1\endcsname\relax
2794       \fi
2795     \fi
2796   \fi
2797   % switch extras
2798   \bbl@usehooks{beforeextras}{}%
2799   \csname extras#1\endcsname\relax
2800   \bbl@usehooks{afterextras}{}%
2801   % > babel-ensure
2802   % > babel-sh-<short>
2803   % > babel-bidi

```

```

2804 % > babel-fontspec
2805 % hyphenation - case mapping
2806 \ifcase\bbbl@opt@hyphenmap\or
2807   \def\BabelLower##1##2{\lccode##1=##2\relax}%
2808   \ifnum\bbbl@hymapsel>4\else
2809     \csname\language @\bbbl@hyphenmap\endcsname
2810   \fi
2811   \chardef\bbbl@opt@hyphenmap\z@
2812 \else
2813   \ifnum\bbbl@hymapsel>\bbbl@opt@hyphenmap\else
2814     \csname\language @\bbbl@hyphenmap\endcsname
2815   \fi
2816 \fi
2817 \global\let\bbbl@hymapsel\@cclv
2818 % hyphenation - patterns
2819 \bbbl@patterns{#1}%
2820 % hyphenation - mins
2821 \babel@savevariable\lefthyphenmin
2822 \babel@savevariable\righthyphenmin
2823 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2824   \set@hyphenmins\tw@\thr@\relax
2825 \else
2826   \expandafter\expandafter\expandafter\set@hyphenmins
2827   \csname #1hyphenmins\endcsname\relax
2828 \fi}

```

`otherlanguage` The `otherlanguage` environment can be used as an alternative to using the `\selectlanguage` declarative command. When you are typesetting a document which mixes left-to-right and right-to-left typesetting you have to use this environment in order to let things work as you expect them to. The `\ignorespaces` command is necessary to hide the environment when it is entered in horizontal mode.

```

2829 \long\def\otherlanguage#1{%
2830   \ifnum\bbbl@hymapsel=\@cclv\let\bbbl@hymapsel\thr@\fi
2831   \csname selectlanguage \endcsname{#1}%
2832   \ignorespaces}

```

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

```

2833 \long\def\endotherlanguage{%
2834   \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`.

```

2835 \expandafter\def\csname otherlanguage*\endcsname#1{%
2836   \ifnum\bbbl@hymapsel=\@cclv\chardef\bbbl@hymapsel4\relax\fi
2837   \foreign@language{#1}}

```

At the end of the environment we need to switch off the extra definitions. The grouping mechanism of the environment will take care of resetting the correct hyphenation rules and “extras”.

```

2838 \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.

```

2839 \providecommand\bbl@beforeforeign{}
2840 \edef\foreignlanguage{%
2841   \noexpand\protect
2842   \expandafter\noexpand\csname foreignlanguage \endcsname}
2843 \expandafter\def\csname foreignlanguage \endcsname{%
2844   \@ifstar\bbl@foreign@s\bbl@foreign@x}
2845 \def\bbl@foreign@x#1#2{%
2846   \begingroup
2847     \let\BabelText\@firstofone
2848     \bbl@beforeforeign
2849     \foreign@language{#1}%
2850     \bbl@usehooks{foreign}{}%
2851     \BabelText{#2}% Now in horizontal mode!
2852   \endgroup}
2853 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \setpar, ?\@@par
2854   \begingroup
2855     {\par}%
2856     \let\BabelText\@firstofone
2857     \foreign@language{#1}%
2858     \bbl@usehooks{foreign*}{}%
2859     \bbl@dirparastext
2860     \BabelText{#2}% Still in vertical mode!
2861     {\par}%
2862   \endgroup}

```

`\foreign@language` This macro does the work for `\foreignlanguage` and the `otherlanguage*` 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`.

```

2863 \def\foreign@language#1{%
2864   % set name
2865   \edef\language#1%
2866   \bbl@fixname\language
2867   \bbl@iflanguage\language%
2868     \expandafter\ifx\csname date\language\endcsname\relax
2869       \bbl@warning % TODO - why a warning, not an error?
2870       {Unknown language `#1'. Either you have\\%
2871         misspelled its name, it has not been installed,\\%

```

```

2872         or you requested it in a previous run. Fix its name,\%
2873         install it or just rerun the file, respectively. In\%
2874         some cases, you may need to remove the aux file.\%
2875         I'll proceed, but expect wrong results.\%
2876         Reported}%
2877     \fi
2878     % set type
2879     \let\bbl@select@type\@ne
2880     \expandafter\bbl@switch\expandafter{\language}%

```

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

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

```

2881 \let\bbl@hyphlist\@empty
2882 \let\bbl@hyphenation@relax
2883 \let\bbl@pttnlist\@empty
2884 \let\bbl@patterns@relax
2885 \let\bbl@hymapsel=\ccclv
2886 \def\bbl@patterns#1{%
2887     \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
2888         \csname l@#1\endcsname
2889         \edef\bbl@tempa{#1}%
2890     \else
2891         \csname l@#1:\f@encoding\endcsname
2892         \edef\bbl@tempa{#1:\f@encoding}%
2893     \fi
2894     \@expandtwoargs\bbl@usehooks{patterns}{#1}{\bbl@tempa}%
2895     % > luatex
2896     \@ifundefined{bbl@hyphenation@}{% Can be \relax!
2897         \begingroup
2898             \bbl@xin@{, \number\language,}{, \bbl@hyphlist}%
2899             \ifin@else
2900                 \@expandtwoargs\bbl@usehooks{hyphenation}{#1}{\bbl@tempa}%
2901                 \hyphenation{%
2902                     \bbl@hyphenation@
2903                     \@ifundefined{bbl@hyphenation@#1}%
2904                         \@empty
2905                     {\space\csname bbl@hyphenation@#1\endcsname}}%
2906                 \xdef\bbl@hyphlist{\bbl@hyphlist\number\language,}%
2907             \fi
2908         \endgroup}}

```

**hyphenrules** The environment `hyphenrules` can be used to select *just* the hyphenation rules. This environment does *not* change `\language` and when the hyphenation rules specified were not loaded it has no effect. Note however, `\lccode's` and font encodings are not set at all, so in most cases you should use other language\*.

```

2909 \def\hyphenrules#1{%
2910     \edef\bbl@tempf{#1}%
2911     \bbl@fixname\bbl@tempf
2912     \bbl@iflanguage\bbl@tempf{%
2913         \expandafter\bbl@patterns\expandafter{\bbl@tempf}%

```



```

2914 \languageshortands{none}%
2915 \expandafter\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
2916 \set@hyphenmins\tw@\thr@@\relax
2917 \else
2918 \expandafter\expandafter\expandafter\set@hyphenmins
2919 \csname\bbl@tempf hyphenmins\endcsname\relax
2920 \fi}}
2921 \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.

```

2922 \def\providehyphenmins#1#2{%
2923 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2924 \@namedef{#1hyphenmins}{#2}%
2925 \fi}

```

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

```

2926 \def\set@hyphenmins#1#2{%
2927 \lefthyphenmin#1\relax
2928 \righthyphenmin#2\relax}

```

`\ProvidesLanguage` The identification code for each file is something that was introduced in L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>. 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.

```

2929 \ifx\ProvidesFile\@undefined
2930 \def\ProvidesLanguage#1[#2 #3 #4]{%
2931 \wlog{Language: #1 #4 #3 <#2>}%
2932 }
2933 \else
2934 \def\ProvidesLanguage#1{%
2935 \begingroup
2936 \catcode`\ 10 %
2937 \@makeother\/%
2938 \@ifnextchar[%]
2939 {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}}
2940 \def\@provideslanguage#1[#2]{%
2941 \wlog{Language: #1 #2}%
2942 \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
2943 \endgroup}
2944 \fi

```

`\LdfInit` This macro is defined in two versions. The first version is to be part of the ‘kernel’ of babel, ie. the part that is loaded in the format; the second version is defined in `babel.def`. The version in the format just checks the category code of the ampersand and then loads `babel.def`.

The category code of the ampersand is restored and the macro calls itself again with the new definition from `babel.def`

```

2945 \def\LdfInit{%
2946 \chardef\atcatcode=\catcode`\@
2947 \catcode`\@=11\relax
2948 \input babel.def\relax
2949 \catcode`\@=\atcatcode \let\atcatcode\relax
2950 \LdfInit}

```

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

```
2951 \ifx\originalTeX\@undefined\let\originalTeX\@empty\fi
```

Because this part of the code can be included in a format, we make sure that the macro which initialises the save mechanism, `\babel@beginsave`, is not considered to be undefined.

```
2952 \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’:

```
2953 \providecommand\setlocale{%
2954   \bbl@error
2955   {Not yet available}%
2956   {Find an armchair, sit down and wait}}
2957 \let\uselocale\setlocale
2958 \let\locale\setlocale
2959 \let\selectlocale\setlocale
2960 \let\textlocale\setlocale
2961 \let\textlanguage\setlocale
2962 \let\languagetext\setlocale
```

## 11.2 Errors

`\@nolanerr` The babel package will signal an error when a documents tries to select a language that hasn’t been defined earlier. When a user selects a language for which no hyphenation patterns were loaded into the format he will be given a warning about that fact. We revert to the patterns for `\language=0` in that case. In most formats that will be (US)english, but it might also be empty.

`\@noopterr` When the package was loaded without options not everything will work as expected. An error message is issued in that case.  
When the format knows about `\PackageError` it must be  $\LaTeX 2_{\epsilon}$ , so we can safely use its error handling interface. Otherwise we’ll have to ‘keep it simple’.

```
2963 \edef\bbl@nulllanguage{\string\language=0}
2964 \ifx\PackageError\@undefined
2965   \def\bbl@error#1#2{%
2966     \begingroup
2967       \newlinechar=`^^J
2968       \def\{^^J(babel) }%
2969       \errhelp{#2}\errmessage{\{#1}%
2970     \endgroup}
2971   \def\bbl@warning#1{%
2972     \begingroup
2973       \newlinechar=`^^J
2974       \def\{^^J(babel) }%
2975       \message{\{#1}%
2976     \endgroup}
2977   \def\bbl@info#1{%
2978     \begingroup
2979       \newlinechar=`^^J
2980       \def\{^^J}%
2981       \wlog{#1}%
2982     \endgroup}
2983 \else
2984   \def\bbl@error#1#2{%
2985     \begingroup
```

```

2986     \def\{\MessageBreak}%
2987     \PackageError{babel}{#1}{#2}%
2988   \endgroup}
2989   \def\bbl@warning#1{%
2990     \begingroup
2991       \def\{\MessageBreak}%
2992       \PackageWarning{babel}{#1}%
2993     \endgroup}
2994   \def\bbl@info#1{%
2995     \begingroup
2996       \def\{\MessageBreak}%
2997       \PackageInfo{babel}{#1}%
2998     \endgroup}
2999 \fi
3000 \@ifpackagewith{babel}{silent}
3001   {\let\bbl@info@gobble
3002    \let\bbl@warning@gobble}
3003   {}
3004 \def\bbl@nocaption{\protect\bbl@nocaption@i}
3005 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
3006   \global\@namedef{#2}{\textbf{?#1?}}%
3007   \@nameuse{#2}%
3008   \bbl@warning{%
3009     \@backslashchar#2 not set. Please, define\\%
3010     it in the preamble with something like:\\%
3011     \string\renewcommand\@backslashchar#2{..}\\%
3012     Reported}}
3013 \def\bbl@tentative{\protect\bbl@tentative@i}
3014 \def\bbl@tentative@i#1{%
3015   \bbl@warning{%
3016     Some functions for '#1' are tentative.\\%
3017     They might not work as expected and their behavior\\%
3018     could change in the future.\\%
3019     Reported}}
3020 \def\@nolanerr#1{%
3021   \bbl@error
3022   {You haven't defined the language #1\space yet}%
3023   {Your command will be ignored, type <return> to proceed}}
3024 \def\@nopatterns#1{%
3025   \bbl@warning
3026   {No hyphenation patterns were preloaded for\\%
3027    the language '#1' into the format.\\%
3028    Please, configure your TeX system to add them and\\%
3029    rebuild the format. Now I will use the patterns\\%
3030    preloaded for \bbl@nulllanguage\space instead}}
3031 \let\bbl@usehooks@gobbletwo
3032 \</kernel>
3033 \<*/patterns>

```

## 12 Loading hyphenation patterns

The following code is meant to be read by  $\text{\LaTeX}$  because it should instruct  $\text{\TeX}$  to read hyphenation patterns. To this end the `docstrip` option `patterns` can be used to include this code in the file `hyphen.cfg`. Code is written with lower level macros.

We want to add a message to the message  $\text{\LaTeX}$  2.09 puts in the `\everyjob` register. This could be done by the following code:

```

\let\orgeveryjob\everyjob
\def\everyjob#1{%
  \orgeveryjob{#1}%
  \orgeveryjob\expandafter{\the\orgeveryjob\immediate\write16{%
    hyphenation patterns for \the\loaded@patterns loaded.}}%
\let\everyjob\orgeveryjob\let\orgeveryjob\@undefined}

```

The code above redefines the control sequence `\everyjob` in order to be able to add something to the current contents of the register. This is necessary because the processing of hyphenation patterns happens long before  $\LaTeX$  fills the register. There are some problems with this approach though.

- When someone wants to use several hyphenation patterns with  $\LaTeX$  the above scheme won't work. The reason is that  $\LaTeX$  overwrites the contents of the `\everyjob` register with its own message.
- Plain  $\TeX$  does not use the `\everyjob` register so the message would not be displayed.

To circumvent this a 'dirty trick' can be used. As this code is only processed when creating a new format file there is one command that is sure to be used, `\dump`. Therefore the original `\dump` is saved in `\org@dump` and a new definition is supplied. To make sure that  $\LaTeX$  2.09 executes the `\@begindocumenthook` we would want to alter `\begin{document}`, but as this done too often already, we add the new code at the front of `\@preamblecmds`. But we can only do that after it has been defined, so we add this piece of code to `\dump`.

This new definition starts by adding an instruction to write a message on the terminal and in the transcript file to inform the user of the preloaded hyphenation patterns. Then everything is restored to the old situation and the format is dumped.

```

3034 <<Make sure ProvidesFile is defined>>
3035 \ProvidesFile{hyphen.cfg}[\<date>] [\<version>] Babel hyphens]
3036 \xdef\bb1@format{\jobname}
3037 \ifx\AtBeginDocument\@undefined
3038   \def\@empty{}
3039   \let\orig@dump\dump
3040   \def\dump{%
3041     \ifx\@ztryfc\@undefined
3042     \else
3043       \toks0=\expandafter{\@preamblecmds}%
3044       \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
3045       \def\@begindocumenthook{}%
3046     \fi
3047     \let\dump\orig@dump\let\orig@dump\@undefined\dump}
3048 \fi
3049 <<Define core switching macros>>

```

`\process@line` Each line in the file `language.dat` is processed by `\process@line` after it is read. The first thing this macro does is to check whether the line starts with `=`. When the first token of a line is an `=`, the macro `\process@synonym` is called; otherwise the macro `\process@language` will continue.

```

3050 \def\process@line#1#2 #3 #4 {%
3051   \ifx=#1%
3052     \process@synonym{#2}%
3053   \else
3054     \process@language{#1#2}{#3}{#4}%
3055   \fi
3056   \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.

```
3057 \toks@{}
3058 \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.

```
3059 \def\process@synonym#1{%
3060   \ifnum\last@language=\m@ne
3061     \toks@expandafter{\the\toks@\relax\process@synonym{#1}}%
3062   \else
3063     \expandafter\chardef\csname l@#1\endcsname\last@language
3064     \wlog{\string\l@#1=\string\language\the\last@language}%
3065     \expandafter\let\csname #1hyphenmins\expandafter\endcsname
3066       \csname\language\hyphenmins\endcsname
3067     \let\bbl@elt\relax
3068     \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}}}%
3069   \fi}
```

`\process@language` The macro `\process@language` is used to process a non-empty line from the ‘configuration file’. It has three arguments, each delimited by white space. The first argument is the ‘name’ of a language; the second is the name of the file that contains the patterns. The optional third argument is the name of a file containing hyphenation exceptions. The first thing to do is call `\addlanguage` to allocate a pattern register and to make that register ‘active’. Then the pattern file is read.

For some hyphenation patterns it is needed to load them with a specific font encoding selected. This can be specified in the file `language.dat` by adding for instance ‘:T1’ to the name of the language. The macro `\bbl@get@enc` extracts the font encoding from the language name and stores it in `\bbl@hyph@enc`. The latter can be used in hyphenation files if you need to set a behavior depending on the given encoding (it is set to empty if no encoding is given).

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

Some pattern files contain changes to the `\lccode` en `\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.

```
3070 \def\process@language#1#2#3{%
3071   \expandafter\addlanguage\csname l@#1\endcsname
```

```

3072 \expandafter\language\csname l@#1\endcsname
3073 \edef\languagename{#1}%
3074 \bbl@hook@everylanguage{#1}%
3075 % > luatex
3076 \bbl@get@enc#1::\@@@
3077 \beginingroup
3078 \lefthyphenmin\m@ne
3079 \bbl@hook@loadpatterns{#2}%
3080 % > luatex
3081 \ifnum\lefthyphenmin=\m@ne
3082 \else
3083 \expandafter\xdef\csname #1hyphenmins\endcsname{%
3084 \the\lefthyphenmin\the\righthyphenmin}%
3085 \fi
3086 \endgroup
3087 \def\bbl@tempa{#3}%
3088 \ifx\bbl@tempa\@empty\else
3089 \bbl@hook@loadexceptions{#3}%
3090 % > luatex
3091 \fi
3092 \let\bbl@elt\relax
3093 \edef\bbl@languages{%
3094 \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
3095 \ifnum\the\language=\z@
3096 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
3097 \set@hyphenmins\tw@\thr@@\relax
3098 \else
3099 \expandafter\expandafter\expandafter\set@hyphenmins
3100 \csname #1hyphenmins\endcsname
3101 \fi
3102 \the\toks@
3103 \toks@{}%
3104 \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.

```

3105 \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.

```

3106 \def\bbl@hook@everylanguage#1{}
3107 \def\bbl@hook@loadpatterns#1{\input #1\relax}
3108 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
3109 \let\bbl@hook@loadkernel\bbl@hook@loadpatterns
3110 \beginingroup
3111 \def\AddBabelHook#1#2{%
3112 \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
3113 \def\next{\toks1}%
3114 \else
3115 \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname###1}%
3116 \fi
3117 \next}
3118 \ifx\directlua\@undefined
3119 \ifx\XeTeXinputencoding\@undefined\else
3120 \input xebabel.def
3121 \fi
3122 \else
3123 \input luababel.def

```

```

3124 \fi
3125 \openin1 = babel-\bbl@format.cfg
3126 \ifeof1
3127 \else
3128 \input babel-\bbl@format.cfg\relax
3129 \fi
3130 \closein1
3131 \endgroup
3132 \bbl@hook@loadkernel{switch.def}

```

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

```

3133 \openin1 = language.dat

```

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

```

3134 \def\languagename{english}%
3135 \ifeof1
3136 \message{I couldn't find the file language.dat,\space
3137          I will try the file hyphen.tex}
3138 \input hyphen.tex\relax
3139 \chardef\l@english\z@
3140 \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`.

```

3141 \last@language\m@ne

```

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

```

3142 \loop

```

While reading from the input, it is useful to switch off recognition of the end-of-line character. This saves us stripping off spaces from the contents of the control sequence.

```

3143 \endlinechar\m@ne
3144 \read1 to \bbl@line
3145 \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.

```

3146 \if T\ifeof1F\fi T\relax
3147 \ifx\bbl@line\@empty\else
3148 \edef\bbl@line{\bbl@line\space\space\space}%
3149 \expandafter\process@line\bbl@line\relax
3150 \fi
3151 \repeat

```

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

```

3152 \begingroup
3153 \def\bbl@elt#1#2#3#4{%
3154 \global\language=#2\relax
3155 \gdef\languagename{#1}%
3156 \def\bbl@elt##1##2##3##4{}}%
3157 \bbl@languages
3158 \endgroup
3159 \fi

```

and close the configuration file.

```
3160 \closein1
```

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

```
3161 \if/\the\toks@/\else
3162   \errhelp{language.dat loads no language, only synonyms}
3163   \errmessage{Orphan language synonym}
3164 \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.

```
3165 \let\bbl@line\@undefined
3166 \let\process@line\@undefined
3167 \let\process@synonym\@undefined
3168 \let\process@language\@undefined
3169 \let\bbl@get@enc\@undefined
3170 \let\bbl@hyph@enc\@undefined
3171 \let\bbl@tempa\@undefined
3172 \let\bbl@hook@loadkernel\@undefined
3173 \let\bbl@hook@everylanguage\@undefined
3174 \let\bbl@hook@loadpatterns\@undefined
3175 \let\bbl@hook@loadexceptions\@undefined
3176 \</patterns>
```

Here the code for `iniTEX` ends.

## 13 Font handling with fontspec

Add the bidi handler just before `luaoftload`, 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].

```
3177 <<(*More package options)>> ≡
3178 \ifodd\bbl@engine
3179   \DeclareOption{bidi=basic-r}%
3180   {\ExecuteOptions{bidi=basic}}
3181   \DeclareOption{bidi=basic}%
3182   {\let\bbl@beforeforeign\leavevmode
3183    % TODO - to locale_props, not as separate attribute
3184    \newattribute\bbl@attr@dir
3185    % I don't like it, hackish:
3186    \frozen@everymath\expandafter{%
3187      \expandafter\bbl@mathboxdir\the\frozen@everymath}%
3188    \frozen@everydisplay\expandafter{%
3189      \expandafter\bbl@mathboxdir\the\frozen@everydisplay}%
3190    \bbl@exp{\output{\bodydir\pagedir\the\output}}%
3191    \AtEndOfPackage{\EnableBabelHook{babel-bidi}}}
3192 \else
3193   \DeclareOption{bidi=basic-r}%
3194   {\ExecuteOptions{bidi=basic}}
3195   \DeclareOption{bidi=basic}%
3196   {\bbl@error
3197    {The bidi method 'basic' is available only in\\%
3198     luatex. I'll continue with 'bidi=default', so\\%
3199     expect wrong results}%
3200    {See the manual for further details.}%}
```



```

3201 \let\bbl@beforeforeign\leavevmode
3202 \AtEndOfPackage{%
3203   \EnableBabelHook{babel-bidi}%
3204   \bbl@xebidipar}}
3205 \def\bbl@loadxebidi#1{%
3206   \ifx\RTLfootnotetext\@undefined
3207     \AtEndOfPackage{%
3208       \EnableBabelHook{babel-bidi}%
3209       \ifx\fontspec\@undefined
3210         \usepackage{fontspec}% bidi needs fontspec
3211       \fi
3212       \usepackage#1{bidi}}%
3213   \fi}
3214 \DeclareOption{bidi=bidi}%
3215   {\bbl@tentative{bidi=bidi}%
3216     \bbl@loadxebidi{}}
3217 \DeclareOption{bidi=bidi-r}%
3218   {\bbl@tentative{bidi=bidi-r}%
3219     \bbl@loadxebidi{[rldocument]}}
3220 \DeclareOption{bidi=bidi-l}%
3221   {\bbl@tentative{bidi=bidi-l}%
3222     \bbl@loadxebidi{}}
3223 \fi
3224 \DeclareOption{bidi=default}%
3225   {\let\bbl@beforeforeign\leavevmode
3226     \ifodd\bbl@engine
3227       \newattribute\bbl@attr@dir
3228       \bbl@exp{\output{\bodydir\pagedir\the\output}}%
3229     \fi
3230     \AtEndOfPackage{%
3231       \EnableBabelHook{babel-bidi}%
3232       \ifodd\bbl@engine\else
3233         \bbl@xebidipar
3234       \fi}}
3235 <</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.

```

3236 <<{*Font selection}>> ≡
3237 \bbl@trace{Font handling with fontspec}
3238 \@onlypreamble\babelfont
3239 \newcommand\babelfont[2][{}]{% 1=langs/scripts 2=fam
3240   \edef\bbl@tempa{#1}%
3241   \def\bbl@tempb{#2}%
3242   \ifx\fontspec\@undefined
3243     \usepackage{fontspec}%
3244   \fi
3245   \EnableBabelHook{babel-fontspec}% Just calls \bbl@switchfont
3246   \bbl@bblfont}
3247 \newcommand\bbl@bblfont[2][{}]{% 1=features 2=fontname
3248   \bbl@ifunset{\bbl@tempb family}{\bbl@providfam{\bbl@tempb}}{}}%
3249 % For the default font, just in case:
3250 \bbl@ifunset{\bbl@lsys\languagename}{\bbl@provide@lsys{\languagename}}{}}%
3251 \expandafter\bbl@ifblank\expandafter{\bbl@tempa}%
3252   {\bbl@csarg\edef{\bbl@tempb dflt@}{<{#1}{#2}}% save bbl@rmdflt@
3253   \bbl@exp{%
3254     \let<\bbl@\bbl@tempb dflt@\languagename>\<\bbl@\bbl@tempb dflt@>%
3255     \\\bbl@font@set<\bbl@\bbl@tempb dflt@\languagename>%
3256     \<\bbl@tempb default>\<\bbl@tempb family>}}%

```

```

3257 {\bbl@foreach\bbl@tempa{% ie bbl@rmdflt@lang / *scrt
3258 \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:

```

3259 \def\bbl@providfam#1{%
3260 \bbl@exp{%
3261 \\\newcommand\<#1default>{% Just define it
3262 \\\bbl@add@list\\bbl@font@fams{#1}%
3263 \\\DeclareRobustCommand\<#1family>{%
3264 \\\not@math@alphabet\<#1family>\relax
3265 \\\fontfamily\<#1default>\\selectfont}%
3266 \\\DeclareTextFontCommand{\<text#1>}{\<#1family>}}

```

The following macro is activated when the hook babel-fontspec is enabled.

```

3267 \def\bbl@switchfont{%
3268 \bbl@ifunset{\bbl@lsys@\language}\bbl@provide@lsys{\language}}}%
3269 \bbl@exp{% eg Arabic -> arabic
3270 \lowercase{\edef\\bbl@tempa{\bbl@cs{sname@\language}}}}%
3271 \bbl@foreach\bbl@font@fams{%
3272 \bbl@ifunset{\bbl@##1dflt@\language}% (1) language?
3273 {\bbl@ifunset{\bbl@##1dflt@*\bbl@tempa}% (2) from script?
3274 {\bbl@ifunset{\bbl@##1dflt@}% 2=F - (3) from generic?
3275 {}% 123=F - nothing!
3276 {\bbl@exp{% 3=T - from generic
3277 \global\let\<bbl@##1dflt@\language>%
3278 \<bbl@##1dflt@>}}}%
3279 {\bbl@exp{% 2=T - from script
3280 \global\let\<bbl@##1dflt@\language>%
3281 \<bbl@##1dflt@*\bbl@tempa>}}}%
3282 {}% 1=T - language, already defined
3283 \def\bbl@tempa{%
3284 \bbl@warning{The current font is not a standard family:\\%
3285 \fontname\font\\%
3286 Script and Language are not applied. Consider\\%
3287 defining a new family with \string\babelfont.\\%
3288 Reported}}}%
3289 \bbl@foreach\bbl@font@fams{% don't gather with prev for
3290 \bbl@ifunset{\bbl@##1dflt@\language}%
3291 {\bbl@cs{famrst@##1}%
3292 \global\bbl@csarg\let{famrst@##1}\relax}%
3293 {\bbl@exp{% order is relevant
3294 \\\bbl@add\\originalTeX{%
3295 \\\bbl@font@rst{\bbl@cs{##1dflt@\language}}%
3296 \<##1default>\<##1family>{##1}}}%
3297 \\\bbl@font@set\<bbl@##1dflt@\language>% the main part!
3298 \<##1default>\<##1family>}}}%
3299 \bbl@ifrestoring{{}\bbl@tempa}}%

```

Now the macros defining the font with fontspec.

When there are repeated keys in fontspec, the last value wins. So, we just place the ini settings at the beginning, and user settings will take precedence. We must deactivate temporarily \bbl@mapselect because \selectfont is called internally when a font is defined.

```

3300 \def\bbl@font@set#1#2#3{% eg \bbl@rmdflt@lang \rmdefault \rmfamily
3301 \bbl@xin@{<>}{#1}%
3302 \ifin@
3303 \bbl@exp{\\\bbl@fontspec@set\\#1\expandafter@gobbletwo#1\\#3}%
3304 \fi
3305 \bbl@exp{%

```

```

3306 \def\#2{#1}% eg, \rmdefault{\bbl@rmdflt@lang}
3307 \bbl@ifsamestring{#2}{\f@family}{\#3\let\bbl@tempa\relax}{}}
3308 % TODO - next should be global?, but even local does its job. I'm
3309 % still not sure -- must investigate:
3310 \def\bbl@fontspec@set#1#2#3#4{% eg \bbl@rmdflt@lang fnt-opt fnt-nme \xxfamily
3311 \let\bbl@tempe\bbl@mapselect
3312 \let\bbl@mapselect\relax
3313 \let\bbl@temp@fam#4% eg, '\rmfamily', to be restored below
3314 \let#4\relax % So that can be used with \newfontfamily
3315 \bbl@exp{%
3316 \let\bbl@temp@pfam\<\bbl@stripslash#4\space>% eg, '\rmfamily '
3317 \<keys_if_exist:nnF>{\fontspec-opentype}%
3318 {Script/\bbl@cs{sname@\language}}}%
3319 {\newfontscript{\bbl@cs{sname@\language}}}%
3320 {\bbl@cs{sotf@\language}}}%
3321 \<keys_if_exist:nnF>{\fontspec-opentype}%
3322 {Language/\bbl@cs{lname@\language}}}%
3323 {\newfontlanguage{\bbl@cs{lname@\language}}}%
3324 {\bbl@cs{lotf@\language}}}%
3325 \newfontfamily\#4%
3326 [\bbl@cs{lsys@\language},#2]{#3}% ie \bbl@exp{.}{#3}
3327 \begingroup
3328 #4%
3329 \xdef#1{\f@family}% eg, \bbl@rmdflt@lang{FreeSerif(0)}
3330 \endgroup
3331 \let#4\bbl@temp@fam
3332 \bbl@exp{\let\<\bbl@stripslash#4\space>\bbl@temp@pfam
3333 \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.

```

3334 \def\bbl@font@rst#1#2#3#4{%
3335 \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.

```

3336 \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 :-).

```

3337 \newcommand\babelFSstore[2][{%
3338 \bbl@ifblank{#1}%
3339 {\bbl@csarg\def{sname@#2}{Latin}}%
3340 {\bbl@csarg\def{sname@#2}{#1}}%
3341 \bbl@provide@dirs{#2}%
3342 \bbl@csarg\ifnum{wdir@#2}>\z@
3343 \let\bbl@beforeforeign\leavevmode
3344 \EnableBabelHook{babel-bidi}%
3345 \fi
3346 \bbl@foreach{#2}{%
3347 \bbl@FSstore{##1}{rm}\rmdefault\bbl@save\rmdefault
3348 \bbl@FSstore{##1}{sf}\sfdefault\bbl@save\sfdefault
3349 \bbl@FSstore{##1}{tt}\ttdefault\bbl@save\TTdefault}}
3350 \def\bbl@FSstore#1#2#3#4{%
3351 \bbl@csarg\edef{#2default#1}{#3}%
3352 \expandafter\addto\csname extras#1\endcsname{%
3353 \let#4#3%
3354 \ifx#3\f@family

```

```

3355     \edef#3{\csname bbl@#2default#1\endcsname}%
3356     \fontfamily{#3}\selectfont
3357   \else
3358     \edef#3{\csname bbl@#2default#1\endcsname}%
3359     \fi}%
3360   \expandafter\addto\csname noextras#1\endcsname{%
3361     \ifx#3\f@family
3362       \fontfamily{#4}\selectfont
3363       \fi
3364     \let#3#4}}
3365 \let\bbl@langfeatures\@empty
3366 \def\babelFSfeatures{% make sure \fontspec is redefined once
3367   \let\bbl@ori@fontspec\fontspec
3368   \renewcommand\fontspec[1][{}]{%
3369     \bbl@ori@fontspec[\bbl@langfeatures##1]}
3370   \let\babelFSfeatures\bbl@FSfeatures
3371   \babelFSfeatures}
3372 \def\bbl@FSfeatures#1#2{%
3373   \expandafter\addto\csname extras#1\endcsname{%
3374     \babel@save\bbl@langfeatures
3375     \edef\bbl@langfeatures{#2,}}
3376   <</Font selection>>

```

## 14 Hooks for XeTeX and LuaTeX

### 14.1 XeTeX

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

$\LaTeX$  sets many “codes” just before loading `hyphen.cfg`. That is not a problem in `luatex`, but in `xetex` they must be reset to the proper value. Most of the work is done in `xe(la)tex.ini`, so here we just “undo” some of the changes done by  $\LaTeX$ . Anyway, for consistency `LuaTeX` also resets the catcodes.

```

3377 <<(*Restore Unicode catcodes before loading patterns)>> ≡
3378   \begingroup
3379     % Reset chars "80-"C0 to category "other", no case mapping:
3380     \catcode`\@=11 \count@=128
3381     \loop\ifnum\count@<192
3382       \global\uccode\count@=0 \global\lccode\count@=0
3383       \global\catcode\count@=12 \global\sffcode\count@=1000
3384       \advance\count@ by 1 \repeat
3385     % Other:
3386     \def\O ##1 {%
3387       \global\uccode"##1=0 \global\lccode"##1=0
3388       \global\catcode"##1=12 \global\sffcode"##1=1000 }%
3389     % Letter:
3390     \def\L ##1 ##2 ##3 {\global\catcode"##1=11
3391       \global\uccode"##1="##2
3392       \global\lccode"##1="##3
3393       % Uppercase letters have sffcode=999:
3394       \ifnum"##1="##3 \else \global\sffcode"##1=999 \fi }%
3395     % Letter without case mappings:
3396     \def\l ##1 {\L ##1 ##1 ##1 }%
3397     \l 00AA
3398     \L 00B5 039C 00B5
3399     \l 00BA
3400     \O 00D7

```

```

3401 \l 00DF
3402 \O 00F7
3403 \L 00FF 0178 00FF
3404 \endgroup
3405 \input #1\relax
3406 <</Restore Unicode catcodes before loading patterns>>

```

Some more common code.

```

3407 <<{*Footnote changes}>> ≡
3408 \bbl@trace{Bidi footnotes}
3409 \ifx\bbl@beforeforeign\leavevmode
3410 \def\bbl@footnote#1#2#3{%
3411 \@ifnextchar[%
3412 {\bbl@footnote@o{#1}{#2}{#3}}%
3413 {\bbl@footnote@x{#1}{#2}{#3}}}
3414 \def\bbl@footnote@x#1#2#3#4{%
3415 \bgroup
3416 \select@language@x{\bbl@main@language}%
3417 \bbl@fn@footnote{#2#1{\ignorespaces#4}#3}%
3418 \egroup}
3419 \def\bbl@footnote@o#1#2#3[#4]#5{%
3420 \bgroup
3421 \select@language@x{\bbl@main@language}%
3422 \bbl@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}%
3423 \egroup}
3424 \def\bbl@footnotetext#1#2#3{%
3425 \@ifnextchar[%
3426 {\bbl@footnotetext@o{#1}{#2}{#3}}%
3427 {\bbl@footnotetext@x{#1}{#2}{#3}}}
3428 \def\bbl@footnotetext@x#1#2#3#4{%
3429 \bgroup
3430 \select@language@x{\bbl@main@language}%
3431 \bbl@fn@footnotetext{#2#1{\ignorespaces#4}#3}%
3432 \egroup}
3433 \def\bbl@footnotetext@o#1#2#3[#4]#5{%
3434 \bgroup
3435 \select@language@x{\bbl@main@language}%
3436 \bbl@fn@footnotetext[#4]{#2#1{\ignorespaces#5}#3}%
3437 \egroup}
3438 \def\BabelFootnote#1#2#3#4{%
3439 \ifx\bbl@fn@footnote\undefined
3440 \let\bbl@fn@footnote\footnote
3441 \fi
3442 \ifx\bbl@fn@footnotetext\undefined
3443 \let\bbl@fn@footnotetext\footnotetext
3444 \fi
3445 \bbl@ifblank{#2}%
3446 {\def#1{\bbl@footnote{\@firstofone}{#3}{#4}}
3447 \@namedef{\bbl@stripslash#1text}%
3448 {\bbl@footnotetext{\@firstofone}{#3}{#4}}}%
3449 {\def#1{\bbl@exp{\bbl@footnote{\foreignlanguage{#2}}}{#3}{#4}}%
3450 \@namedef{\bbl@stripslash#1text}%
3451 {\bbl@exp{\bbl@footnotetext{\foreignlanguage{#2}}}{#3}{#4}}}%
3452 \fi
3453 <</Footnote changes>>

```

Now, the code.

```

3454 <{*xetex}>
3455 \def\BabelStringsDefault{unicode}

```

```

3456 \let\xebbl@stop\relax
3457 \AddBabelHook{xetex}{encodedcommands}{%
3458   \def\bbl@tempa{#1}%
3459   \ifx\bbl@tempa\@empty
3460     \XeTeXinputencoding"bytes"%
3461   \else
3462     \XeTeXinputencoding"#1"%
3463   \fi
3464   \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
3465 \AddBabelHook{xetex}{stopcommands}{%
3466   \xebbl@stop
3467   \let\xebbl@stop\relax}
3468 \def\bbl@intraspace#1 #2 #3\@@{%
3469   \bbl@csarg\gdef\xeisp@bbl@cs{sbcpr@languagename}}%
3470   {\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}}
3471 \def\bbl@intrapenalty#1\@@{%
3472   \bbl@csarg\gdef\xeipn@bbl@cs{sbcpr@languagename}}%
3473   {\XeTeXlinebreakpenalty #1\relax}}
3474 \AddBabelHook{xetex}{loadkernel}{%
3475   <<Restore Unicode catcodes before loading patterns>>}}
3476 \ifx\DisableBabelHook\undefined\endinput\fi
3477 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
3478 \DisableBabelHook{babel-fontspec}
3479 <<Font selection>>
3480 \input txtbabel.def
3481 </xetex>

```

## 14.2 Layout

*In progress.*

Note elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titleps, 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 pdfTeX and xetex.

```

3482 <*texxet>
3483 \bbl@trace{Redefinitions for bidi layout}
3484 \def\bbl@sspre@caption{%
3485   \bbl@exp{\everyhbox{\bbl@textdir\bbl@cs{wdir@bbl@main@language}}}}
3486 \ifx\bbl@opt@layout\@nnil\endinput\fi % No layout
3487 \def\bbl@startskip{\ifcase\bbl@thepardir\leftskip\else\rightskip\fi}
3488 \def\bbl@endskip{\ifcase\bbl@thepardir\rightskip\else\leftskip\fi}
3489 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
3490   \def\@hangfrom#1{%
3491     \setbox\@tempboxa\hbox{#1}%
3492     \hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
3493     \noindent\box\@tempboxa}
3494 \def\raggedright{%
3495   \let\@centercr
3496   \bbl@startskip\z@skip
3497   \@rightskip\@flushglue
3498   \bbl@endskip\@rightskip
3499   \parindent\z@
3500   \parfillskip\bbl@startskip}
3501 \def\raggedleft{%

```

```

3502 \let\\\@centercr
3503 \bbl@startskip\@flushglue
3504 \bbl@endskip\z@skip
3505 \parindent\z@
3506 \parfillskip\bbl@endskip}
3507 \fi
3508 \IfBabelLayout{lists}
3509 {\def\list#1#2{%
3510 \ifnum \@listdepth >5\relax
3511 \@toodeep
3512 \else
3513 \global\advance\@listdepth\@ne
3514 \fi
3515 \rightmargin\z@
3516 \listparindent\z@
3517 \itemindent\z@
3518 \csname @list\romannumeral\the\@listdepth\endcsname
3519 \def\@itemlabel{#1}%
3520 \let\makelabel\@mklab
3521 \@nmbrlistfalse
3522 #2\relax
3523 \@trivlist
3524 \parskip\parsep
3525 \parindent\listparindent
3526 \advance\linewidth-\rightmargin
3527 \advance\linewidth-\leftmargin
3528 \advance\@totalleftmargin
3529 \ifcase\bbl@thepardir\leftmargin\else\rightmargin\fi
3530 \parshape\@ne\@totalleftmargin\linewidth
3531 \ignorespaces}%
3532 \ifcase\bbl@engine
3533 \def\labelenumii{}\thenumii}%
3534 \def\p@enumiii{\p@enumii}\thenumii}%
3535 \fi
3536 \def\@verbatim{%
3537 \trivlist \item\relax
3538 \if@minipage\else\vskip\parskip\fi
3539 \bbl@startskip\textwidth
3540 \advance\bbl@startskip-\linewidth
3541 \bbl@endskip\z@skip
3542 \parindent\z@
3543 \parfillskip\@flushglue
3544 \parskip\z@skip
3545 \@@par
3546 \language\l@nohyphenation
3547 \@tempwafalse
3548 \def\par{%
3549 \if@tempwa
3550 \leavevmode\null
3551 \@@par\penalty\interlinepenalty
3552 \else
3553 \@tempwattrue
3554 \ifhmode\@@par\penalty\interlinepenalty\fi
3555 \fi}%
3556 \let\do\@makeother \dospecials
3557 \obeylines \verbatim@font \@noligs
3558 \everypar\expandafter{\the\everypar\unpenalty}}
3559 {}
3560 \IfBabelLayout{contents}

```

```

3561 {\def\dottedtocline#1#2#3#4#5{%
3562   \ifnum#1>\c@tocdepth\else
3563     \vskip \z@ \@plus.2\p@
3564     {\bbl@startskip#2\relax
3565      \bbl@endskip\@tocrmarg
3566      \parfillskip-\bbl@endskip
3567      \parindent#2\relax
3568      \@afterindenttrue
3569      \interlinepenalty\M
3570      \leavevmode
3571      \@tempdima#3\relax
3572      \advance\bbl@startskip\@tempdima
3573      \null\nobreak\hskip-\bbl@startskip
3574      {#4}\nobreak
3575      \leaders\hbox{%
3576        $\m@th\mkern\@dotsep mu\hbox{.}\mkern\@dotsep mu$}%
3577      \hfill\nobreak
3578      \hb@xt@\@pnumwidth{\hfil\normalfont\normalcolor#5}%
3579      \par}%
3580   \fi}}
3581 {}
3582 \IfBabelLayout{columns}
3583 {\def\@outputdblcol{%
3584   \if@firstcolumn
3585     \global\@firstcolumnfalse
3586     \global\setbox\@leftcolumn\copy\@outputbox
3587     \splitmaxdepth\maxdimen
3588     \vbadness\maxdimen
3589     \setbox\@outputbox\vbox{\unvbox\@outputbox\unskip}%
3590     \setbox\@outputbox\vsplit\@outputbox to\maxdimen
3591     \toks@ \expandafter{\topmark}%
3592     \xdef\@firstcoltopmark{\the\toks@}%
3593     \toks@ \expandafter{\splitfirstmark}%
3594     \xdef\@firstcolfirstmark{\the\toks@}%
3595     \ifx\@firstcolfirstmark\@empty
3596       \global\let\@setmarks\relax
3597     \else
3598       \gdef\@setmarks{%
3599         \let\firstmark\@firstcolfirstmark
3600         \let\topmark\@firstcoltopmark}%
3601     \fi
3602   \else
3603     \global\@firstcolumntrue
3604     \setbox\@outputbox\vbox{%
3605       \hb@xt@\textwidth{%
3606         \hskip\columnwidth
3607         \hfil
3608         {\normalcolor\vrule \@width\columnseprule}%
3609         \hfil
3610         \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
3611         \hskip-\textwidth
3612         \hb@xt@\columnwidth{\box\@outputbox \hss}%
3613         \hskip\columnsep
3614         \hskip\columnwidth}}}%
3615     \@combinedblfloats
3616     \@setmarks
3617     \@outputpage
3618     \begingroup
3619     \@dblfloatplacement

```



```

3620      \@startdblcolumn
3621      \@whiles\if@fcolmade \fi{\@outputpage
3622      \@startdblcolumn}%
3623      \endgroup
3624      \fi}}%
3625  {}
3626 <<Footnote changes>>
3627 \IfBabelLayout{footnotes}%
3628  {\BabelFootnote\footnote\language\language{}{}}%
3629  \BabelFootnote\localfootnote\language\language{}{}}%
3630  \BabelFootnote\mainfootnote{}{}{}}
3631 {}

```

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.

```

3632 \IfBabelLayout{counters}%
3633  {\let\bbl@latinarabic=\@arabic
3634   \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
3635   \let\bbl@asciroman=\@roman
3636   \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
3637   \let\bbl@asciiRoman=\@Roman
3638   \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}}{}
3639 </texet>

```

### 14.3 LuaTeX

The new loader for luatex is based solely on `language.dat`, which is read on the fly. The code shouldn't be executed when the format is build, so we check if `\AddBabelHook` is defined. Then comes a modified version of the loader in `hyphen.cfg` (without the `hyphenmins` stuff, which is under the direct control of `babel`).

The names `\l@<language>` are defined and take some value from the beginning because all `ldf` files assume this for the corresponding language to be considered valid, but patterns are not loaded (except the first one). This is done later, when the language is first selected (which usually means when the `ldf` finishes). If a language has been loaded, `\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`). 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.

```

3640 (*luatex)
3641 \ifx\AddBabelHook\undefined
3642 \bbl@trace{Read language.dat}
3643 \begingroup
3644 \toks@{}
3645 \count@ \z@ % 0=start, 1=0th, 2=normal
3646 \def\bbl@process@line#1#2 #3 #4 {%
3647   \ifx=#1%
3648     \bbl@process@synonym{#2}%
3649   \else
3650     \bbl@process@language{#1#2}{#3}{#4}%
3651   \fi
3652   \ignorespaces}
3653 \def\bbl@manylang{%
3654   \ifnum\bbl@last>\@ne
3655     \bbl@info{Non-standard hyphenation setup}%
3656   \fi
3657   \let\bbl@manylang\relax}
3658 \def\bbl@process@language#1#2#3{%
3659   \ifcase\count@
3660     \@ifundefined{zth#1}{\count@\tw@}{\count@\@ne}%
3661   \or
3662     \count@\tw@
3663   \fi
3664   \ifnum\count@=\tw@
3665     \expandafter\addlanguage\csname l@#1\endcsname
3666     \language\allocationnumber
3667     \chardef\bbl@last\allocationnumber
3668     \bbl@manylang
3669     \let\bbl@elt\relax
3670     \xdef\bbl@languages{%
3671       \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
3672   \fi
3673   \the\toks@
3674   \toks@{}}
3675 \def\bbl@process@synonym@aux#1#2{%
3676   \global\expandafter\chardef\csname l@#1\endcsname#2\relax
3677   \let\bbl@elt\relax
3678   \xdef\bbl@languages{%
3679     \bbl@languages\bbl@elt{#1}{#2}{}}}%
3680 \def\bbl@process@synonym#1{%
3681   \ifcase\count@
3682     \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
3683   \or
3684     \@ifundefined{zth#1}{\bbl@process@synonym@aux{#1}{0}}{}%
3685   \else
3686     \bbl@process@synonym@aux{#1}{\the\bbl@last}%
3687   \fi}
3688 \ifx\bbl@languages\undefined % Just a (sensible?) guess
3689   \chardef\l@english\z@
3690   \chardef\l@USenglish\z@
3691   \chardef\bbl@last\z@
3692   \global\@namedef{bbl@hyphendata@0}{\hyphen.tex}{}
3693   \gdef\bbl@languages{%
3694     \bbl@elt{english}{0}{\hyphen.tex}{}%
3695     \bbl@elt{USenglish}{0}{}}
3696 \else
3697   \global\let\bbl@languages@format\bbl@languages
3698   \def\bbl@elt#1#2#3#4{% Remove all except language 0

```

```

3699     \ifnum#2>\z@\else
3700         \noexpand\bbbl@elt{#1}{#2}{#3}{#4}%
3701     \fi}%
3702     \xdef\bbbl@languages{\bbbl@languages}%
3703 \fi
3704 \def\bbbl@elt#1#2#3#4{\@namedef{zth@#1}{}} % Define flags
3705 \bbbl@languages
3706 \openin1=language.dat
3707 \ifeof1
3708     \bbbl@warning{I couldn't find language.dat. No additional\%
3709         patterns loaded. Reported}%
3710 \else
3711     \loop
3712         \endlinechar\m@ne
3713         \read1 to \bbbl@line
3714         \endlinechar\^^M
3715         \if T\ifeof1F\fi T\relax
3716         \ifx\bbbl@line\@empty\else
3717             \edef\bbbl@line{\bbbl@line\space\space\space}%
3718             \expandafter\bbbl@process@line\bbbl@line\relax
3719         \fi
3720     \repeat
3721 \fi
3722 \endgroup
3723 \bbbl@trace{Macros for reading patterns files}
3724 \def\bbbl@get@enc#1:#2:#3\@@{\def\bbbl@hyph@enc{#2}}
3725 \ifx\babelcatcodetablenum\@undefined
3726     \def\babelcatcodetablenum{5211}
3727 \fi
3728 \def\bbbl@luapatterns#1#2{%
3729     \bbbl@get@enc#1::\@@@
3730     \setbox\z@\hbox\bgroup
3731         \begingroup
3732             \ifx\catcodetable\@undefined
3733                 \let\savecatcodetable\luatexsavecatcodetable
3734                 \let\initcatcodetable\luatexinitcatcodetable
3735                 \let\catcodetable\luatexcatcodetable
3736             \fi
3737             \savecatcodetable\babelcatcodetablenum\relax
3738             \initcatcodetable\numexpr\babelcatcodetablenum+1\relax
3739             \catcodetable\numexpr\babelcatcodetablenum+1\relax
3740             \catcode`\#=6 \catcode`\$=3 \catcode`\&=4 \catcode`\^=7
3741             \catcode`\_ =8 \catcode`\{=1 \catcode`\}=2 \catcode`\~=13
3742             \catcode`\@=11 \catcode`\^^I=10 \catcode`\^^J=12
3743             \catcode`\<=12 \catcode`\>=12 \catcode`\*=12 \catcode`\.=12
3744             \catcode`\-=12 \catcode`\/=12 \catcode`\[=12 \catcode`\]=12
3745             \catcode`\`=12 \catcode`\'=12 \catcode`\`=12
3746             \input #1\relax
3747             \catcodetable\babelcatcodetablenum\relax
3748         \endgroup
3749     \def\bbbl@tempa{#2}%
3750     \ifx\bbbl@tempa\@empty\else
3751         \input #2\relax
3752     \fi
3753 \egroup}%
3754 \def\bbbl@patterns@lua#1{%
3755     \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
3756         \csname l@#1\endcsname
3757     \edef\bbbl@tempa{#1}%

```

```

3758 \else
3759 \csname l@#1:\f@encoding\endcsname
3760 \edef\bbl@tempa{#1:\f@encoding}%
3761 \fi\relax
3762 \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
3763 \@ifundefined{bbl@hyphendata@the\language}%
3764 {\def\bbl@elt##1##2##3##4{%
3765 \ifnum##2=\csname l@bbl@tempa\endcsname % #2=spanish, dutch:OT1...
3766 \def\bbl@tempb{##3}%
3767 \ifx\bbl@tempb\empty\else % if not a synonymous
3768 \def\bbl@tempc{##3}{##4}%
3769 \fi
3770 \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
3771 \fi}%
3772 \bbl@languages
3773 \@ifundefined{bbl@hyphendata@the\language}%
3774 {\bbl@info{No hyphenation patterns were set for\%
3775 language '\bbl@tempa'. Reported}}%
3776 {\expandafter\expandafter\expandafter\bbl@luapatterns
3777 \csname bbl@hyphendata@the\language\endcsname}}}%
3778 \endinput\fi
3779 \begingroup
3780 \catcode`\%=12
3781 \catcode`\'=12
3782 \catcode`\%=12
3783 \catcode`\:=12
3784 \directlua{
3785 Babel = Babel or {}
3786 function Babel.bytes(line)
3787 return line:gsub(".",
3788 function (chr) return unicode.utf8.char(string.byte(chr)) end)
3789 end
3790 function Babel.begin_process_input()
3791 if luatexbase and luatexbase.add_to_callback then
3792 luatexbase.add_to_callback('process_input_buffer',
3793 Babel.bytes, 'Babel.bytes')
3794 else
3795 Babel.callback = callback.find('process_input_buffer')
3796 callback.register('process_input_buffer', Babel.bytes)
3797 end
3798 end
3799 function Babel.end_process_input ()
3800 if luatexbase and luatexbase.remove_from_callback then
3801 luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
3802 else
3803 callback.register('process_input_buffer', Babel.callback)
3804 end
3805 end
3806 function Babel.addpatterns(pp, lg)
3807 local lg = lang.new(lg)
3808 local pats = lang.patterns(lg) or ''
3809 lang.clear_patterns(lg)
3810 for p in pp:gmatch('[^%s]+') do
3811 ss = ''
3812 for i in string.utfcharacters(p:gsub('%d', '')) do
3813 ss = ss .. '%d?' .. i
3814 end
3815 ss = ss:gsub('^%%d%?%', '%%.') .. '%d?'
3816 ss = ss:gsub('%.%%d%?$', '%%.')

```

```

3817     pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
3818     if n == 0 then
3819         tex.sprint(
3820             [[\string\csname\space bbl@info\endcsname{New pattern: }]
3821             .. p .. [{}]])
3822     pats = pats .. ' ' .. p
3823     else
3824         tex.sprint(
3825             [[\string\csname\space bbl@info\endcsname{Renew pattern: }]
3826             .. p .. [{}]])
3827     end
3828 end
3829 lang.patterns(lg, pats)
3830 end
3831 }
3832 \endgroup
3833 \ifx\newattribute\@undefined\else
3834   \newattribute\bbl@attr@locale
3835   \AddBabelHook{luatex}{beforeextras}{%
3836     \setattribute\bbl@attr@locale\localeid}
3837 \fi
3838 \def\BabelStringsDefault{unicode}
3839 \let\luabbl@stop\relax
3840 \AddBabelHook{luatex}{encodedcommands}{%
3841   \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
3842   \ifx\bbl@tempa\bbl@tempb\else
3843     \directlua{Babel.begin_process_input()}%
3844     \def\luabbl@stop{%
3845       \directlua{Babel.end_process_input()}}%
3846   \fi}%
3847 \AddBabelHook{luatex}{stopcommands}{%
3848   \luabbl@stop
3849   \let\luabbl@stop\relax}
3850 \AddBabelHook{luatex}{patterns}{%
3851   \@ifundefined{bbl@hyphendata@the\language}%
3852   {\def\bbl@elt##1##2##3##4{%
3853     \ifnum##2=\csname l@##2\endcsname % #2=spanish, dutch:OT1...
3854     \def\bbl@tempb{##3}%
3855     \ifx\bbl@tempb\@empty\else % if not a synonymous
3856       \def\bbl@tempc{##3}{##4}%
3857     \fi
3858     \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
3859     \fi}%
3860   \bbl@languages
3861   \@ifundefined{bbl@hyphendata@the\language}%
3862   {\bbl@info{No hyphenation patterns were set for\%
3863     language '#2'. Reported}}%
3864   {\expandafter\expandafter\expandafter\bbl@luapatterns
3865     \csname bbl@hyphendata@the\language\endcsname}}}%
3866 \@ifundefined{bbl@patterns@}{}%
3867 \begingroup
3868   \bbl@xin@{,\number\language,}{,\bbl@pttnlist}%
3869   \ifin\else
3870     \ifx\bbl@patterns@\@empty\else
3871       \directlua{ Babel.addpatterns(
3872         [[\bbl@patterns@]], \number\language) }%
3873     \fi
3874     \@ifundefined{bbl@patterns@#1}%
3875     \@empty

```

```

3876         {\directlua{ Babel.addpatterns(
3877             [[\space\csname bbl@patterns@#1\endcsname]],
3878             \number\language) }}%
3879         \xdef\bbl@pttnlist{\bbl@pttnlist\language,}%
3880         \fi
3881     \endgroup}}
3882 \AddBabelHook{luatex}{everylanguage}{%
3883     \def\process@language##1##2##3{%
3884         \def\process@line####1####2 ####3 ####4 {}}
3885 \AddBabelHook{luatex}{loadpatterns}{%
3886     \input #1\relax
3887     \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
3888         {{#1}}}}
3889 \AddBabelHook{luatex}{loadexceptions}{%
3890     \input #1\relax
3891     \def\bbl@tempb##1##2{{##1}{##2}}%
3892     \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
3893         {\expandafter\expandafter\expandafter\bbl@tempb
3894             \csname bbl@hyphendata@the\language\endcsname}}

```

`\babelpatterns` This macro adds patterns. Two macros are used to store them: `\bbl@patterns@` for the global ones and `\bbl@patterns@<lang>` for language ones. We make sure there is a space between words when multiple commands are used.

```

3895 \@onlypreamble\babelpatterns
3896 \AtEndOfPackage{%
3897     \newcommand\babelpatterns[2][\@empty]{%
3898         \ifx\bbl@patterns@relax
3899             \let\bbl@patterns@\@empty
3900         \fi
3901         \ifx\bbl@pttnlist\@empty\else
3902             \bbl@warning{%
3903                 You must not intermingle \string\selectlanguage\space and\\%
3904                 \string\babelpatterns\space or some patterns will not\\%
3905                 be taken into account. Reported}%
3906             \fi
3907             \ifx\@empty#1%
3908                 \protected@edef\bbl@patterns@{\bbl@patterns@\space#2}%
3909             \else
3910                 \edef\bbl@tempb{\zap@space#1 \@empty}%
3911                 \bbl@for\bbl@tempa\bbl@tempb{%
3912                     \bbl@fixname\bbl@tempa
3913                     \bbl@iflanguage\bbl@tempa{%
3914                         \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
3915                             \@ifundefined{bbl@patterns@\bbl@tempa}%
3916                             \@empty
3917                             {\csname bbl@patterns@\bbl@tempa\endcsname\space}%
3918                             #2}}}%
3919                 \fi}}

```

## 14.4 Southeast Asian scripts

*In progress.* Replace regular (ie, implicit) discretionaries by spaceskips, based on the previous glyph (which I think makes sense, because the hyphen and the previous char go always together). Other discretionaries are not touched.

For the moment, only 3 SA languages are activated by default (see Unicode UAX 14).

```

3920 \def\bbl@intraspace#1 #2 #3\@@{%
3921     \directlua{

```

```

3922 Babel = Babel or {}
3923 Babel.intraspaces = Babel.intraspaces or {}
3924 Babel.intraspaces['\csname bbl@sbcpr@\language\endcsname'] = %
3925     {b = #1, p = #2, m = #3}
3926 Babel.locale_props[\the\localeid].intraspace = %
3927     {b = #1, p = #2, m = #3}
3928 }}
3929 \def\bbl@intrapenalty#1@@{%
3930 \directlua{
3931     Babel = Babel or {}
3932     Babel.intrapenalties = Babel.intrapenalties or {}
3933     Babel.intrapenalties['\csname bbl@sbcpr@\language\endcsname'] = #1
3934     Babel.locale_props[\the\localeid].intrapenalty = #1
3935 }}
3936 \begingroup
3937 \catcode`\%=12
3938 \catcode`\^=14
3939 \catcode`\'=12
3940 \catcode`\~=12
3941 \gdef\bbl@seaintraspace{^
3942 \let\bbl@seaintraspace\relax
3943 \directlua{
3944     Babel = Babel or {}
3945     Babel.sea_enabled = true
3946     Babel.sea_ranges = Babel.sea_ranges or {}
3947     function Babel.set_chranges (script, chrng)
3948         local c = 0
3949         for s, e in string.gmatch(chrng..' ', '(.-%.%.(-)%s') do
3950             Babel.sea_ranges[script..c]={tonumber(s,16), tonumber(e,16)}
3951             c = c + 1
3952         end
3953     end
3954     function Babel.sea_disc_to_space (head)
3955         local sea_ranges = Babel.sea_ranges
3956         local last_char = nil
3957         local quad = 655360      ^^ 10 pt = 655360 = 10 * 65536
3958         for item in node.traverse(head) do
3959             local i = item.id
3960             if i == node.id'glyph' then
3961                 last_char = item
3962             elseif i == 7 and item.subtype == 3 and last_char
3963                 and last_char.char > 0x0C99 then
3964                 quad = font.getfont(last_char.font).size
3965                 for lg, rg in pairs(sea_ranges) do
3966                     if last_char.char > rg[1] and last_char.char < rg[2] then
3967                         lg = lg:sub(1, 4)
3968                         local intraspace = Babel.intraspaces[lg]
3969                         local intrapenalty = Babel.intrapenalties[lg]
3970                         local n
3971                         if intrapenalty ~= 0 then
3972                             n = node.new(14, 0)      ^^ penalty
3973                             n.penalty = intrapenalty
3974                             node.insert_before(head, item, n)
3975                         end
3976                         n = node.new(12, 13)      ^^ (glue, spaceskip)
3977                         node.setglue(n, intraspace.b * quad,
3978                                     intraspace.p * quad,
3979                                     intraspace.m * quad)
3980                         node.insert_before(head, item, n)

```

```

3981         node.remove(head, item)
3982     end
3983 end
3984 end
3985 end
3986 end
3987 }^^
3988 \bbl@luahyphenate}
3989 \catcode`\%=14
3990 \gdef\bbl@cjkintraspacespace{%
3991 \let\bbl@cjkintraspacespace\relax
3992 \directlua{
3993     Babel = Babel or {}
3994     require'babel-data-cjk.lua'
3995     Babel.cjk_enabled = true
3996     function Babel.cjk_linebreak(head)
3997         local GLYPH = node.id'glyph'
3998         local last_char = nil
3999         local quad = 655360      % 10 pt = 655360 = 10 * 65536
4000         local last_class = nil
4001         local last_lang = nil
4002
4003         for item in node.traverse(head) do
4004             if item.id == GLYPH then
4005
4006                 local lang = item.lang
4007
4008                 local LOCALE = node.get_attribute(item,
4009                     luatexbase.registernumber'bbl@attr@locale')
4010                 local props = Babel.locale_props[LOCALE]
4011
4012                 class = Babel.cjk_class[item.char].c
4013
4014                 if class == 'cp' then class = 'cl' end % )] as CL
4015                 if class == 'id' then class = 'I' end
4016
4017                 if class and last_class and Babel.cjk_breaks[last_class][class] then
4018                     br = Babel.cjk_breaks[last_class][class]
4019                 else
4020                     br = 0
4021                 end
4022
4023                 if br == 1 and props.linebreak == 'c' and
4024                     lang ~= \the\l@nohyphenation\space and
4025                     last_lang ~= \the\l@nohyphenation then
4026                     local intrapenalty = props.intrapenalty
4027                     if intrapenalty ~= 0 then
4028                         local n = node.new(14, 0)      % penalty
4029                         n.penalty = intrapenalty
4030                         node.insert_before(head, item, n)
4031                     end
4032                     local intraspacespace = props.intraspacespace
4033                     local n = node.new(12, 13)          % (glue, spaceskip)
4034                     node.setglue(n, intraspacespace.b * quad,
4035                         intraspacespace.p * quad,
4036                         intraspacespace.m * quad)
4037                     node.insert_before(head, item, n)
4038                 end
4039             end

```



```

4040         quad = font.getfont(item.font).size
4041         last_class = class
4042         last_lang = lang
4043         else % if penalty, glue or anything else
4044             last_class = nil
4045         end
4046     end
4047     lang.hyphenate(head)
4048 end
4049 }%
4050 \bbl@luahyphenate}
4051 \gdef\bbl@luahyphenate{%
4052 \let\bbl@luahyphenate\relax
4053 \directlua{
4054     luatexbase.add_to_callback('hyphenate',
4055     function (head, tail)
4056         if Babel.cjk_enabled then
4057             Babel.cjk_linebreak(head)
4058         end
4059         lang.hyphenate(head)
4060         if Babel.sea_enabled then
4061             Babel.sea_disc_to_space(head)
4062         end
4063     end,
4064     'Babel.hyphenate')
4065 }
4066 }
4067 \endgroup

```

## 14.5 CJK line breaking

Minimal line breaking for CJK scripts, mainly intended by 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 (but does not strictly follows) on the Unicode algorithm.

We first need a little table with the corresponding line breaking properties. A few characters have an additional key for the width (fullwidth vs. halfwidth), not yet used. There is a separate file, defined below.

*Work in progress.*

Common stuff.

```

4068 \AddBabelHook{luatex}{loadkernel}{%
4069 <<Restore Unicode catcodes before loading patterns>>}
4070 \ifx\DisableBabelHook\undefined\endinput\fi
4071 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
4072 \DisableBabelHook{babel-fontspec}
4073 <<Font selection>>

```

**Temporary fix** for luatex <1.10, which sometimes inserted a spurious closing dir node with a \textdir within \hboxes. This will be eventually removed.

```

4074 \def\bbl@luafixboxdir{%
4075 \setbox\z@\hbox{\textdir TLT}%
4076 \directlua{
4077     function Babel.first_dir(head)
4078         for item in node.traverse_id(node.id'dir', head) do
4079             return item
4080         end
4081         return nil

```

```

4082     end
4083     if Babel.first_dir(tex.box[0].head) then
4084         function Babel.fixboxdirs(head)
4085             local fd = Babel.first_dir(head)
4086             if fd and fd.dir:sub(1,1) == '-' then
4087                 head = node.remove(head, fd)
4088             end
4089             return head
4090         end
4091     end
4092 }
4093 \AtBeginDocument{\bbl@luafixboxdir}

```

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

```

4094 \newcommand\babelcharproperty[1]{%
4095   \count@=#1\relax
4096   \ifvmode
4097     \expandafter\bbl@chprop
4098   \else
4099     \bbl@error{\string\babelcharproperty\space can be used only in\\%
4100               vertical mode (preamble or between paragraphs)}%
4101     {See the manual for futher info}%
4102   \fi}
4103 \newcommand\bbl@chprop[3][\the\count@]{%
4104   \@tempcnta=#1\relax
4105   \bbl@ifunset{bbl@chprop@#2}%
4106   {\bbl@error{No property named '#2'. Allowed values are\\%
4107             direction (bc), mirror (bmg), and linebreak (lb)}%
4108     {See the manual for futher info}}%
4109   {%
4110   \loop
4111     \@nameuse{bbl@chprop@#2}{#3}%
4112   \ifnum\count@<\@tempcnta
4113     \advance\count@\@ne
4114   \repeat}
4115 \def\bbl@chprop@direction#1{%
4116   \directlua{
4117     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
4118     Babel.characters[\the\count@]['d'] = '#1'
4119   }}
4120 \let\bbl@chprop@bc\bbl@chprop@direction
4121 \def\bbl@chprop@mirror#1{%
4122   \directlua{
4123     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
4124     Babel.characters[\the\count@]['m'] = '\number#1'
4125   }}
4126 \let\bbl@chprop@bmg\bbl@chprop@mirror
4127 \def\bbl@chprop@linebreak#1{%
4128   \directlua{
4129     Babel.Babel.cjk_characters[\the\count@] = Babel.Babel.cjk_characters[\the\count@] or {}
4130     Babel.Babel.cjk_characters[\the\count@]['c'] = '#1'
4131   }}
4132 \let\bbl@chprop@lb\bbl@chprop@linebreak

```

## 14.6 Layout

Work in progress.

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.) and with `bidi=basic-r`, 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`.

```

4133 \bbl@trace{Redefinitions for bidi layout}
4134 \ifx\@eqnnum\undefined\else
4135   \ifx\bbl@attr@dir\undefined\else
4136     \edef\@eqnnum{%
4137       \unexpanded{\ifcase\bbl@attr@dir\else\bbl@textdir\@ne\fi}%
4138       \unexpanded\expandafter{\@eqnnum}}%
4139   \fi
4140 \fi
4141 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
4142 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
4143   \def\bbl@nextfake#1{% non-local changes, use always inside a group!
4144     \bbl@exp{%
4145       \mathdir\the\bodydir
4146       #1%           Once entered in math, set boxes to restore values
4147       \<ifmmode>%
4148       \everyvbox{%
4149         \the\everyvbox
4150         \bodydir\the\bodydir
4151         \mathdir\the\mathdir
4152         \everyhbox{\the\everyhbox}%
4153         \everyvbox{\the\everyvbox}}%
4154       \everyhbox{%
4155         \the\everyhbox
4156         \bodydir\the\bodydir
4157         \mathdir\the\mathdir
4158         \everyhbox{\the\everyhbox}%
4159         \everyvbox{\the\everyvbox}}%
4160       \<fi>}}%
4161   \def\@hangfrom#1{%
4162     \setbox\@tempboxa\hbox{\{#1\}}%
4163     \hangindent\wd\@tempboxa
4164     \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
4165       \shapemode\@ne
4166     \fi
4167     \noindent\box\@tempboxa}
4168 \fi
4169 \IfBabelLayout{tabular}
4170   {\bbl@replace\@tabular{\$}\bbl@nextfake$}%
4171   \let\bbl@tabular\@tabular
4172   \AtBeginDocument{%
4173     \ifx\bbl@tabular\@tabular\else
4174       \bbl@replace\@tabular{\$}\bbl@nextfake$}%
4175   \fi}}
4176 {}
4177 \IfBabelLayout{lists}
4178   {\expandafter\def\expandafter\bbl@toreplace\expandafter{\list{##1}{##2}}%
4179   \bbl@replace\bbl@toreplace{\parshape}\bbl@listparshape}%
4180   \bbl@exp{\def\\list##1##2{\the\toks@}}% \toks@ <- implicit result
4181   \def\bbl@listparshape#1#2#3{%
4182     \parshape #1 #2 #3 %

```

```

4183 \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
4184 \shapemode\tw@
4185 \fi}}
4186 {}

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-r, but
there are some additional readjustments for bidi=default.

4187 \IfBabelLayout{counters}%
4188 {\def\@textsuperscript#1{% lua has separate settings for math
4189 \m@th
4190 \mathdir\pagedir % required with basic-r; ok with default, too
4191 \ensuremath{^{\mbox{\fontsize\sf@size\z@ #1}}}}}%
4192 \let\bbl@latinarabic=\@arabic
4193 \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
4194 \@ifpackagewith{babel}{bidi=default}%
4195 {\let\bbl@asciroman=\@roman
4196 \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
4197 \let\bbl@asciRoman=\@Roman
4198 \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciRoman#1}}}%
4199 \def\labelenumii{}\theenumii}%
4200 \def\p@enumiii{\p@enumii}\theenumii{}\}}{}
4201 <<Footnote changes>>
4202 \IfBabelLayout{footnotes}%
4203 {\BabelFootnote\footnote\language\language{}{}}%
4204 \BabelFootnote\localfootnote\language\language{}{}}%
4205 \BabelFootnote\mainfootnote{}{}}{}
4206 {}

```

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.

```

4207 \IfBabelLayout{extras}%
4208 {\def\underline#1{%
4209 \relax
4210 \ifmmode\@underline{#1}%
4211 \else\bbl@nextfake$\@underline{\hbox{#1}}\m@th$\relax\fi}%
4212 \DeclareRobustCommand{\LaTeXe}{\mbox{\m@th
4213 \if b\expandafter\car\@series\@nil\boldmath\fi
4214 \babelsublr{%
4215 \LaTeX\kern.15em2\bbl@nextfake$_{\textstyle\varepsilon}$}}}
4216 {}
4217 </luatex>

```

## 14.7 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's not shown here. See the generated file.

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.

```

4218 (*basic-r)
4219 Babel = Babel or {}
4220
4221 Babel.bidi_enabled = true
4222
4223 require('babel-data-bidi.lua')
4224
4225 local characters = Babel.characters
4226 local ranges = Babel.ranges
4227
4228 local DIR = node.id("dir")
4229
4230 local function dir_mark(head, from, to, outer)
4231   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
4232   local d = node.new(DIR)
4233   d.dir = '+' .. dir
4234   node.insert_before(head, from, d)
4235   d = node.new(DIR)
4236   d.dir = '-' .. dir
4237   node.insert_after(head, to, d)
4238 end
4239
4240 function Babel.bidi(head, ispar)
4241   local first_n, last_n      -- first and last char with nums
4242   local last_es              -- an auxiliary 'last' used with nums
4243   local first_d, last_d      -- first and last char in L/R block
4244   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):

```

4245   local strong = ('TRT' == tex.pardir) and 'r' or 'l'
4246   local strong_lr = (strong == 'l') and 'l' or 'r'
4247   local outer = strong
4248
4249   local new_dir = false
4250   local first_dir = false
4251   local inmath = false
4252
4253   local last_lr
4254

```

```

4255 local type_n = ''
4256
4257 for item in node.traverse(head) do
4258
4259     -- three cases: glyph, dir, otherwise
4260     if item.id == node.id'glyph'
4261     or (item.id == 7 and item.subtype == 2) then
4262
4263         local itemchar
4264         if item.id == 7 and item.subtype == 2 then
4265             itemchar = item.replace.char
4266         else
4267             itemchar = item.char
4268         end
4269         local chardata = characters[itemchar]
4270         dir = chardata and chardata.d or nil
4271         if not dir then
4272             for nn, et in ipairs(ranges) do
4273                 if itemchar < et[1] then
4274                     break
4275                 elseif itemchar <= et[2] then
4276                     dir = et[3]
4277                     break
4278                 end
4279             end
4280         end
4281         dir = dir or 'l'
4282         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).

```

4283     if new_dir then
4284         attr_dir = 0
4285         for at in node.traverse(item.attr) do
4286             if at.number == luatexbase.registernumber'bbl@attr@dir' then
4287                 attr_dir = at.value % 3
4288             end
4289         end
4290         if attr_dir == 1 then
4291             strong = 'r'
4292         elseif attr_dir == 2 then
4293             strong = 'al'
4294         else
4295             strong = 'l'
4296         end
4297         strong_lr = (strong == 'l') and 'l' or 'r'
4298         outer = strong_lr
4299         new_dir = false
4300     end
4301
4302     if dir == 'nsm' then dir = strong end -- W1

```

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

```

4303     dir_real = dir -- We need dir_real to set strong below
4304     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:

```

4305     if strong == 'al' then
4306         if dir == 'en' then dir = 'an' end           -- W2
4307         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
4308         strong_lr = 'r'                               -- W3
4309     end

```

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

```

4310     elseif item.id == node.id'dir' and not inmath then
4311         new_dir = true
4312         dir = nil
4313     elseif item.id == node.id'math' then
4314         inmath = (item.subtype == 0)
4315     else
4316         dir = nil           -- Not a char
4317     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>.

```

4318     if dir == 'en' or dir == 'an' or dir == 'et' then
4319         if dir ~= 'et' then
4320             type_n = dir
4321         end
4322         first_n = first_n or item
4323         last_n = last_es or item
4324         last_es = nil
4325     elseif dir == 'es' and last_n then -- W3+W6
4326         last_es = item
4327     elseif dir == 'cs' then           -- it's right - do nothing
4328     elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
4329         if strong_lr == 'r' and type_n ~= '' then
4330             dir_mark(head, first_n, last_n, 'r')
4331         elseif strong_lr == 'l' and first_d and type_n == 'an' then
4332             dir_mark(head, first_n, last_n, 'r')
4333             dir_mark(head, first_d, last_d, outer)
4334             first_d, last_d = nil, nil
4335         elseif strong_lr == 'l' and type_n ~= '' then
4336             last_d = last_n
4337         end
4338         type_n = ''
4339         first_n, last_n = nil, nil
4340     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:

```

4341     if dir == 'l' or dir == 'r' then
4342         if dir ~= outer then
4343             first_d = first_d or item
4344             last_d = item
4345         elseif first_d and dir ~= strong_lr then

```

```

4346         dir_mark(head, first_d, last_d, outer)
4347         first_d, last_d = nil, nil
4348     end
4349 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.

```

4350     if dir and not last_lr and dir ~= 'l' and outer == 'r' then
4351         item.char = characters[item.char] and
4352             characters[item.char].m or item.char
4353     elseif (dir or new_dir) and last_lr ~= item then
4354         local mir = outer .. strong_lr .. (dir or outer)
4355         if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
4356             for ch in node.traverse(node.next(last_lr)) do
4357                 if ch == item then break end
4358                 if ch.id == node.id'glyph' then
4359                     ch.char = characters[ch.char].m or ch.char
4360                 end
4361             end
4362         end
4363     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).

```

4364     if dir == 'l' or dir == 'r' then
4365         last_lr = item
4366         strong = dir_real          -- Don't search back - best save now
4367         strong_lr = (strong == 'l') and 'l' or 'r'
4368     elseif new_dir then
4369         last_lr = nil
4370     end
4371 end

```

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

```

4372     if last_lr and outer == 'r' then
4373         for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
4374             ch.char = characters[ch.char].m or ch.char
4375         end
4376     end
4377     if first_n then
4378         dir_mark(head, first_n, last_n, outer)
4379     end
4380     if first_d then
4381         dir_mark(head, first_d, last_d, outer)
4382     end

```

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

```

4383     return node.prev(head) or head
4384 end
4385 </basic-r>

```

And here the Lua code for bidi=basic:

```

4386 <(*basic)
4387 Babel = Babel or {}
4388

```



```

4389 -- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
4390
4391 Babel.fontmap = Babel.fontmap or {}
4392 Babel.fontmap[0] = {}      -- l
4393 Babel.fontmap[1] = {}      -- r
4394 Babel.fontmap[2] = {}      -- al/an
4395
4396 Babel.bidi_enabled = true
4397
4398 require('babel-data-bidi.lua')
4399
4400 local characters = Babel.characters
4401 local ranges = Babel.ranges
4402
4403 local DIR = node.id('dir')
4404 local GLYPH = node.id('glyph')
4405
4406 local function insert_implicit(head, state, outer)
4407   local new_state = state
4408   if state.sim and state.eim and state.sim ~= state.eim then
4409     dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
4410     local d = node.new(DIR)
4411     d.dir = '+' .. dir
4412     node.insert_before(head, state.sim, d)
4413     local d = node.new(DIR)
4414     d.dir = '-' .. dir
4415     node.insert_after(head, state.eim, d)
4416   end
4417   new_state.sim, new_state.eim = nil, nil
4418   return head, new_state
4419 end
4420
4421 local function insert_numeric(head, state)
4422   local new
4423   local new_state = state
4424   if state.san and state.ean and state.san ~= state.ean then
4425     local d = node.new(DIR)
4426     d.dir = '+TLT'
4427     _, new = node.insert_before(head, state.san, d)
4428     if state.san == state.sim then state.sim = new end
4429     local d = node.new(DIR)
4430     d.dir = '-TLT'
4431     _, new = node.insert_after(head, state.ean, d)
4432     if state.ean == state.eim then state.eim = new end
4433   end
4434   new_state.san, new_state.ean = nil, nil
4435   return head, new_state
4436 end
4437
4438 -- TODO - \hbox with an explicit dir can lead to wrong results
4439 -- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>. A small attempt
4440 -- was s made to improve the situation, but the problem is the 3-dir
4441 -- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
4442 -- well.
4443
4444 function Babel.bidi(head, ispar, hdir)
4445   local d -- d is used mainly for computations in a loop
4446   local prev_d = ''
4447   local new_d = false

```

```

4448
4449 local nodes = {}
4450 local outer_first = nil
4451 local inmath = false
4452
4453 local glue_d = nil
4454 local glue_i = nil
4455
4456 local has_en = false
4457 local first_et = nil
4458
4459 local ATDIR = luatexbase.registernumber'bbl@attr@dir'
4460
4461 local save_outer
4462 local temp = node.get_attribute(head, ATDIR)
4463 if temp then
4464     temp = temp % 3
4465     save_outer = (temp == 0 and 'l') or
4466                 (temp == 1 and 'r') or
4467                 (temp == 2 and 'al')
4468 elseif ispar then -- Or error? Shouldn't happen
4469     save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
4470 else -- Or error? Shouldn't happen
4471     save_outer = ('TRT' == hdir) and 'r' or 'l'
4472 end
4473 -- when the callback is called, we are just _after_ the box,
4474 -- and the textdir is that of the surrounding text
4475 -- if not ispar and hdir ~= tex.textdir then
4476 --     save_outer = ('TRT' == hdir) and 'r' or 'l'
4477 -- end
4478 local outer = save_outer
4479 local last = outer
4480 -- 'al' is only taken into account in the first, current loop
4481 if save_outer == 'al' then save_outer = 'r' end
4482
4483 local fontmap = Babel.fontmap
4484
4485 for item in node.traverse(head) do
4486
4487     -- In what follows, #node is the last (previous) node, because the
4488     -- current one is not added until we start processing the neutrals.
4489
4490     -- three cases: glyph, dir, otherwise
4491     if item.id == GLYPH
4492         or (item.id == 7 and item.subtype == 2) then
4493
4494         local d_font = nil
4495         local item_r
4496         if item.id == 7 and item.subtype == 2 then
4497             item_r = item.replace -- automatic discs have just 1 glyph
4498         else
4499             item_r = item
4500         end
4501         local chardata = characters[item_r.char]
4502         d = chardata and chardata.d or nil
4503         if not d or d == 'nsm' then
4504             for nn, et in ipairs(ranges) do
4505                 if item_r.char < et[1] then
4506                     break

```

```

4507         elseif item_r.char <= et[2] then
4508             if not d then d = et[3]
4509             elseif d == 'nsm' then d_font = et[3]
4510             end
4511             break
4512         end
4513     end
4514 end
4515 d = d or 'l'
4516
4517 -- A short 'pause' in bidi for mapfont
4518 d_font = d_font or d
4519 d_font = (d_font == 'l' and 0) or
4520           (d_font == 'nsm' and 0) or
4521           (d_font == 'r' and 1) or
4522           (d_font == 'al' and 2) or
4523           (d_font == 'an' and 2) or nil
4524 if d_font and fontmap and fontmap[d_font][item_r.font] then
4525     item_r.font = fontmap[d_font][item_r.font]
4526 end
4527
4528 if new_d then
4529     table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
4530     if inmath then
4531         attr_d = 0
4532     else
4533         attr_d = node.get_attribute(item, ATDIR)
4534         attr_d = attr_d % 3
4535     end
4536     if attr_d == 1 then
4537         outer_first = 'r'
4538         last = 'r'
4539     elseif attr_d == 2 then
4540         outer_first = 'r'
4541         last = 'al'
4542     else
4543         outer_first = 'l'
4544         last = 'l'
4545     end
4546     outer = last
4547     has_en = false
4548     first_et = nil
4549     new_d = false
4550 end
4551
4552 if glue_d then
4553     if (d == 'l' and 'l' or 'r') ~= glue_d then
4554         table.insert(nodes, {glue_i, 'on', nil})
4555     end
4556     glue_d = nil
4557     glue_i = nil
4558 end
4559
4560 elseif item.id == DIR then
4561     d = nil
4562     new_d = true
4563
4564 elseif item.id == node.id'glue' and item.subtype == 13 then
4565     glue_d = d

```

```

4566     glue_i = item
4567     d = nil
4568
4569     elseif item.id == node.id'math' then
4570         inmath = (item.subtype == 0)
4571
4572     else
4573         d = nil
4574     end
4575
4576     -- AL <= EN/ET/ES      -- W2 + W3 + W6
4577     if last == 'al' and d == 'en' then
4578         d = 'an'          -- W3
4579     elseif last == 'al' and (d == 'et' or d == 'es') then
4580         d = 'on'          -- W6
4581     end
4582
4583     -- EN + CS/ES + EN      -- W4
4584     if d == 'en' and #nodes >= 2 then
4585         if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
4586             and nodes[#nodes-1][2] == 'en' then
4587             nodes[#nodes][2] = 'en'
4588         end
4589     end
4590
4591     -- AN + CS + AN          -- W4 too, because uax9 mixes both cases
4592     if d == 'an' and #nodes >= 2 then
4593         if (nodes[#nodes][2] == 'cs')
4594             and nodes[#nodes-1][2] == 'an' then
4595             nodes[#nodes][2] = 'an'
4596         end
4597     end
4598
4599     -- ET/EN                -- W5 + W7->l / W6->on
4600     if d == 'et' then
4601         first_et = first_et or (#nodes + 1)
4602     elseif d == 'en' then
4603         has_en = true
4604         first_et = first_et or (#nodes + 1)
4605     elseif first_et then    -- d may be nil here !
4606         if has_en then
4607             if last == 'l' then
4608                 temp = 'l'    -- W7
4609             else
4610                 temp = 'en'   -- W5
4611             end
4612         else
4613             temp = 'on'       -- W6
4614         end
4615         for e = first_et, #nodes do
4616             if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
4617         end
4618         first_et = nil
4619         has_en = false
4620     end
4621
4622     if d then
4623         if d == 'al' then
4624             d = 'r'

```

```

4625         last = 'al'
4626     elseif d == 'l' or d == 'r' then
4627         last = d
4628     end
4629     prev_d = d
4630     table.insert(nodes, {item, d, outer_first})
4631 end
4632
4633     outer_first = nil
4634
4635 end
4636
4637 -- TODO -- repeated here in case EN/ET is the last node. Find a
4638 -- better way of doing things:
4639 if first_et then      -- dir may be nil here !
4640     if has_en then
4641         if last == 'l' then
4642             temp = 'l'    -- W7
4643         else
4644             temp = 'en'   -- W5
4645         end
4646     else
4647         temp = 'on'      -- W6
4648     end
4649     for e = first_et, #nodes do
4650         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
4651     end
4652 end
4653
4654 -- dummy node, to close things
4655 table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
4656
4657 ----- NEUTRAL -----
4658
4659 outer = save_outer
4660 last = outer
4661
4662 local first_on = nil
4663
4664 for q = 1, #nodes do
4665     local item
4666
4667     local outer_first = nodes[q][3]
4668     outer = outer_first or outer
4669     last = outer_first or last
4670
4671     local d = nodes[q][2]
4672     if d == 'an' or d == 'en' then d = 'r' end
4673     if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
4674
4675     if d == 'on' then
4676         first_on = first_on or q
4677     elseif first_on then
4678         if last == d then
4679             temp = d
4680         else
4681             temp = outer
4682         end
4683         for r = first_on, q - 1 do

```

```

4684         nodes[r][2] = temp
4685         item = nodes[r][1]    -- MIRRORING
4686         if item.id == GLYPH and temp == 'r' then
4687             item.char = characters[item.char].m or item.char
4688         end
4689     end
4690     first_on = nil
4691 end
4692
4693     if d == 'r' or d == 'l' then last = d end
4694 end
4695
4696 ----- IMPLICIT, REORDER -----
4697
4698 outer = save_outer
4699 last = outer
4700
4701 local state = {}
4702 state.has_r = false
4703
4704 for q = 1, #nodes do
4705
4706     local item = nodes[q][1]
4707
4708     outer = nodes[q][3] or outer
4709
4710     local d = nodes[q][2]
4711
4712     if d == 'nsm' then d = last end          -- W1
4713     if d == 'en' then d = 'an' end
4714     local isdir = (d == 'r' or d == 'l')
4715
4716     if outer == 'l' and d == 'an' then
4717         state.san = state.san or item
4718         state.ean = item
4719     elseif state.san then
4720         head, state = insert_numeric(head, state)
4721     end
4722
4723     if outer == 'l' then
4724         if d == 'an' or d == 'r' then      -- im -> implicit
4725             if d == 'r' then state.has_r = true end
4726             state.sim = state.sim or item
4727             state.eim = item
4728         elseif d == 'l' and state.sim and state.has_r then
4729             head, state = insert_implicit(head, state, outer)
4730         elseif d == 'l' then
4731             state.sim, state.eim, state.has_r = nil, nil, false
4732         end
4733     else
4734         if d == 'an' or d == 'l' then
4735             if nodes[q][3] then -- nil except after an explicit dir
4736                 state.sim = item -- so we move sim 'inside' the group
4737             else
4738                 state.sim = state.sim or item
4739             end
4740             state.eim = item
4741         elseif d == 'r' and state.sim then
4742             head, state = insert_implicit(head, state, outer)

```

```

4743     elseif d == 'r' then
4744         state.sim, state.eim = nil, nil
4745     end
4746 end
4747
4748 if isdir then
4749     last = d           -- Don't search back - best save now
4750 elseif d == 'on' and state.san then
4751     state.san = state.san or item
4752     state.ean = item
4753 end
4754
4755 end
4756
4757 return node.prev(head) or head
4758 end
4759 </basic>

```

## 15 Data for CJK

It is a boring file and it's not shown here. See the generated file.

## 16 The ‘nil’ language

This 'language' does nothing, except setting the hyphenation patterns to nohyphenation. For this language currently no special definitions are needed or available.

The macro `\LdfInit` takes care of preventing that this file is loaded more than once, checking the category code of the `@` sign, etc.

```

4760 ⟨*nil⟩
4761 \ProvidesLanguage{nil}[(⟨date⟩) ⟨⟨version⟩⟩ Nil language]
4762 \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.

```

4763 \ifx\l@nil\undefined
4764   \newlanguage\l@nil
4765   \@namedef{bbl@hyphendata@the\l@nil}{\{\}\{\}\}% Remove warning
4766 \fi

```

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

4767 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}

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

```
\captionnil
\datenil 4768 \let\captionsnil\@empty
4769 \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.

```
4770 \ldf@finish{nil}
4771 </nil>
```

## 17 Support for Plain T<sub>E</sub>X (plain.def)

### 17.1 Not renaming hyphen.tex

As Don Knuth has declared that the filename `hyphen.tex` may only be used to designate *his* version of the american English hyphenation patterns, a new solution has to be found in order to be able to load hyphenation patterns for other languages in a plain-based T<sub>E</sub>X-format. When asked he responded:

That file name is “sacred”, and if anybody changes it they will cause severe upward/downward compatibility headaches.

People can have a file `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`

```
4772 <(*bplain | blplain)>
4773 \catcode`\{=1 % left brace is begin-group character
4774 \catcode`\}=2 % right brace is end-group character
4775 \catcode`\#=6 % hash mark is macro parameter character
```

Now let’s see if a file called `hyphen.cfg` can be found somewhere on T<sub>E</sub>X’s input path by trying to open it for reading...

```
4776 \openin 0 hyphen.cfg
```

If the file wasn’t found the following test turns out true.

```
4777 \ifeof0
4778 \else
```

When `hyphen.cfg` could be opened we make sure that *it* will be read instead of the file `hyphen.tex` which should (according to Don Knuth’s ruling) contain the american English hyphenation patterns and nothing else.

We do this by first saving the original meaning of `\input` (and I use a one letter control sequence for that so as not to waste multi-letter control sequence on this in the format).

```
4779 \let\input
```

Then `\input` is defined to forget about its argument and load `hyphen.cfg` instead.

```
4780 \def\input #1 {%
4781   \let\input\input
4782   \input{#1}
}
```

Once that’s done the original meaning of `\input` can be restored and the definition of `\input` can be forgotten.

```
4783 \let\input\input
4784 }
4785 \fi
4786 </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`.

```
4787 <bplain>\input plain.tex
4788 <blplain>\input blplain.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.

```
4789 \def\fmtname{babel-plain}
4790 \def\fmtname{babel-lplain}
```

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

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

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

```
4791 \def\plain
4792 \def\@empty{}
4793 \def\loadlocalcfg#1{%
4794   \openin0#1.cfg
4795   \ifeof0
4796     \closein0
4797   \else
4798     \closein0
4799     {\immediate\write16{*****}%
4800      \immediate\write16{* Local config file #1.cfg used}%
4801      \immediate\write16{*}%
4802     }
4803     \input #1.cfg\relax
4804   \fi
4805   \@endoflfd}
```

## 17.3 General tools

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

```
4806 \long\def\@firstofone#1{#1}
4807 \long\def\@firstoftwo#1#2{#1}
4808 \long\def\@secondoftwo#1#2{#2}
4809 \def\@nnil{\@nil}
4810 \def\@gobbletwo#1#2{}
4811 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
4812 \def\@star@or@long#1{%
4813   \@ifstar
4814   {\let\l@ngrel@x\relax#1}%
4815   {\let\l@ngrel@x\long#1}}
4816 \let\l@ngrel@x\relax
4817 \def\@car#1#2\@nil{#1}
4818 \def\@cdr#1#2\@nil{#2}
4819 \let\@typeset@protect\relax
4820 \let\protected@edef\edef
4821 \long\def\@gobble#1{}
4822 \edef\@backslashchar{\expandafter\@gobble\string\}
4823 \def\strip@prefix#1>{}
4824 \def\g@addto@macro#1#2{%
4825   \toks@\expandafter{#1#2}%
4826   \xdef#1{\the\toks@}}
4827 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
4828 \def\@nameuse#1{\csname #1\endcsname}
4829 \def\@ifundefined#1{%
4830   \expandafter\ifx\csname#1\endcsname\relax
4831     \expandafter\@firstoftwo
4832   \else
```

```

4833 \expandafter\@secondoftwo
4834 \fi}
4835 \def\@expandtwoargs#1#2#3{%
4836 \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
4837 \def\zap@space#1 #2{%
4838 #1%
4839 \ifx#2\@empty\else\expandafter\zap@space\fi
4840 #2}

```

$\LaTeX_2\epsilon$  has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

4841 \ifx\@preamblecmds\@undefined
4842 \def\@preamblecmds{}
4843 \fi
4844 \def\@onlypreamble#1{%
4845 \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
4846 \@preamblecmds\do#1}}
4847 \@onlypreamble\@onlypreamble

```

Mimick  $\LaTeX$ 's `\AtBeginDocument`; for this to work the user needs to add `\begindocument` to his file.

```

4848 \def\begindocument{%
4849 \@begindocumenthook
4850 \global\let\@begindocumenthook\@undefined
4851 \def\do##1{\global\let##1\@undefined}%
4852 \@preamblecmds
4853 \global\let\do\noexpand}
4854 \ifx\@begindocumenthook\@undefined
4855 \def\@begindocumenthook{}
4856 \fi
4857 \@onlypreamble\@begindocumenthook
4858 \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 `\@endoflfd`.

```

4859 \def\AtEndOfPackage#1{\g@addto@macro\@endoflfd{#1}}
4860 \@onlypreamble\AtEndOfPackage
4861 \def\@endoflfd{}
4862 \@onlypreamble\@endoflfd
4863 \let\bbl@afterlang\@empty
4864 \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.

```

4865 \ifx\if@files\@undefined
4866 \expandafter\let\csname if@files\expandafter\endcsname
4867 \csname iffalse\endcsname
4868 \fi

```

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

```

4869 \def\newcommand{\@star@or@long\new@command}
4870 \def\new@command#1{%
4871 \@testopt{\@newcommand#1}0}
4872 \def\@newcommand#1[#2]{%
4873 \@ifnextchar [{\@xargdef#1[#2]}%
4874 {\@argdef#1[#2]}}
4875 \long\def\@argdef#1[#2]#3{%
4876 \@yargdef#1\@ne{#2}{#3}}

```

```

4877 \long\def\xargdef#1[#2][#3]#4{%
4878   \expandafter\def\expandafter#1\expandafter{%
4879     \expandafter\@protected@testopt\expandafter #1%
4880     \csname\string#1\expandafter\endcsname{#3}}}%
4881   \expandafter\@yargdef \csname\string#1\endcsname
4882   \tw@{#2}{#4}}
4883 \long\def\@yargdef#1#2#3{%
4884   \@tempcnta#3\relax
4885   \advance \@tempcnta \@ne
4886   \let\@hash@\relax
4887   \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
4888   \@tempcntb #2%
4889   \@whilenum\@tempcntb <\@tempcnta
4890   \do{%
4891     \edef\reserved@a{\reserved@a\@hash@the\@tempcntb}%
4892     \advance\@tempcntb \@ne}%
4893   \let\@hash@##%
4894   \l@ngrelx\expandafter\def\expandafter#1\reserved@a}
4895 \def\providecommand{\@star@or@long\provide@command}
4896 \def\provide@command#1{%
4897   \begingroup
4898     \escapechar\m@ne\xdef\@gtempa{{\string#1}}%
4899   \endgroup
4900   \expandafter\@ifundefined\@gtempa
4901     {\def\reserved@a{\new@command#1}}%
4902     {\let\reserved@a\relax
4903      \def\reserved@a{\new@command\reserved@a}}%
4904   \reserved@a}%
4905 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
4906 \def\declare@robustcommand#1{%
4907   \edef\reserved@a{\string#1}%
4908   \def\reserved@b{#1}%
4909   \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
4910   \edef#1{%
4911     \ifx\reserved@a\reserved@b
4912       \noexpand\x@protect
4913       \noexpand#1%
4914     \fi
4915     \noexpand\protect
4916     \expandafter\@gobble\string#1 \endcsname
4917   }%
4918   \expandafter\new@command\csname
4919     \expandafter\@gobble\string#1 \endcsname
4920 }
4921 }
4922 \def\x@protect#1{%
4923   \ifx\protect\@typeset@protect\else
4924     \@x@protect#1%
4925   \fi
4926 }
4927 \def\@x@protect#1\fi#2#3{%
4928   \fi\protect#1%
4929 }

```

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`.

```

4930 \def\bb1@tempa{\csname newif\endcsname\ifin@}
4931 \ifx\in@\undefined
4932   \def\in@#1#2{%
4933     \def\in@##1##2##3\in@{%
4934       \ifx\in@##2\in@false\else\in@true\fi}%
4935     \in@#2#1\in@\in@}
4936 \else
4937   \let\bb1@tempa\@empty
4938 \fi
4939 \bb1@tempa

```

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

```

4940 \def\@ifpackagewith#1#2#3#4{#3}

```

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

```

4941 \def\@ifl@aded#1#2#3#4{}

```

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

```

4942 \ifx\@tempcnta\@undefined
4943   \csname newcount\endcsname\@tempcnta\relax
4944 \fi
4945 \ifx\@tempcntb\@undefined
4946   \csname newcount\endcsname\@tempcntb\relax
4947 \fi

```

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

```

4948 \ifx\bye\@undefined
4949   \advance\count10 by -2\relax
4950 \fi
4951 \ifx\@ifnextchar\@undefined
4952   \def\@ifnextchar#1#2#3{%
4953     \let\reserved@d=#1%
4954     \def\reserved@a{#2}\def\reserved@b{#3}%
4955     \futurelet\@let@token\@ifnch}
4956   \def\@ifnch{%
4957     \ifx\@let@token\@sptoken
4958       \let\reserved@c\@xifnch
4959     \else
4960       \ifx\@let@token\reserved@d
4961         \let\reserved@c\reserved@a
4962       \else
4963         \let\reserved@c\reserved@b
4964       \fi
4965     \fi
4966     \reserved@c}
4967   \def\:{\let\@sptoken= }\: % this makes \@sptoken a space token
4968   \def\:{\@xifnch} \expandafter\def\:{\futurelet\@let@token\@ifnch}
4969 \fi
4970 \def\@testopt#1#2{%

```

```

4971 \@ifnextchar[{\#1}{\#1[\#2]}}
4972 \def\@protected@testopt#1{%
4973 \ifx\protect\@typeset@protect
4974 \expandafter\@testopt
4975 \else
4976 \@@x@protect#1%
4977 \fi}
4978 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
4979 #2\relax}\fi}
4980 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
4981 \else\expandafter\@gobble\fi{#1}}

```

## 17.4 Encoding related macros

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

```

4982 \def\DeclareTextCommand{%
4983 \@dec@text@cmd\providecommand
4984 }
4985 \def\ProvideTextCommand{%
4986 \@dec@text@cmd\providecommand
4987 }
4988 \def\DeclareTextSymbol#1#2#3{%
4989 \@dec@text@cmd\chardef#1{#2}#3\relax
4990 }
4991 \def\@dec@text@cmd#1#2#3{%
4992 \expandafter\def\expandafter#2%
4993 \expandafter{%
4994 \csname#3-cmd\expandafter\endcsname
4995 \expandafter#2%
4996 \csname#3\string#2\endcsname
4997 }%
4998 % \let\@ifdefinable\@rc@ifdefinable
4999 \expandafter#1\csname#3\string#2\endcsname
5000 }
5001 \def\@current@cmd#1{%
5002 \ifx\protect\@typeset@protect\else
5003 \noexpand#1\expandafter\@gobble
5004 \fi
5005 }
5006 \def\@changed@cmd#1#2{%
5007 \ifx\protect\@typeset@protect
5008 \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
5009 \expandafter\ifx\csname ?\string#1\endcsname\relax
5010 \expandafter\def\csname ?\string#1\endcsname{%
5011 \@changed@x@err{#1}%
5012 }%
5013 \fi
5014 \global\expandafter\let
5015 \csname\cf@encoding\string#1\expandafter\endcsname
5016 \csname ?\string#1\endcsname
5017 \fi
5018 \csname\cf@encoding\string#1%
5019 \expandafter\endcsname
5020 \else
5021 \noexpand#1%
5022 \fi
5023 }
5024 \def\@changed@x@err#1{%

```

```

5025 \errhelp{Your command will be ignored, type <return> to proceed}%
5026 \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
5027 \def\DeclareTextCommandDefault#1{%
5028 \DeclareTextCommand#1?%
5029 }
5030 \def\ProvideTextCommandDefault#1{%
5031 \ProvideTextCommand#1?%
5032 }
5033 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
5034 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
5035 \def\DeclareTextAccent#1#2#3{%
5036 \DeclareTextCommand#1{#2}[1]{\accent#3 ##1}
5037 }
5038 \def\DeclareTextCompositeCommand#1#2#3#4{%
5039 \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
5040 \edef\reserved@b{\string##1}%
5041 \edef\reserved@c{%
5042 \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
5043 \ifx\reserved@b\reserved@c
5044 \expandafter\expandafter\expandafter\ifx
5045 \expandafter\@car\reserved@a\relax\relax\@nil
5046 \@text@composite
5047 \else
5048 \edef\reserved@b##1{%
5049 \def\expandafter\noexpand
5050 \csname#2\string#1\endcsname####1{%
5051 \noexpand\@text@composite
5052 \expandafter\noexpand\csname#2\string#1\endcsname
5053 ####1\noexpand\@empty\noexpand\@text@composite
5054 {##1}%
5055 }%
5056 }%
5057 \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
5058 \fi
5059 \expandafter\def\csname\expandafter\string\csname
5060 #2\endcsname\string#1-\string#3\endcsname{#4}
5061 \else
5062 \errhelp{Your command will be ignored, type <return> to proceed}%
5063 \errmessage{\string\DeclareTextCompositeCommand\space used on
5064 inappropriate command \protect#1}
5065 \fi
5066 }
5067 \def\@text@composite#1#2#3\@text@composite{%
5068 \expandafter\@text@composite@x
5069 \csname\string#1-\string#2\endcsname
5070 }
5071 \def\@text@composite@x#1#2{%
5072 \ifx#1\relax
5073 #2%
5074 \else
5075 #1%
5076 \fi
5077 }
5078 %
5079 \def\@strip@args#1:#2-#3\@strip@args{#2}
5080 \def\DeclareTextComposite#1#2#3#4{%
5081 \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
5082 \bgroup
5083 \lcode`\@=#4%

```

```

5084     \lowercase{%
5085     \egroup
5086     \reserved@a @%
5087     }%
5088 }
5089 %
5090 \def\UseTextSymbol#1#2{%
5091 %   \let\@curr@enc\cf@encoding
5092 %   \@use@text@encoding{#1}%
5093 %   #2%
5094 %   \@use@text@encoding\@curr@enc
5095 }
5096 \def\UseTextAccent#1#2#3{%
5097 %   \let\@curr@enc\cf@encoding
5098 %   \@use@text@encoding{#1}%
5099 %   #2{\@use@text@encoding\@curr@enc\selectfont#3}%
5100 %   \@use@text@encoding\@curr@enc
5101 }
5102 \def\@use@text@encoding#1{%
5103 %   \edef\f@encoding{#1}%
5104 %   \xdef\font@name{%
5105 %       \csname\curr@fontshape/\f@size\endcsname
5106 %   }%
5107 %   \pickup@font
5108 %   \font@name
5109 %   \@enc@update
5110 }
5111 \def\DeclareTextSymbolDefault#1#2{%
5112     \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
5113 }
5114 \def\DeclareTextAccentDefault#1#2{%
5115     \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
5116 }
5117 \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.

```

5118 \DeclareTextAccent{"}{OT1}{127}
5119 \DeclareTextAccent{'}{OT1}{19}
5120 \DeclareTextAccent{^}{OT1}{94}
5121 \DeclareTextAccent{`}{OT1}{18}
5122 \DeclareTextAccent{~}{OT1}{126}

```

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

```

5123 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
5124 \DeclareTextSymbol{\textquotedblright}{OT1}{`\"}
5125 \DeclareTextSymbol{\textquoteleft}{OT1}{`\'}
5126 \DeclareTextSymbol{\textquoteright}{OT1}{`\'}
5127 \DeclareTextSymbol{\i}{OT1}{16}
5128 \DeclareTextSymbol{\ss}{OT1}{25}

```

For a couple of languages we need the  $\text{\LaTeX}$ -control sequence `\scriptsize` to be available. Because plain  $\text{\TeX}$  doesn't have such a sophisticated font mechanism as  $\text{\LaTeX}$  has, we just `\let` it to `\sevenrm`.

```

5129 \ifx\scriptsize\undefined
5130     \let\scriptsize\sevenrm
5131 \fi
5132 </plain>

```

## 18 Acknowledgements

I would like to thank all who volunteered as  $\beta$ -testers for their time. Michel Goossens supplied contributions for most of the other languages. Nico Poppelier helped polish the text of the documentation and supplied parts of the macros for the Dutch language. Paul Wackers and Werenfried Spit helped find and repair bugs. During the further development of the babel system I received much help from Bernd Raichle, for which I am grateful.

## References

- [1] Huda Smitshuijzen Abifares, *Arabic Typography*, Saqi, 2001.
- [2] Donald E. Knuth, *The T<sub>E</sub>Xbook*, Addison-Wesley, 1986.
- [3] Leslie Lamport, *L<sup>A</sup>T<sub>E</sub>X, A document preparation System*, Addison-Wesley, 1986.
- [4] K.F. Treebus. *Tekstwijzer, een gids voor het grafisch verwerken van tekst*, SDU Uitgeverij ('s-Gravenhage, 1988).
- [5] Hubert Partl, *German T<sub>E</sub>X*, *TUGboat* 9 (1988) #1, p. 70–72.
- [6] Leslie Lamport, in: T<sub>E</sub>Xhax Digest, Volume 89, #13, 17 February 1989.
- [7] Johannes Braams, Victor Eijkhout and Nico Poppelier, *The development of national L<sup>A</sup>T<sub>E</sub>X styles*, *TUGboat* 10 (1989) #3, p. 401–406.
- [8] Yannis Haralambous, *Fonts & Encodings*, O'Reilly, 2007.
- [9] Jukka K. Korpela, *Unicode Explained*, O'Reilly, 2006.
- [10] Ken Lunde, *CJKV Information Processing*, O'Reilly, 2nd ed., 2009.
- [11] Joachim Schrod, *International L<sup>A</sup>T<sub>E</sub>X is ready to use*, *TUGboat* 11 (1990) #1, p. 87–90.
- [12] Apostolos Syropoulos, Antonis Tsolomitis and Nick Sofroniu, *Digital typography using L<sup>A</sup>T<sub>E</sub>X*, Springer, 2002, p. 301–373.