

# Babel

Version 3.35

2019/10/15

*Original author*

Johannes L. Braams

*Current maintainer*

Javier Bezos

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

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

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

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

# Contents

<b>I</b>	<b>User guide</b>	<b>4</b>
<b>1</b>	<b>The user interface</b>	<b>4</b>
1.1	Monolingual documents . . . . .	4
1.2	Multilingual documents . . . . .	5
1.3	Modifiers . . . . .	6
1.4	xelatex and luatex . . . . .	7
1.5	Troubleshooting . . . . .	8
1.6	Plain . . . . .	8
1.7	Basic language selectors . . . . .	8
1.8	Auxiliary language selectors . . . . .	9
1.9	More on selection . . . . .	10
1.10	Shorthands . . . . .	11
1.11	Package options . . . . .	14
1.12	The base option . . . . .	16
1.13	ini files . . . . .	17
1.14	Selecting fonts . . . . .	24
1.15	Modifying a language . . . . .	26
1.16	Creating a language . . . . .	27
1.17	Digits . . . . .	29
1.18	Getting the current language name . . . . .	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 . . . . .	38
1.25	Unicode character properties in luatex . . . . .	39
1.26	Tips, workarounds, known 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 . . . . .	42
<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 . . . . .	44
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 . . . . .	48
<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 . . . . .	57
<b>8</b>	<b>The Package File (<math>\LaTeX</math>, babel.sty)</b>	<b>58</b>
8.1	base . . . . .	58
8.2	key=value options and other general option . . . . .	60
8.3	Conditional loading of shorthands . . . . .	62
8.4	Language options . . . . .	63
<b>9</b>	<b>The kernel of Babel (babel.def, common)</b>	<b>66</b>
9.1	Tools . . . . .	66
9.2	Hooks . . . . .	68
9.3	Setting up language files . . . . .	70
9.4	Shorthands . . . . .	73
9.5	Language attributes . . . . .	82
9.6	Support for saving macro definitions . . . . .	84
9.7	Short tags . . . . .	85
9.8	Hyphens . . . . .	85
9.9	Multiencoding strings . . . . .	87
9.10	Macros common to a number of languages . . . . .	93
9.11	Making glyphs available . . . . .	93
9.11.1	Quotation marks . . . . .	93
9.11.2	Letters . . . . .	94
9.11.3	Shorthands for quotation marks . . . . .	95
9.11.4	Umlauts and tremas . . . . .	96
9.12	Layout . . . . .	97
9.13	Load engine specific macros . . . . .	98
9.14	Creating languages . . . . .	98
<b>10</b>	<b>The kernel of Babel (babel.def, only <math>\LaTeX</math>)</b>	<b>109</b>
10.1	The redefinition of the style commands . . . . .	109
10.2	Cross referencing macros . . . . .	109
10.3	Marks . . . . .	112
10.4	Preventing clashes with other packages . . . . .	114
10.4.1	ifthen . . . . .	114
10.4.2	varioref . . . . .	114
10.4.3	hhline . . . . .	115
10.4.4	hyperref . . . . .	115
10.4.5	fancyhdr . . . . .	116
10.5	Encoding and fonts . . . . .	116
10.6	Basic bidi support . . . . .	118
10.7	Local Language Configuration . . . . .	121
<b>11</b>	<b>Multiple languages (switch.def)</b>	<b>122</b>
11.1	Selecting the language . . . . .	123
11.2	Errors . . . . .	131
<b>12</b>	<b>Loading hyphenation patterns</b>	<b>133</b>
<b>13</b>	<b>Font handling with fontspec</b>	<b>137</b>

<b>14</b>	<b>Hooks for XeTeX and LuaTeX</b>	<b>142</b>
14.1	XeTeX . . . . .	142
14.2	Layout . . . . .	144
14.3	LuaTeX . . . . .	146
14.4	Southeast Asian scripts . . . . .	151
14.5	CJK line breaking . . . . .	154
14.6	Layout . . . . .	155
14.7	Auto bidi with basic and basic-r . . . . .	157
<b>15</b>	<b>Data for CJK</b>	<b>168</b>
<b>16</b>	<b>The ‘nil’ language</b>	<b>168</b>
<b>17</b>	<b>Support for Plain T<sub>E</sub>X (plain.def)</b>	<b>169</b>
17.1	Not renaming hyphen.tex . . . . .	169
17.2	Emulating some L <sup>A</sup> T <sub>E</sub> X features . . . . .	170
17.3	General tools . . . . .	170
17.4	Encoding related macros . . . . .	174
<b>18</b>	<b>Acknowledgements</b>	<b>177</b>

## Troubleshooting

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

## Part I

# User guide

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

## 1 The user interface

### 1.1 Monolingual documents

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

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

```
\documentclass{article}

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

\usepackage[french]{babel}

\begin{document}

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

\end{document}
```

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

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

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

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

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

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

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

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

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

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

## 1.2 Multilingual documents

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

### 1.3 Modifiers

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

---

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

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

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

## 1.4 xelatex and lualatex

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

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

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

```
\documentclass{article}

\usepackage[vietnamese,danish]{babel}

\begin{document}

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

\selectlanguage{vietnamese}

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

\end{document}
```

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

```
\documentclass{article}

\usepackage[russian]{babel}

\babelfont{rm}{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 L<sup>A</sup>T<sub>E</sub>X (ie, `\usepackage{⟨language⟩}`) is deprecated and you will get the error:<sup>2</sup>

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

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

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

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

## 1.6 Plain

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

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

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

## 1.7 Basic language selectors

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

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

`\selectlanguage` {⟨language⟩}

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

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

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

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

```
\selectlanguage{german}
```

This command can be used as environment, too.

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

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

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

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

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

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

## 1.8 Auxiliary language selectors

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

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

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

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

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

Spaces after the environment are ignored.

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

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

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

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

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

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

## 1.9 More on selection

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

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

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

**EXAMPLE** With

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

you can write

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

and

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

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

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

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

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

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

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

```
\babelensure{polish}
```

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

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

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

## 1.10 Shorthands

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

**NOTE** Note the following:

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

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

A typical error when using shorthands is the following:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## 1.11 Package options

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

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

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

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



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

## 1.12 The base option

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

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

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

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

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

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

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

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

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

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

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

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

### 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]{lᦺ lᦴ lᦶ lᦸ lᦺ lᦴ lᦶ lᦸ} % Random
```

Khemer 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 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>	as	Assamese
agq	Aghem	asa	Asu
ak	Akan	ast	Asturian <sup>ul</sup>
am	Amharic <sup>ul</sup>	az-Cyrl	Azerbaijani
ar	Arabic <sup>ul</sup>	az-Latn	Azerbaijani
ar-DZ	Arabic <sup>ul</sup>	az	Azerbaijani <sup>ul</sup>
ar-MA	Arabic <sup>ul</sup>	bas	Basaa
ar-SY	Arabic <sup>ul</sup>	be	Belarusian <sup>ul</sup>

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

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

vun	Vunjo	zh-Hans-HK	Chinese
wae	Walser	zh-Hans-MO	Chinese
xog	Soga	zh-Hans-SG	Chinese
yav	Yangben	zh-Hans	Chinese
yi	Yiddish	zh-Hant-HK	Chinese
yo	Yoruba	zh-Hant-MO	Chinese
yue	Cantonese	zh-Hant	Chinese
zgh	Standard Moroccan Tamazight	zh	Chinese
		zu	Zulu

---

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

---

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

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

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

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

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



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

## 1.14 Selecting fonts

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

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

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

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

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

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

```
\documentclass{article}

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

\babelprovide[import]{hebrew}

\babelfont{rm}{FreeSerif}

\begin{document}

Svenska \foreignlanguage{hebrew}{עִבְרִית} svenska.

\end{document}
```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**TROUBLESHOOTING** *Package fontspec Warning: 'Language 'LANG' not available for font 'FONT' with script 'SCRIPT' 'Default' language used instead'.* This 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.

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

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

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

**NOTE** These macros (`\captions⟨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⟩*}

If the language `\marg{language-name}` has not been defined and there are no options, it creates an “empty” one in the following way: 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, some options allow to fill the language, and babel warns you about what to do if there is a missing string. Very likely you will find alerts like that in the log file:

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

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

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

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

The main language is not changed (danish in this example). So, you must add

`\selectlanguage{arhinish}` or other selectors where necessary.

If the language has been loaded as an argument in `\documentclass` or `\usepackage`, then `\babelprovide` redefines the requested data.

`import=` *⟨language-tag⟩*

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

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

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

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

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

There are about 200 ini files, with data taken from the ldf files and the CLDR provided by Unicode. Not all languages in the latter are complete, and therefore neither are the ini

files. A few languages will show a warning about the current lack of suitability of the date format (hindi, french, breton, and occitan).

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

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

Loads only the strings. For example:

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

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

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

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

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

A special value is `+`, which allocates a new language (in the  $\text{T}_{\text{E}}\text{X}$  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=** `\<script-name>`

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

**language=** `\<language-name>`

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

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

**mapfont=** direction

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

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

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

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

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

`\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  $\TeX$  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  $\TeX$  are entered as `-`, and (2) *optional* or *soft hyphens*, which are entered as `\-`. Strictly, a *soft hyphen* is not a hyphen, but just a breaking opportunity or, in  $\TeX$  terms, a “discretionary”; a *hard hyphen* is a hyphen with a breaking opportunity after it. A further type is a *non-breaking hyphen*, a hyphen without a breaking opportunity. In  $\TeX$ , `-` and `\-` forbid further breaking opportunities in the word. This is the desired behavior very often, but not always, and therefore many languages provide shorthands for these cases. Unfortunately, this has not been done consistently: for example, `-` in Dutch, Portuguese, Catalan or Danish is a hard hyphen, while in German, Spanish, Norwegian, Slovak or Russian is a soft hyphen. Furthermore, some of them even redefine `\-`, so that you cannot insert a soft hyphen without breaking opportunities in the rest of the word. Therefore, some macros are provided with a set of basic “hyphens” which can be used by themselves, to define a user shorthand, or even in language files.

- `\babelhyphen{soft}` and `\babelhyphen{hard}` are self explanatory.
- `\babelhyphen{repeat}` inserts a hard hyphen which is repeated at the beginning of the next line, as done in languages like Polish, Portuguese and Spanish.
- `\babelhyphen{nobreak}` inserts a hard hyphen without a break after it (even if a space follows).
- `\babelhyphen{empty}` inserts a break opportunity without a hyphen at all.
- `\babelhyphen{\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 to enable it, there is a starred counterpart: `\babelhyphen*\{soft\}` (which in most cases is equivalent to the original `\-`), `\babelhyphen*\{hard\}`, etc.

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

There are also some differences with  $\LaTeX$ : (1) the character used is that set for the current font, while in  $\LaTeX$  it is hardwired to `-` (a typical value); (2) the hyphen to be used in fonts

with a negative `\hyphenchar` is `-`, like in  $\TeX$ , but it can be changed to another value by redefining `\babelnullhyphen`; (3) a break after the hyphen is forbidden if preceded by a glue  $>0$  pt (at the beginning of a word, provided it is not immediately preceded by, say, a parenthesis).

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

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

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

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

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

**`\babelpatterns`** [`\language`], [`\language`], ...]{`\patterns`}

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

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

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

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

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

## 1.20 Selecting scripts

Currently babel provides no standard interface to select scripts, because they are best selected with either `\fontencoding` (low-level) or a language name (high-level). Even the

<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.



Latin script may require different encodings (ie, sets of glyphs) depending on the language, and therefore such a switch would be in a sense incomplete.<sup>18</sup>

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

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

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

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

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

## 1.21 Selecting directions

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

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

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

There are some package options controlling bidi writing.

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

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

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

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

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

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

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

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

**EXAMPLE** The following text comes from the Arabic Wikipedia (article about Arabia).

Copy-pasting some text from the Wikipedia is a good way to test this feature.

Remember `basic-r` is available in luatex only.

```
\documentclass{article}

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

\babelprovide[import, main]{arabic}

\babelfont{rm}{FreeSerif}

\begin{document}

    وقد عرفت شبه جزيرة العرب طيلة العصر الهيليني (الاجريقي) بـ
    Arabia أو Aravia (بالاغريقية Αραβία)، استخدم الرومان ثلاث
    بادئات بـ "Arabia" على ثلاث مناطق من شبه الجزيرة العربية، إلا أنها
    حقيقةً كانت أكبر مما تعرف عليه اليوم.

\end{document}
```

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

```
\documentclass{book}

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

```
\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 `r l` 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**  $\{\langle cmd \rangle\}\{\langle local-language \rangle\}\{\langle before \rangle\}\{\langle after \rangle\}$

**New 3.17** Something like:

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

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

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

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

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

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

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

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

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

## 1.22 Language attributes

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

## 1.23 Hooks

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

**\AddBabelHook** [*<lang>*]{*<name>*}{*<event>*}{*<code>*}

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## 1.24 Languages supported by babel with ldf files

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

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

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

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

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

## 1.25 Unicode character properties in luatex

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

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

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

- If you use the document class book *and* you use  $\backslash\text{ref}$  inside the argument of  $\backslash\text{chapter}$  (or just use  $\backslash\text{ref}$  inside  $\backslash\text{MakeUppercase}$ ),  $\LaTeX$  will keep complaining about an undefined label. To prevent such problems, you could revert to using uppercase labels, you can use  $\backslash\text{lowercase}\{\backslash\text{ref}\{\text{foo}\}\}$  inside the argument of  $\backslash\text{chapter}$ , or, if you will not use shorthands in labels, set the safe option to none or bib.
- Both ltxdoc and babel use  $\backslash\text{AtBeginDocument}$  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 T<sub>E</sub>X only takes into account the values when the paragraph is hyphenated, i.e., when it has been finished.<sup>21</sup> So, if you write a chunk of French text with `\foreignlanguage`, the apostrophes might not be taken into account. This is a limitation of T<sub>E</sub>X, not of babel. Alternatively, you may use `\useshortands` to activate ' and `\defineshortand`, or redefine `\textquoteright` (the latter is called by the non-ASCII right quote).
- `\bibitem` is out of sync with `\selectlanguage` in the .aux file. The reason is `\bibitem` uses `\immediate` (and others, in fact), while `\selectlanguage` doesn't. There is no known workaround.
- Babel does not take into account `\normalsfcodes` and (non-)French spacing is not always properly (un)set by languages. However, problems are unlikely to happen and therefore this part remains untouched in version 3.9 (but it is in the 'to do' list).
- Using a character mathematically active (ie, with math code "8000) as a shorthand can make T<sub>E</sub>X enter in an infinite loop in some rare cases. (Another issue in the 'to do' list, although there is a partial solution.)

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

**csquotes** Logical markup for quotes.

**iflang** Tests correctly the current language.

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

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

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

**biblatex** Programmable bibliographies and citations.

**bicaption** Bilingual captions.

**babelbib** Multilingual bibliographies.

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

Ligatures can be disabled.

**substitutefont** Combines fonts in several encodings.

**mkpattern** Generates hyphenation patterns.

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

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

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

## 1.27 Current and future work

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

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

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.” may be referred to as either “ítem 3.” or “3.<sup>er</sup> ítem”, and so on.

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

## 1.28 Tentative and experimental code

See the code section for  $\backslash\text{foreignlanguage}^*$  (a new starred version of  $\backslash\text{foreignlanguage}$ ).

### Old stuff

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

- $\backslash\text{babelFSstore}\{\langle\text{babel-language}\rangle\}$  sets the current three basic families (rm, sf, tt) as the default for the language given.
- $\backslash\text{babelFSdefault}\{\langle\text{babel-language}\rangle\}\{\langle\text{fontspec-features}\rangle\}$  patches  $\backslash\text{fontspec}$  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`

$\text{T}_{\text{E}}\text{X}$  and most engines based on it ( $\text{pdfT}_{\text{E}}\text{X}$ ,  $\text{xetex}$ ,  $\epsilon\text{-T}_{\text{E}}\text{X}$ , the main exception being  $\text{luatex}$ ) require hyphenation patterns to be preloaded when a format is created (eg,  $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ ,  $\text{XeL}_{\text{A}}\text{T}_{\text{E}}\text{X}$ ,  $\text{pdfL}_{\text{A}}\text{T}_{\text{E}}\text{X}$ ).  $\text{babel}$  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  $\text{luatex}$ , 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  $\text{luatex-hyphen}$ , 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**

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

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

**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  $\text{\TeX}$  environment has to record for which languages he has hyphenation patterns *and* in which files these are stored<sup>25</sup>. When hyphenation exceptions are stored in a separate file this can be indicated by naming that file *after* the file with the hyphenation patterns.

The file can contain empty lines and comments, as well as lines which start with an equals (=) sign. Such a line will instruct  $\text{\LaTeX}$  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:

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

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

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

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

Some recommendations:

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

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

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

### 3.1 Guidelines for contributed languages

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

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

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

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

### 3.2 Basic macros

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

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

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

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

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

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

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

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

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

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

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

### 3.3 Skeleton

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

```

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

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

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

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

```

```

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

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

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

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

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

\EndBabelCommands

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

\ldf@finish{<language>}

```

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

```

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

```

### 3.4 Support for active characters

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

`\initiate@active@char`

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

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

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

`\declare@shorthand`

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



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

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

### 3.5 Support for saving macro definitions

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

`\babel@save`

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

`\babel@savevariable`

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

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

### 3.6 Support for extending macros

`\addto`

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

### 3.7 Macros common to a number of languages

`\bbl@allowhyphens`

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

`\allowhyphens`

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

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

`\set@low@box`

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

`\save@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.

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



### 3.8 Encoding-dependent strings

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

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

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

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

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

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

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

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

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

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

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

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

```

\SetString{\chaptername}{utf8-string}

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

\EndBabelCommands

```

A real example is:

```

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

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

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

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

\StartBabelCommands{german,austrian}{date}
\SetString\monthiiiname{Februar}
\SetString\monthiiiname{M\"a\"rz}
\SetString\monthivname{April}
\SetString\monthvname{Mai}
\SetString\monthviname{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  $\backslash date \langle language \rangle$  exists).

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

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

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

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

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

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

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

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

#1 is replaced by the roman numeral.

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

```
\StartBabelCommands{turkish}{}[ot1enc, fontenc=OT1]  
\SetCase  
  {\uccode"10=`I\relax}  
  {\lccode`I="10\relax}  
  
\StartBabelCommands{turkish}{}[unicode, fontenc=TU EU1 EU2, charset=utf8]  
\SetCase  
  {\uccode`i=`İ\relax  
   \uccode`ı=`I\relax}  
  {\lccode`İ=`i\relax  
   \lccode`I=`ı\relax}  
  
\StartBabelCommands{turkish}{}  
\SetCase  
  {\uccode`i="9D\relax  
   \uccode"19=`I\relax}  
  {\lccode"9D=`i\relax  
   \lccode`I="19\relax}
```

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

```
\EndBabelCommands
```

(Note the mapping for OT1 is not complete.)

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

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

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

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

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

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

## 4 Changes

### 4.1 Changes in `babel` version 3.9

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

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

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

## Part II

# Source code

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

## 5 Identification and loading of required files

*Code documentation is still under revision.*

The babel package after unpacking consists of the following files:

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

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

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

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

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

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

## 6 locale directory

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

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

This is a preliminary documentation.

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

Most keys are self-explanatory.

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

**version** of the ini file

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

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

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

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

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

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

## 7 Tools

```
1 <<version=3.35>>
2 <<date=2019/10/15>>
```

**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, we take extra care to ‘throw’ it over the `\else` and `\fi` parts of an `\if`-statement<sup>31</sup>. These macros will break if another `\if... \fi` statement appears in one of the arguments and it is not enclosed in braces.

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

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

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

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

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

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

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

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

```

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

```

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

```

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

```

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

```

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

```

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

```

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

```

`\bbl@replace`

```

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

```

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



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

```

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

```

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

```

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

```

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

```

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

```

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

```

157 <<*Load patterns in luatex>> ≡
158 \ifx\directlua\@undefined\else
159   \ifx\bb1\luapatterns\@undefined
160     \input luababel.def
161   \fi
162 \fi
163 <</Load patterns in luatex>>

```

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

```

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

```

## 7.1 Multiple languages

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

```

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

```

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

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

```

174 <<*Define core switching macros>> ≡
175 \ifx\newlanguage\@undefined
176   \csname newcount\endcsname\last@language
177   \def\addlanguage#1{%
178     \global\advance\last@language\@ne
179     \ifnum\last@language<\@cc1vi
180       \else

```

```

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

```

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

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

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

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

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

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

### 8.1 base

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

```

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

```

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

```

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

```

```

262         or nil)
263     %
264     luatexbase.add_to_callback('hpack_filter',
265         Babel.pre_otfload_h,
266         'Babel.pre_otfload_h',
267         luatexbase.priority_in_callback('hpack_filter',
268             '\bbl@harfprefline')
269         or luatexbase.priority_in_callback('hpack_filter',
270             'luaotfload.node_processor')
271         or nil)
272     }%
273     \@ifpackageloaded{harfload}%
274     {\directlua{ Babel.mirroring_enabled = false }}%
275     {}}
276     \let\bbl@tempa\relax
277     \@ifpackagewith{babel}{bidi=basic}%
278     {\def\bbl@tempa{basic}}%
279     {\@ifpackagewith{babel}{bidi=basic-r}%
280     {\def\bbl@tempa{basic-r}}%
281     {}}
282     \ifx\bbl@tempa\relax\else
283         \let\bbl@beforeforeign\leavevmode
284         \AtEndOfPackage{\EnableBabelHook{babel-bidi}}%
285         \RequirePackage{luatexbase}%
286         \directlua{
287             require('babel-data-bidi.lua')
288             require('babel-bidi-\bbl@tempa.lua')
289         }
290         \bbl@activate@preotf
291     \fi
292 \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.

```

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

```

308 \bbl@trace{key=value and another general options}

```

```

309 \bbl@csarg\let{tempa\expandafter}\csname opt@babel.sty\endcsname
310 \def\bbl@tempb#1.#2{%
311   #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%
312 \def\bbl@tempd#1.#2\@nnil{%
313   \ifx\@empty#2%
314     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
315   \else
316     \in@{=}{#1}\ifin@
317     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
318   \else
319     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
320   \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
321   \fi
322 \fi}
323 \let\bbl@tempc\@empty
324 \bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty\@nnil}
325 \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.

```

326 \DeclareOption{KeepShorthandsActive}{}
327 \DeclareOption{activeacute}{}
328 \DeclareOption{activegrave}{}
329 \DeclareOption{debug}{}
330 \DeclareOption{noconfigs}{}
331 \DeclareOption{showlanguages}{}
332 \DeclareOption{silent}{}
333 \DeclareOption{mono}{}
334 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
335 % Don't use. Experimental:
336 \newif\ifbbl@single
337 \DeclareOption{selectors=off}{\bbl@singletrue}
338 <<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.

```

339 \let\bbl@opt@shorthands\@nnil
340 \let\bbl@opt@config\@nnil
341 \let\bbl@opt@main\@nnil
342 \let\bbl@opt@headfoot\@nnil
343 \let\bbl@opt@layout\@nnil

```

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

```

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

```

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

```

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

```

362 \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=...

```

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

```

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

```

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

```

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

```

384   \edef\bbl@opt@shorthands{%
385     \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.

```

386   \bbl@ifshorthand{'}%
387   {\PassOptionsToPackage{activeacute}{babel}}{}
388   \bbl@ifshorthand{`}%

```

```

389     {\PassOptionsToPackage{activegrave}{babel}}{}
390 \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.

```

391 \ifx\bbl@opt@headfoot\@nnil\else
392   \g@addto@macro\@resetactivechars{%
393     \set@typeset@protect
394     \expandafter\select@language@x\expandafter{\bbl@opt@headfoot}%
395     \let\protect\noexpand}
396 \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.

```

397 \ifx\bbl@opt@safe\@undefined
398   \def\bbl@opt@safe{BR}
399 \fi
400 \ifx\bbl@opt@main\@nnil\else
401   \edef\bbl@language@opts{%
402     \ifx\bbl@language@opts\@empty\else\bbl@language@opts,\fi
403     \bbl@opt@main}
404 \fi

```

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

```

405 \bbl@trace{Defining IfBabelLayout}
406 \ifx\bbl@opt@layout\@nnil
407   \newcommand\IfBabelLayout[3]{#3}%
408 \else
409   \newcommand\IfBabelLayout[1]{%
410     \@expandtwoargs\in@{.#1.}{.\bbl@opt@layout.}%
411     \ifin@
412       \expandafter\@firstoftwo
413     \else
414       \expandafter\@secondoftwo
415     \fi}
416 \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).

```

417 \bbl@trace{Language options}
418 \let\bbl@afterlang\relax
419 \let\BabelModifiers\relax
420 \let\bbl@loaded\@empty
421 \def\bbl@load@language#1{%
422   \InputIfFileExists{#1.ldf}%
423   {\edef\bbl@loaded{\CurrentOption
424     \ifx\bbl@loaded\@empty\else,\bbl@loaded\fi}%
425     \expandafter\let\expandafter\bbl@afterlang
426       \csname\CurrentOption.ldf-h@@k\endcsname
427     \expandafter\let\expandafter\BabelModifiers
428       \csname bbl@mod@\CurrentOption\endcsname}%
429   {\bbl@error{%
430     Unknown option '\CurrentOption'. Either you misspelled it\\%

```



```

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

```

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

```

435 \def\bbl@try@load@lang#1#2#3{%
436   \IfFileExists{\CurrentOption.ldf}%
437   {\bbl@load@language{\CurrentOption}}%
438   {#1\bbl@load@language{#2}#3}}
439 \DeclareOption{afrikaans}{\bbl@try@load@lang{}{dutch}{}}
440 \DeclareOption{brazil}{\bbl@try@load@lang{}{portuges}{}}
441 \DeclareOption{brazilian}{\bbl@try@load@lang{}{portuges}{}}
442 \DeclareOption{hebrew}{%
443   \input{rlbabel.def}%
444   \bbl@load@language{hebrew}}
445 \DeclareOption{hungarian}{\bbl@try@load@lang{}{magyar}{}}
446 \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}{}}
447 \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}{}}
448 \DeclareOption{polutonikogreek}{%
449   \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}}
450 \DeclareOption{portuguese}{\bbl@try@load@lang{}{portuges}{}}
451 \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}{}}
452 \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}{}}
453 \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.

```

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

```

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

```

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

```

```

475      {\noexpand\bbload@language{\bb@tempa}}}%
476      \bb@tempb}%
477      \@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.

```

478 \bb@foreach\@classoptionslist{%
479   \bb@ifunset{ds@#1}%
480   {\IfFileExists{#1.ldf}%
481    {\DeclareOption{#1}{\bbload@language{#1}}}%
482    {}}%
483   {}}

```

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

```

484 \ifx\bbload@opt@main\@nnil\else
485   \expandafter
486   \let\expandafter\bbloadmain\csname ds@\bbload@opt@main\endcsname
487   \DeclareOption{\bbload@opt@main}{}
488 \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):

```

489 \def\AfterBabelLanguage#1{%
490   \bb@ifsamestring\CurrentOption{#1}{\global\bbadd\bb@afterlang}{}}
491 \DeclareOption*{}
492 \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.

```

493 \ifx\bbload@opt@main\@nnil
494   \edef\bbload@tempa{\@classoptionslist,\bbload@language@opts}
495   \let\bbload@tempc\@empty
496   \bb@for\bbload@tempb\bbload@tempa{%
497     \bbloadxin@{\bbload@tempb},{\bbload@loaded},}%
498     \ifin\edef\bbload@tempc{\bbload@tempb}\fi}
499   \def\bbload@tempa#1,#2\@nnil{\def\bbload@tempb{#1}}
500   \expandafter\bbload@tempa\bbload@loaded,\@nnil
501   \ifx\bbload@tempb\bbload@tempc\else
502     \bbload@warning{%
503       Last declared language option is '\bbload@tempc',\%
504       but the last processed one was '\bbload@tempb'.\%
505       The main language cannot be set as both a global\%
506       and a package option. Use 'main=\bbload@tempc' as\%
507       option. Reported}%
508   \fi
509 \else
510   \DeclareOption{\bbload@opt@main}{\bbload@loadmain}
511   \ExecuteOptions{\bbload@opt@main}
512   \DeclareOption*{}
513   \ProcessOptions*
514 \fi
515 \def\AfterBabelLanguage{%

```

```

516 \bbl@error
517   {Too late for \string\AfterBabelLanguage}%
518   {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.

```

519 \ifx\bbl@main@language\@undefined
520   \bbl@info{%
521     You haven't specified a language. I'll use 'nil'\%
522     as the main language. Reported}
523   \bbl@load@language{nil}
524 \fi
525 \</package>
526 \<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

```

527 \ifx\ldf@quit\@undefined
528 \else
529   \expandafter\endinput
530 \fi
531 \<<Make sure ProvidesFile is defined>>
532 \ProvidesFile{babel.def}[\<<date>>] \<<version>> Babel common definitions]
533 \<<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 an alternative mechanism is provided. For the moment, only `\babeloptionstrings` and `\babeloptionmath` are provided, which can be defined before loading babel.

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

```

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

```

```

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

And continue.
556 \ifx\bbl@switchflag\@undefined % Prevent double input
557 \let\bbl@switchflag\relax
558 \input switch.def\relax
559 \fi
560 \bbl@trace{Compatibility with language.def}
561 \ifx\bbl@languages\@undefined
562 \ifx\directlua\@undefined
563 \openin1 = language.def
564 \ifeof1
565 \closein1
566 \message{I couldn't find the file language.def}
567 \else
568 \closein1
569 \begingroup
570 \def\addlanguage#1#2#3#4#5{%
571 \expandafter\ifx\csname lang@#1\endcsname\relax\else
572 \global\expandafter\let\csname l@#1\endcsname
573 \csname lang@#1\endcsname
574 \fi}%
575 \def\uselanguage#1{}\}%
576 \input language.def
577 \endgroup
578 \fi
579 \fi
580 \chardef\l@english\z@
581 \fi
582 <<Load patterns in luatex>>
583 <<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.

```

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

```

```

588 \ifx#1\relax
589 \def#1{#2}%
590 \else
591 {\toks@\expandafter{#1#2}%
592 \xdef#1{\the\toks@}}%
593 \fi
594 \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.

```

595 \def\bbl@withactive#1#2{%
596 \begingroup
597 \lccode`~=#2\relax
598 \lowercase{\endgroup#1~}}

```

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

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

```

599 \def\bbl@redefine#1{%
600 \edef\bbl@tempa{\bbl@stripslash#1}%
601 \expandafter\let\csname org@\bbl@tempa\endcsname#1%
602 \expandafter\def\csname\bbl@tempa\endcsname}

```

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

```
603 \@onlypreamble\bbl@redefine
```

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

```

604 \def\bbl@redefine@long#1{%
605 \edef\bbl@tempa{\bbl@stripslash#1}%
606 \expandafter\let\csname org@\bbl@tempa\endcsname#1%
607 \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
608 \@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`.

```

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

```

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

```
616 \@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 command used by babel to execute hooks defined for an event.

```

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

```

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

```

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

```

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

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

```

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

```

```

655 \edef\bbl@tempa{\bbl@tempb#1\@empty}%
656 \def\bbl@tempb##1=##2\@{\@namedef{\bbl@ens@##1}{##2}}%
657 \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
658 \def\bbl@tempc{\bbl@ensure}%
659 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
660 \expandafter{\bbl@ens@include}}%
661 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
662 \expandafter{\bbl@ens@exclude}}%
663 \toks@\expandafter{\bbl@tempc}%
664 \bbl@exp{%
665 \endgroup
666 \def<bbl@e@#2>{\the\toks@{\bbl@ens@fontenc}}}%
667 \def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
668 \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
669 \ifx##1\undefined % 3.32 - Don't assume the macros exists
670 \edef##1{\noexpand\bbl@nocaption
671 {\bbl@stripslash##1}{\language\bbl@stripslash##1}}%
672 \fi
673 \ifx##1\@empty\else
674 \in@{##1}{#2}%
675 \ifin@ \else
676 \bbl@ifunset{\bbl@ensure@\language}%
677 {\bbl@exp{%
678 \\\DeclareRobustCommand<bbl@ensure@\language>[1]{%
679 \\\foreignlanguage{\language}%
680 {\ifx\relax#3\else
681 \\\fontencoding{#3}\selectfont
682 \fi
683 #####1}}}%
684 }%
685 \toks@\expandafter{##1}%
686 \edef##1{%
687 \bbl@csarg\noexpand{ensure@\language}%
688 {\the\toks@}}%
689 \fi
690 \expandafter\bbl@tempb
691 \fi}%
692 \expandafter\bbl@tempb\bbl@captionslist\today\@empty
693 \def\bbl@tempa##1{% elt for include list
694 \ifx##1\@empty\else
695 \bbl@csarg\in@{ensure@\language\expandafter}\expandafter{##1}%
696 \ifin@ \else
697 \bbl@tempb##1\@empty
698 \fi
699 \expandafter\bbl@tempa
700 \fi}%
701 \bbl@tempa#1\@empty}
702 \def\bbl@captionslist{%
703 \prefacename\refname\abstractname\bibname\chaptername\appendixname
704 \contentsname\listfigurename\listtablename\indexname\figurename
705 \tablename\partname\encname\ccname\headtoname\pagename\seename
706 \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 @-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.

```

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

```

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

```

734 \def\ldf@quit#1{%
735   \expandafter\main@language\expandafter{#1}%
736   \catcode`\@=\atcatcode \let\atcatcode\relax
737   \catcode`\==\eqcatcode \let\eqcatcode\relax
738   \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.



```

739 \def\bbl@afterldf#1{%
740   \bbl@afterlang
741   \let\bbl@afterlang\relax
742   \let\BabelModifiers\relax
743   \let\bbl@screset\relax}%
744 \def\ldf@finish#1{%
745   \loadlocalcfg{#1}%
746   \bbl@afterldf{#1}%
747   \expandafter\main@language\expandafter{#1}%
748   \catcode`\@=\atcatcode \let\atcatcode\relax
749   \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`.

```

750 \@onlypreamble\LdfInit
751 \@onlypreamble\ldf@quit
752 \@onlypreamble\ldf@finish

```

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

```

753 \def\main@language#1{%
754   \def\bbl@main@language{#1}%
755   \let\language\bbl@main@language
756   \bbl@id@assign
757   \chardef\localeid\@nameuse{bbl@id@\@language}%
758   \bbl@patterns{\language}}

```

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

```

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

```

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

```

776 \def\select@language@x#1{%
777   \ifcase\bbl@select@type
778     \bbl@ifsamestring\language{#1}{\select@language{#1}}%
779   \else
780     \select@language{#1}%
781   \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  $\TeX$  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.

```
782 \bbl@trace{Shorhands}
783 \def\bbl@add@special#1{% 1:a macro like "\", \?, etc.
784   \bbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
785   \bbl@ifunset{@sanitize}{\bbl@add\@sanitize{\@makeother#1}}%
786   \ifx\nfss@catcodes\undefined\else % TODO - same for above
787     \begingroup
788       \catcode`#1\active
789       \nfss@catcodes
790       \ifnum\catcode`#1=\active
791         \endgroup
792         \bbl@add\nfss@catcodes{\@makeother#1}%
793       \else
794         \endgroup
795       \fi
796   \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.

```
797 \def\bbl@remove@special#1{%
798   \begingroup
799   \def\x##1##2{\ifnum`#1=`##2\noexpand\@empty
800     \else\noexpand##1\noexpand##2\fi}%
801   \def\do{\x\do}%
802   \def\@makeother{\x\@makeother}%
803   \edef\x{\endgroup
804     \def\noexpand\dospecials{\dospecials}%
805     \expandafter\ifx\cename @sanitize\endcename\relax\else
806       \def\noexpand\@sanitize{\@sanitize}%
807     \fi}%
808   \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).

```

809 \def\bbl@active@def#1#2#3#4{%
810   \@namedef{#3#1}{%
811     \expandafter\ifx\csname#2@sh@#1\endcsname\relax
812       \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
813     \else
814       \bbl@afterfi\csname#2@sh@#1\endcsname
815     \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.

```

816   \long\@namedef{#3@arg#1}##1{%
817     \expandafter\ifx\csname#2@sh@#1\string##1\endcsname\relax
818       \bbl@afterelse\csname#4#1\endcsname##1%
819     \else
820       \bbl@afterfi\csname#2@sh@#1\string##1\endcsname
821     \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.

```

822 \def\initiate@active@char#1{%
823   \bbl@ifunset{active@char\string#1}%
824   {\bbl@withactive
825     {\expandafter\@initiate@active@char\expandafter}#1\string#1}%
826   {}

```

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).

```

827 \def\@initiate@active@char#1#2#3{%
828   \bbl@csarg\edef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%
829   \ifx#1\@undefined
830     \bbl@csarg\edef{oridef@#2}{\let\noexpand#1\noexpand\@undefined}%
831   \else
832     \bbl@csarg\let{oridef@#2}#1%
833     \bbl@csarg\edef{oridef@#2}{%
834       \let\noexpand#1
835       \expandafter\noexpand\csname bbl@oridef@#2\endcsname}%
836   \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*").

```

837   \ifx#1#3\relax
838     \expandafter\let\csname normal@char#2\endcsname#3%
839   \else
840     \bbl@info{Making #2 an active character}%
841     \ifnum\mathcode`#2=\ifodd\bbl@engine"1000000 \else"8000 \fi
842     \@namedef{normal@char#2}{%
843       \textormath{#3}{\csname bbl@oridef@#2\endcsname}}%
844   \else
845     \@namedef{normal@char#2}{#3}%

```

846     \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).

```
847     \bbl@restoreactive{#2}%
848     \AtBeginDocument{%
849         \catcode`#2\active
850         \if@filesw
851             \immediate\write\@mainaux{\catcode`\string#2\active}%
852         \fi}%
853     \expandafter\bbl@add@special\csname#2\endcsname
854     \catcode`#2\active
855     \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⟩).

```
856     \let\bbl@tempa\@firstoftwo
857     \if\string^#2%
858         \def\bbl@tempa{\noexpand\textormath}%
859     \else
860         \ifx\bbl@mathnormal\@undefined\else
861             \let\bbl@tempa\bbl@mathnormal
862         \fi
863     \fi
864     \expandafter\edef\csname active@char#2\endcsname{%
865         \bbl@tempa
866         {\noexpand\if@safe@actives
867             \noexpand\expandafter
868             \expandafter\noexpand\csname normal@char#2\endcsname
869             \noexpand\else
870                 \noexpand\expandafter
871                 \expandafter\noexpand\csname bbl@doactive#2\endcsname
872             \noexpand\fi}%
873         {\expandafter\noexpand\csname normal@char#2\endcsname}}%
874     \bbl@csarg\edef{doactive#2}{%
875         \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!).

```
876     \bbl@csarg\edef{active@#2}{%
877         \noexpand\active@prefix\noexpand#1%
878         \expandafter\noexpand\csname active@char#2\endcsname}%
879     \bbl@csarg\edef{normal@#2}{%
880         \noexpand\active@prefix\noexpand#1%
881         \expandafter\noexpand\csname normal@char#2\endcsname}%
882     \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.

```
883 \bbl@active@def#2\user@group{user@active}{language@active}%
884 \bbl@active@def#2\language@group{language@active}{system@active}%
885 \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.

```
886 \expandafter\edef\csname\user@group @sh@#2@@\endcsname
887 {\expandafter\noexpand\csname normal@char#2\endcsname}%
888 \expandafter\edef\csname\user@group @sh@#2@string\protect\endcsname
889 {\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.

```
890 \if\string'#2%
891 \let\prim@s\bbl@prim@s
892 \let\active@math@prime#1%
893 \fi
894 \bbl@usehooks{initiateactive}{\{#1\}{#2\}{#3\}}
```

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

```
895 <<*More package options>> ≡
896 \DeclareOption{math=active}{}
897 \DeclareOption{math=normal}{\def\bbl@mathnormal{\noexpand\textormath}}
898 <</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*.

```
899 \@ifpackagewith{babel}{KeepShorthandsActive}%
900 {\let\bbl@restoreactive\@gobble}%
901 {\def\bbl@restoreactive#1{%
902   \bbl@exp{%
903     \\AfterBabelLanguage\\CurrentOption
904     {\catcode`#1=\the\catcode`#1\relax}%
905     \\AtEndOfPackage
906     {\catcode`#1=\the\catcode`#1\relax}}}%
907   \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.

```
908 \def\bbl@sh@select#1#2{%
909   \expandafter\ifx\csname#1sh@#2@sel\endcsname\relax
910     \bbl@afterelse\bbl@scndcs
911   \else
```

```

912 \bbl@afterfi\csname#1@sh@#2@sel\endcsname
913 \fi}

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

914 \begingroup
915 \bbl@ifunset{ifincsname}%
916 {\gdef\active@prefix#1{%
917 \ifx\protect\@typeset@protect
918 \else
919 \ifx\protect\@unexpandable@protect
920 \noexpand#1%
921 \else
922 \protect#1%
923 \fi
924 \expandafter\@gobble
925 \fi}}
926 {\gdef\active@prefix#1{%
927 \ifincsname
928 \string#1%
929 \expandafter\@gobble
930 \else
931 \ifx\protect\@typeset@protect
932 \else
933 \ifx\protect\@unexpandable@protect
934 \noexpand#1%
935 \else
936 \protect#1%
937 \fi
938 \expandafter\expandafter\expandafter\@gobble
939 \fi
940 \fi}}
941 \endgroup

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

942 \newif\if@safe@actives
943 \@safe@activesfalse

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

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

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

945 \def\bbl@activate#1{%
946 \bbl@withactive{\expandafter\let\expandafter}#1%
947 \csname bbl@active@\string#1\endcsname}
948 \def\bbl@deactivate#1{%
949 \bbl@withactive{\expandafter\let\expandafter}#1%
950 \csname bbl@normal@\string#1\endcsname}

```

<code>\bbl@firstcs</code> <code>\bbl@scndcs</code>	<p>These macros have two arguments. They use one of their arguments to build a control sequence from.</p> <pre> 951 \def\bbl@firstcs#1#2{\csname#1\endcsname} 952 \def\bbl@scndcs#1#2{\csname#2\endcsname} </pre>
<code>\declare@shorthand</code>	<p>The command <code>\declare@shorthand</code> is used to declare a shorthand on a certain level. It takes three arguments:</p> <ol style="list-style-type: none"> <li>1. a name for the collection of shorthands, i.e. ‘system’, or ‘dutch’;</li> <li>2. the character (sequence) that makes up the shorthand, i.e. <code>~</code> or <code>"a</code>;</li> <li>3. the code to be executed when the shorthand is encountered.</li> </ol> <pre> 953 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil} 954 \def\@decl@short#1#2#3\@nil#4{% 955   \def\bbl@tempa{#3}% 956   \ifx\bbl@tempa\@empty 957     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs 958     \bbl@ifunset{#1@sh@\string#2@}\}% 959     {\def\bbl@tempa{#4}% 960      \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa 961      \else 962        \bbl@info 963        {Redefining #1 shorthand \string#2\\ 964         in language \CurrentOption}% 965        \fi}% 966     \@namedef{#1@sh@\string#2@}{#4}% 967   \else 968     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs 969     \bbl@ifunset{#1@sh@\string#2@\string#3@}\}% 970     {\def\bbl@tempa{#4}% 971      \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa 972      \else 973        \bbl@info 974        {Redefining #1 shorthand \string#2\string#3\\ 975         in language \CurrentOption}% 976        \fi}% 977     \@namedef{#1@sh@\string#2@\string#3@}{#4}% 978   \fi} </pre>
<code>\textormath</code>	<p>Some of the shorthands that will be declared by the language definition files have to be usable in both text and mathmode. To achieve this the helper macro <code>\textormath</code> is provided.</p> <pre> 979 \def\textormath{% 980   \ifmmode 981     \expandafter\@secondoftwo 982   \else 983     \expandafter\@firstoftwo 984   \fi} </pre>
<code>\user@group</code> <code>\language@group</code> <code>\system@group</code>	<p>The current concept of ‘shorthands’ supports three levels or groups of shorthands. For each level the name of the level or group is stored in a macro. The default is to have a user group; use language group ‘english’ and have a system group called ‘system’.</p> <pre> 985 \def\user@group{user} 986 \def\language@group{english} 987 \def\system@group{system} </pre>

`\useshortands` 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.

```

988 \def\useshortands{%
989   \@ifstar\bb1@usesesh@s{\bb1@usesesh@x{}}
990 \def\bb1@usesesh@s#1{%
991   \bb1@usesesh@x
992   {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bb1@activate{#1}}}%
993   {#1}}
994 \def\bb1@usesesh@x#1#2{%
995   \bb1@ifshorthand{#2}%
996   {\def\user@group{user}%
997    \initiate@active@char{#2}%
998    #1%
999    \bb1@activate{#2}}%
1000  {\bb1@error
1001   {Cannot declare a shorthand turned off (\string#2)}
1002   {Sorry, but you cannot use shorthands which have been\%
1003    turned off in the package options}}}
```

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

```

1004 \def\user@language@group{user@\language@group}
1005 \def\bb1@set@user@generic#1#2{%
1006   \bb1@ifunset{user@generic@active#1}%
1007   {\bb1@active@def#1\user@language@group{user@active}{user@generic@active}%
1008    \bb1@active@def#1\user@group{user@generic@active}{language@active}%
1009    \expandafter\edef\csname#2@sh@#1@@\endcsname{%
1010     \expandafter\noexpand\csname normal@char#1\endcsname}%
1011     \expandafter\edef\csname#2@sh@#1@\string\protect\endcsname{%
1012      \expandafter\noexpand\csname user@active#1\endcsname}}%
1013   \@empty}
1014 \newcommand\defineshorthand[3][user]{%
1015   \edef\bb1@tempa{\zap@space#1 \@empty}%
1016   \bb1@for\bb1@tempb\bb1@tempa{%
1017     \if*\expandafter\car\bb1@tempb\@nil
1018       \edef\bb1@tempb{user\expandafter@gobble\bb1@tempb}%
1019       \@expandtwoargs
1020       \bb1@set@user@generic{\expandafter\string\car#2\@nil}\bb1@tempb
1021     \fi
1022     \declare@shorthand{\bb1@tempb}{#2}{#3}}}
```

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

```

1023 \def\languageshortands#1{\def\language@group{#1}}
```

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

```

1024 \def\aliasshorthand#1#2{%
1025   \bb1@ifshorthand{#2}%
1026   {\expandafter\ifx\csname active@char\string#2\endcsname\relax
1027     \ifx\document\@notprerr
```



```

1028      \@notshorthand{#2}%
1029      \else
1030      \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".

1031      \expandafter\let\csname active@char\string#2\expandafter\endcsname
1032      \csname active@char\string#1\endcsname
1033      \expandafter\let\csname normal@char\string#2\expandafter\endcsname
1034      \csname normal@char\string#1\endcsname
1035      \bbl@activate{#2}%
1036      \fi
1037      \fi}%
1038      {\bbl@error
1039      {Cannot declare a shorthand turned off (\string#2)}
1040      {Sorry, but you cannot use shorthands which have been\\%
1041      turned off in the package options}}}

```

\@notshorthand

```

1042 \def\@notshorthand#1{%
1043   \bbl@error{%
1044     The character '\string #1' should be made a shorthand character;\\%
1045     add the command \string\usesshorthands\string{#1\string} to
1046     the preamble.\\%
1047     I will ignore your instruction}%
1048   {You may proceed, but expect unexpected results}}

```

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

```

1049 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
1050 \DeclareRobustCommand*\shorthandoff{%
1051   \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
1052 \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.

```

1053 \def\bbl@switch@sh#1#2{%
1054   \ifx#2\@nnil\else
1055     \bbl@ifunset{\bbl@active@\string#2}%
1056     {\bbl@error
1057       {I cannot switch '\string#2' on or off--not a shorthand}%
1058       {This character is not a shorthand. Maybe you made\\%
1059       a typing mistake? I will ignore your instruction}}%
1060     {\ifcase#1%
1061       \catcode`#2\relax
1062       \or
1063       \catcode`#2\active
1064       \or
1065       \csname bbl@oricat@\string#2\endcsname

```

```

1066      \csname bbl@oridef@string#2\endcsname
1067      \fi}%
1068      \bbl@afterfi\bbl@switch@sh#1%
1069      \fi}

```

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

```

1070 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
1071 \def\bbl@putsh#1{%
1072   \bbl@ifunset{\bbl@active@string#1}%
1073   {\bbl@putsh@i#1\empty\@nnil}%
1074   {\csname bbl@active@string#1\endcsname}}
1075 \def\bbl@putsh@i#1#2\@nnil{%
1076   \csname\language @sh@string#1@%
1077     \ifx\empty#2\else string#2\fi\endcsname}
1078 \ifx\bbl@opt@shorthands\@nnil\else
1079   \let\bbl@s@initiate@active@char\initiate@active@char
1080   \def\initiate@active@char#1{%
1081     \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
1082   \let\bbl@s@switch@sh\bbl@switch@sh
1083   \def\bbl@switch@sh#1#2{%
1084     \ifx#2\@nnil\else
1085       \bbl@afterfi
1086       \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
1087       \fi}
1088   \let\bbl@s@activate\bbl@activate
1089   \def\bbl@activate#1{%
1090     \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
1091   \let\bbl@s@deactivate\bbl@deactivate
1092   \def\bbl@deactivate#1{%
1093     \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
1094 \fi

```

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

```

1095 \newcommand\ifbabelshorthand[3]{\bbl@ifunset{\bbl@active@string#1}{#3}{#2}}

```

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

```

1096 \def\bbl@prim@s{%
1097   \prime\futurelet\@let@token\bbl@pr@m@s}
1098 \def\bbl@if@primes#1#2{%
1099   \ifx#1\@let@token
1100     \expandafter\@firstoftwo
1101   \else\ifx#2\@let@token
1102     \bbl@afterelse\expandafter\@firstoftwo
1103   \else
1104     \bbl@afterfi\expandafter\@secondoftwo
1105   \fi\fi}
1106 \begingroup
1107   \catcode`\^=7 \catcode`\*=\active \lccode`\*=\^
1108   \catcode`\'=12 \catcode`\"=\active \lccode`\"=\'
1109   \lowercase{%
1110     \gdef\bbl@pr@m@s{%
1111       \bbl@if@primes" '%
1112       \pr@@@s

```

```

1113      {\bbl@if@primes*^{\pr@@@t\egroup}}
1114 \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).

```

1115 \initiate@active@char{~}
1116 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
1117 \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.

```

1118 \expandafter\def\csname OT1dqpos\endcsname{127}
1119 \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

```

1120 \ifx\f@encoding\@undefined
1121   \def\f@encoding{OT1}
1122 \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.

```

1123 \bbl@trace{Language attributes}
1124 \newcommand\languageattribute[2]{%
1125   \def\bbl@tempc{#1}%
1126   \bbl@fixname\bbl@tempc
1127   \bbl@iflanguage\bbl@tempc{%
1128     \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.

```

1129     \ifx\bbl@known@attribs\@undefined
1130       \in@false
1131     \else

```

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

```

1132       \bbl@xin@{\bbl@tempc-##1,}{\bbl@known@attribs,}%
1133     \fi

```

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

```

1134     \ifin@
1135       \bbl@warning{%
1136         You have more than once selected the attribute '##1'\%
1137         for language #1. Reported}%
1138     \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.

```

1139     \bbl@exp{%
1140         \\bbl@add@list\\bbl@known@attribs{\bbl@tempc-##1}}%
1141     \edef\bbl@tempa{\bbl@tempc-##1}%
1142     \expandafter\bbl@ifknown@ttrib\expandafter{\bbl@tempa}\bbl@attributes%
1143     {\csname\bbl@tempc @attr@##1\endcsname}%
1144     {\@attrerr{\bbl@tempc}{##1}}%
1145     \fi}}

```

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

```

1146 \@onlypreamble\languageattribute

```

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

```

1147 \newcommand*{\@attrerr}[2]{%
1148     \bbl@error
1149     {The attribute #2 is unknown for language #1.}%
1150     {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}`.

```

1151 \def\bbl@declare@ttribute#1#2#3{%
1152     \bbl@xin@{,#2,}{,\BabelModifiers,}%
1153     \ifin@
1154         \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
1155     \fi
1156     \bbl@add@list\bbl@attributes{#1-#2}%
1157     \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.

```

1158 \def\bbl@ifattributeset#1#2#3#4{%

```

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

```

1159     \ifx\bbl@known@attribs\@undefined
1160         \in@false
1161     \else

```

The we need to check the list of known attributes.

```

1162     \bbl@xin@{,#1-#2,}{,\bbl@known@attribs,}%
1163     \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`'.

```

1164     \ifin@
1165         \bbl@afterelse#3%
1166     \else
1167         \bbl@afterfi#4%
1168     \fi
1169 }

```

`\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.

```
1170 \def\bbl@ifknown@ttrib#1#2{%
```

We first assume the attribute is unknown.

```
1171 \let\bbl@tempa\@secondoftwo
```

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

```
1172 \bbl@loopx\bbl@tempb{#2}{%
```

```
1173 \expandafter\in@\expandafter{\expandafter,\bbl@tempb,}{, #1,}%
```

```
1174 \ifin@
```

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

```
1175 \let\bbl@tempa\@firstoftwo
```

```
1176 \else
```

```
1177 \fi}%
```

Finally we execute `\bbl@tempa`.

```
1178 \bbl@tempa
```

```
1179 }
```

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

```
1180 \def\bbl@clear@ttribs{%
```

```
1181 \ifx\bbl@attributes\undefined\else
```

```
1182 \bbl@loopx\bbl@tempa{\bbl@attributes}{%
```

```
1183 \expandafter\bbl@clear@ttrib\bbl@tempa.
```

```
1184 }%
```

```
1185 \let\bbl@attributes\undefined
```

```
1186 \fi}
```

```
1187 \def\bbl@clear@ttrib#1-#2.{%
```

```
1188 \expandafter\let\csname#1@attr#2\endcsname\undefined}
```

```
1189 \AtBeginDocument{\bbl@clear@ttribs}
```

## 9.6 Support for saving macro definitions

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

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

```
\babel@beginsave 1190 \bbl@trace{Macros for saving definitions}
```

```
1191 \def\babel@beginsave{\babel@savecnt\z@}
```

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

```
1192 \newcount\babel@savecnt
```

```
1193 \babel@beginsave
```

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

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

sequence, the restore commands are appended to `\originalTeX` and the counter is incremented.

```
1194 \def\babel@save#1{%
1195   \expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax
1196   \toks@\expandafter{\originalTeX\let#1=}%
1197   \bbl@exp{%
1198     \def\\originalTeX{\the\toks@<\babel@number\babel@savecnt>\relax}}%
1199   \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.

```
1200 \def\babel@savevariable#1{%
1201   \toks@\expandafter{\originalTeX #1=}%
1202   \bbl@exp{\def\\originalTeX{\the\toks@\the#1\relax}}}
```

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

```
1203 \def\bbl@frenchspacing{%
1204   \ifnum\the\sfcodes\@m
1205     \let\bbl@nonfrenchspacing\relax
1206   \else
1207     \frenchspacing
1208     \let\bbl@nonfrenchspacing\nonfrenchspacing
1209   \fi}
1210 \let\bbl@nonfrenchspacing\nonfrenchspacing
```

## 9.7 Short tags

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

```
1211 \bbl@trace{Short tags}
1212 \def\babeltags#1{%
1213   \edef\bbl@tempa{\zap@space#1 \@empty}%
1214   \def\bbl@tempb##1=##2\@{#}%
1215   \edef\bbl@tempc{%
1216     \noexpand\newcommand
1217     \expandafter\noexpand\csname ##1\endcsname{%
1218       \noexpand\protect
1219       \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
1220     \noexpand\newcommand
1221     \expandafter\noexpand\csname text##1\endcsname{%
1222       \noexpand\foreignlanguage{##2}}
1223     \bbl@tempc}%
1224   \bbl@for\bbl@tempa\bbl@tempa{%
1225     \expandafter\bbl@tempb\bbl@tempa\@{#}}
```

## 9.8 Hyphens

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

```
1226 \bbl@trace{Hyphens}
```

```

1227 \@onlypreamble\babelhyphenation
1228 \AtEndOfPackage{%
1229   \newcommand\babelhyphenation[2][\@empty]{%
1230     \ifx\bbbl@hyphenation@relax
1231       \let\bbbl@hyphenation@\@empty
1232     \fi
1233     \ifx\bbbl@hyphlist\@empty\else
1234       \bbbl@warning{%
1235         You must not intermingle \string\selectlanguage\space and\\%
1236         \string\babelhyphenation\space or some exceptions will not\\%
1237         be taken into account. Reported}%
1238     \fi
1239     \ifx\@empty#1%
1240       \protected@edef\bbbl@hyphenation@{\bbbl@hyphenation@\space#2}%
1241     \else
1242       \bbbl@vforeach{#1}{%
1243         \def\bbbl@tempa{##1}%
1244         \bbbl@fixname\bbbl@tempa
1245         \bbbl@iflanguage\bbbl@tempa{%
1246           \bbbl@csarg\protected@edef{hyphenation@\bbbl@tempa}{%
1247             \bbbl@ifunset{bbbl@hyphenation@\bbbl@tempa}%
1248             \@empty
1249             {\csname bbl@hyphenation@\bbbl@tempa\endcsname\space}%
1250             #2}}}%
1251       \fi}}

```

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

```

1252 \def\bbbl@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}
1253 \def\bbbl@t@one{T1}
1254 \def\allowhyphens{\ifx\cf@encoding\bbbl@t@one\else\bbbl@allowhyphens\fi}

```

`\babelhyphen` Macros to insert common hyphens. Note the space before `@` in `\babelhyphen`. Instead of protecting it with `\DeclareRobustCommand`, which could insert a `\relax`, we use the same procedure as shorthands, with `\active@prefix`.

```

1255 \newcommand\babelnullhyphen{\char\hyphenchar\font}
1256 \def\babelhyphen{\active@prefix\babelhyphen\bbbl@hyphen}
1257 \def\bbbl@hyphen{%
1258   \@ifstar{\bbbl@hyphen@i @}{\bbbl@hyphen@i \@empty}}
1259 \def\bbbl@hyphen@i#1#2{%
1260   \bbbl@ifunset{bbbl@hy#1#2\@empty}%
1261   {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{-}{#2}}}%
1262   {\csname bbl@hy#1#2\@empty\endcsname}}

```

The following two commands are used to wrap the “hyphen” and set the behavior of the rest of the word – the version with a single `@` is used when further hyphenation is allowed, while that with `@@` if no more hyphens are allowed. In both cases, if the hyphen is preceded by a positive space, breaking after the hyphen is disallowed.

There should not be a discretionary after a hyphen at the beginning of a word, so it is prevented if preceded by a skip. Unfortunately, this does handle cases like “(-suffix)”. `\nobreak` is always preceded by `\leavevmode`, in case the shorthand starts a paragraph.

```

1263 \def\bbbl@usehyphen#1{%
1264   \leavevmode
1265   \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
1266   \nobreak\hskip\z@skip}
1267 \def\bbbl@@usehyphen#1{%
1268   \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

<sup>33</sup> $\mathrm{T}_{\mathrm{E}}\mathrm{X}$  begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.

The following macro inserts the hyphen char.

```
1269 \def\bbl@hyphenchar{%
1270   \ifnum\hyphenchar\font=\m@ne
1271     \babeinullhyphen
1272   \else
1273     \char\hyphenchar\font
1274   \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.

```
1275 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1276 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1277 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
1278 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
1279 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}
1280 \def\bbl@hy@nobreak{\mbox{\bbl@hyphenchar}}
1281 \def\bbl@hy@repeat{%
1282   \bbl@usehyphen{%
1283     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
1284 \def\bbl@hy@repeat{%
1285   \bbl@usehyphen{%
1286     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
1287 \def\bbl@hy@empty{\hskip\z@skip}
1288 \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.

```
1289 \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.

```
1290 \bbl@trace{Multiencoding strings}
1291 \def\bbl@tglobal#1{\global\let#1#1}
1292 \def\bbl@recatcode#1{%
1293   \@tempcnta="7F
1294   \def\bbl@tempa{%
1295     \ifnum\@tempcnta>"FF\else
1296       \catcode\@tempcnta=#1\relax
1297       \advance\@tempcnta\@ne
1298       \expandafter\bbl@tempa
1299     \fi}%
1300   \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).

```
1301 \ifpackagewith{babel}{nocase}%
1302   {\let\bbl@patchucl\relax}%
1303   {\def\bbl@patchucl{%
1304     \global\let\bbl@patchucl\relax
1305     \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@ucl}}%
1306     \gdef\bbl@ucl##1{%
1307       \let\bbl@encoded\bbl@encoded@ucl
1308       \bbl@ifunset{\language @bbl@ucl}% and resumes it
1309       {##1}%
1310       {\let\bbl@tempa##1\relax % Used by LANG@bbl@ucl
1311         \csname\language @bbl@ucl\endcsname}%
1312       {\bbl@tolower\@empty}{\bbl@toupper\@empty}}}%
1313   \gdef\bbl@tolower{\csname\language @bbl@lc\endcsname}%
1314   \gdef\bbl@toupper{\csname\language @bbl@uc\endcsname}}%
1315 <<(*More package options)>> ≡
1316 \DeclareOption{nocase}{}
1317 <</More package options>>
```

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

```
1318 <<(*More package options)>> ≡
1319 \let\bbl@opt@strings\@nnil % accept strings=value
1320 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1321 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1322 \def\BabelStringsDefault{generic}
1323 <</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.

```
1324 \@onlypreamble\StartBabelCommands
1325 \def\StartBabelCommands{%
1326   \begingroup
1327   \bbl@recatcode{11}%
1328   <<Macros local to BabelCommands>>
1329   \def\bbl@provstring##1##2{%
1330     \providecommand##1{##2}%
1331     \bbl@tglobal##1}%
1332   \global\let\bbl@scafter\@empty
1333   \let\StartBabelCommands\bbl@startcmds
1334   \ifx\BabelLanguages\relax
1335     \let\BabelLanguages\CurrentOption
1336   \fi
1337   \begingroup
1338   \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
1339   \StartBabelCommands}
1340 \def\bbl@startcmds{%
1341   \ifx\bbl@screset\@nnil\else
1342     \bbl@usehooks{stopcommands}{}%
1343   \fi
1344   \endgroup
1345   \begingroup
1346   \@ifstar
1347   {\ifx\bbl@opt@strings\@nnil
```

```

1348     \let\bbl@opt@strings\BabelStringsDefault
1349     \fi
1350     \bbl@startcmds@i}%
1351     \bbl@startcmds@i}
1352 \def\bbl@startcmds@i#1#2{%
1353   \edef\bbl@L{\zap@space#1 \@empty}%
1354   \edef\bbl@G{\zap@space#2 \@empty}%
1355   \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.

```

1356 \newcommand\bbl@startcmds@ii[1][\@empty]{%
1357   \let\SetString@gobbletwo
1358   \let\bbl@stringdef@gobbletwo
1359   \let\AfterBabelCommands@gobble
1360   \ifx\@empty#1%
1361     \def\bbl@sc@label{generic}%
1362     \def\bbl@encstring##1##2{%
1363       \ProvideTextCommandDefault##1{##2}%
1364       \bbl@tglobal##1%
1365       \expandafter\bbl@tglobal\csname\string?\string##1\endcsname}%
1366       \let\bbl@sctest\in@true
1367     \else
1368       \let\bbl@sc@charset\space % <- zapped below
1369       \let\bbl@sc@fontenc\space % <- " "
1370       \def\bbl@tempa##1=##2\@nil{%
1371         \bbl@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
1372         \bbl@vforeach{label=#1}{\bbl@tempa##1\@nil}%
1373         \def\bbl@tempa##1 ##2{% space -> comma
1374           ##1%
1375           \ifx\@empty##2\else\ifx,##1,\else,\fi\bbl@afterfi\bbl@tempa##2\fi}%
1376         \edef\bbl@sc@fontenc{\expandafter\bbl@tempa\bbl@sc@fontenc\@empty}%
1377         \edef\bbl@sc@label{\expandafter\zap@space\bbl@sc@label\@empty}%
1378         \edef\bbl@sc@charset{\expandafter\zap@space\bbl@sc@charset\@empty}%
1379         \def\bbl@encstring##1##2{%
1380           \bbl@foreach\bbl@sc@fontenc{%
1381             \bbl@ifunset{T@####1}%
1382             {}%
1383             {\ProvideTextCommand##1{####1}{##2}%
1384             \bbl@tglobal##1%
1385             \expandafter
1386             \bbl@tglobal\csname####1\string##1\endcsname}}}%
1387         \def\bbl@sctest{%
1388           \bbl@xin@{\bbl@opt@strings,}{\bbl@sc@label,\bbl@sc@fontenc,}}%
1389       \fi
1390       \ifx\bbl@opt@strings\@nnil % ie, no strings key -> defaults
1391       \else\ifx\bbl@opt@strings\relax % ie, strings=encoded
1392         \let\AfterBabelCommands\bbl@aftercmds
1393         \let\SetString\bbl@setstring
1394         \let\bbl@stringdef\bbl@encstring

```

```

1395 \else % ie, strings=value
1396 \bbl@sctest
1397 \ifin@
1398 \let\AfterBabelCommands\bbl@aftercmds
1399 \let\SetString\bbl@setstring
1400 \let\bbl@stringdef\bbl@provstring
1401 \fi\fi\fi
1402 \bbl@scswitch
1403 \ifx\bbl@G\@empty
1404 \def\SetString##1##2{%
1405 \bbl@error{Missing group for string \string##1}%
1406 {You must assign strings to some category, typically\\%
1407 captions or extras, but you set none}}%
1408 \fi
1409 \ifx\@empty#1%
1410 \bbl@usehooks{defaultcommands}{}%
1411 \else
1412 \@expandtwoargs
1413 \bbl@usehooks{encodedcommands}{\bbl@sc@charset}{\bbl@sc@fontenc}}%
1414 \fi}

```

There are two versions of `\bbl@scswitch`. The first version is used when `ldfs` are read, and it makes sure `\langle group \rangle \langle language \rangle` is reset, but only once (`\bbl@screset` is used to keep track of this). The second version is used in the preamble and packages loaded after `babel` and does nothing. The macro `\bbl@forlang` loops `\bbl@L` but its body is executed only if the value is in `\BabelLanguages` (inside `babel`) or `\date \langle language \rangle` is defined (after `babel` has been loaded). There are also two version of `\bbl@forlang`. The first one skips the current iteration if the language is not in `\BabelLanguages` (used in `ldfs`), and the second one skips undefined languages (after `babel` has been loaded).

```

1415 \def\bbl@forlang#1#2{%
1416 \bbl@for#1\bbl@L{%
1417 \bbl@xin@{, #1,}{, \BabelLanguages,}%
1418 \ifin@#2\relax\fi}}
1419 \def\bbl@scswitch{%
1420 \bbl@forlang\bbl@tempa{%
1421 \ifx\bbl@G\@empty\else
1422 \ifx\SetString\@gobbletwo\else
1423 \edef\bbl@GL{\bbl@G\bbl@tempa}%
1424 \bbl@xin@{, \bbl@GL,}{, \bbl@screset,}%
1425 \ifin@\else
1426 \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
1427 \xdef\bbl@screset{\bbl@screset, \bbl@GL}%
1428 \fi
1429 \fi
1430 \fi}}
1431 \AtEndOfPackage{%
1432 \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{}{#2}}}%
1433 \let\bbl@scswitch\relax}
1434 \@onlypreamble\EndBabelCommands
1435 \def\EndBabelCommands{%
1436 \bbl@usehooks{stopcommands}{}%
1437 \endgroup
1438 \endgroup
1439 \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.

```

1440 \def\bbl@setstring#1#2{%
1441   \bbl@forlang\bbl@tempa{%
1442     \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
1443     \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
1444     {\global\expandafter % TODO - con \bbl@exp ?
1445       \bbl@add\csname\bbl@G\bbl@tempa\expandafter\endcsname\expandafter
1446       {\expandafter\bbl@scset\expandafter#1\csname\bbl@LC\endcsname}}%
1447     }%
1448   \def\BabelString{#2}%
1449   \bbl@usehooks{stringprocess}{}%
1450   \expandafter\bbl@stringdef
1451     \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`.

```

1452 \ifx\bbl@opt@strings\relax
1453   \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
1454   \bbl@patchuclc
1455   \let\bbl@encoded\relax
1456   \def\bbl@encoded@uclc#1{%
1457     \@inmathwarn#1%
1458     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
1459       \expandafter\ifx\csname ?\string#1\endcsname\relax
1460         \TextSymbolUnavailable#1%
1461       \else
1462         \csname ?\string#1\endcsname
1463       \fi
1464     \else
1465       \csname\cf@encoding\string#1\endcsname
1466     \fi}
1467 \else
1468   \def\bbl@scset#1#2{\def#1{#2}}
1469 \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.

```

1470 <<*Macros local to BabelCommands>> ≡
1471 \def\SetStringLoop##1##2{%
1472   \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
1473   \count@\z@
1474   \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
1475     \advance\count@\@ne
1476     \toks@\expandafter{\bbl@tempa}%
1477     \bbl@exp{%
1478       \\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
1479       \count@=\the\count@\relax}}%
1480 <</Macros local to BabelCommands>>

```

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

```

1481 \def\bbl@aftercmds#1{%
1482   \toks@\expandafter{\bbl@scafter#1}%
1483   \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.

```

1484 <<(*Macros local to BabelCommands)>> ≡
1485   \newcommand\SetCase[3][]{%
1486     \bbl@patchuclc
1487     \bbl@forlang\bbl@tempa{%
1488       \expandafter\bbl@encstring
1489       \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
1490       \expandafter\bbl@encstring
1491       \csname\bbl@tempa @bbl@uc\endcsname{##2}%
1492       \expandafter\bbl@encstring
1493       \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
1494 <</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.

```

1495 <<(*Macros local to BabelCommands)>> ≡
1496   \newcommand\SetHyphenMap[1]{%
1497     \bbl@forlang\bbl@tempa{%
1498       \expandafter\bbl@stringdef
1499       \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
1500 <</Macros local to BabelCommands)>>

```

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

```

1501 \newcommand\BabelLower[2]{% one to one.
1502   \ifnum\lccode#1=#2\else
1503     \babel@savevariable{\lccode#1}%
1504     \lccode#1=#2\relax
1505   \fi}
1506 \newcommand\BabelLowerMM[4]{% many-to-many
1507   \@tempcnta=#1\relax
1508   \@tempcntb=#4\relax
1509   \def\bbl@tempa{%
1510     \ifnum\@tempcnta>#2\else
1511       \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
1512       \advance\@tempcnta#3\relax
1513       \advance\@tempcntb#3\relax
1514       \expandafter\bbl@tempa
1515     \fi}%
1516   \bbl@tempa}
1517 \newcommand\BabelLowerMO[4]{% many-to-one
1518   \@tempcnta=#1\relax
1519   \def\bbl@tempa{%
1520     \ifnum\@tempcnta>#2\else
1521       \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
1522       \advance\@tempcnta#3
1523       \expandafter\bbl@tempa
1524     \fi}%
1525   \bbl@tempa}

```

The following package options control the behavior of hyphenation mapping.

```

1526 <<(*More package options)>> ≡

```

```

1527 \DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
1528 \DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\ne@}
1529 \DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
1530 \DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@@}
1531 \DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}
1532 <</More package options>>

```

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

```

1533 \AtEndOfPackage{%
1534   \ifx\bbl@opt@hyphenmap\undefined
1535     \bbl@xin{,}{\bbl@language@opts}%
1536     \chardef\bbl@opt@hyphenmap\ifin4\else\ne\fi
1537   \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.

```

1538 \bbl@trace{Macros related to glyphs}
1539 \def\set@low@box#1{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}%
1540   \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@%
1541   \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.

```

1542 \def\save@sf@q#1{\leavevmode
1543   \begingroup
1544   \edef\@SF{\spacefactor\the\spacefactor}#1\@SF
1545   \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.

```

1546 \ProvideTextCommand{\quotedblbase}{OT1}{%
1547   \save@sf@q{\set@low@box{\textquotedblright\}}%
1548   \box\z@\kern-.04em\bbl@allowhyphens}}

```

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

```

1549 \ProvideTextCommandDefault{\quotedblbase}{%
1550   \UseTextSymbol{OT1}{\quotedblbase}}

```

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

```

1551 \ProvideTextCommand{\quotesinglbase}{OT1}{%
1552   \save@sf@q{\set@low@box{\textquoteright\}}%
1553   \box\z@\kern-.04em\bbl@allowhyphens}}

```

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

```

1554 \ProvideTextCommandDefault{\quotesinglbase}{%
1555   \UseTextSymbol{OT1}{\quotesinglbase}}

```

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

```
\guillemotright 1556 \ProvideTextCommand{\guillemotleft}{OT1}{%
1557   \ifmmode
1558     \ll
1559   \else
1560     \save@sf@q{\nobreak
1561       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bbl@allowhyphens}%
1562   \fi}
1563 \ProvideTextCommand{\guillemotright}{OT1}{%
1564   \ifmmode
1565     \gg
1566   \else
1567     \save@sf@q{\nobreak
1568       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bbl@allowhyphens}%
1569   \fi}
```

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

```
1570 \ProvideTextCommandDefault{\guillemotleft}{%
1571   \UseTextSymbol{OT1}{\guillemotleft}}
1572 \ProvideTextCommandDefault{\guillemotright}{%
1573   \UseTextSymbol{OT1}{\guillemotright}}
```

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

```
\guilsinglright 1574 \ProvideTextCommand{\guilsinglleft}{OT1}{%
1575   \ifmmode
1576     <%
1577   \else
1578     \save@sf@q{\nobreak
1579       \raise.2ex\hbox{$\scriptscriptstyle<$}\bbl@allowhyphens}%
1580   \fi}
1581 \ProvideTextCommand{\guilsinglright}{OT1}{%
1582   \ifmmode
1583     >%
1584   \else
1585     \save@sf@q{\nobreak
1586       \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
1587   \fi}
```

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

```
1588 \ProvideTextCommandDefault{\guilsinglleft}{%
1589   \UseTextSymbol{OT1}{\guilsinglleft}}
1590 \ProvideTextCommandDefault{\guilsinglright}{%
1591   \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.

```
\IJ 1592 \DeclareTextCommand{\ij}{OT1}{%
1593   i\kern-0.02em\bbl@allowhyphens j}
1594 \DeclareTextCommand{\IJ}{OT1}{%
1595   I\kern-0.02em\bbl@allowhyphens J}
1596 \DeclareTextCommand{\ij}{T1}{\char188}
1597 \DeclareTextCommand{\IJ}{T1}{\char156}
```

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

```
1598 \ProvideTextCommandDefault{\ij}{%
1599   \UseTextSymbol{OT1}{\ij}}
1600 \ProvideTextCommandDefault{\IJ}{%
1601   \UseTextSymbol{OT1}{\IJ}}
```

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

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

```
1602 \def\crrtic@{\hrule height0.1ex width0.3em}
1603 \def\crttic@{\hrule height0.1ex width0.33em}
1604 \def\ddj@{%
1605   \setbox0\hbox{d}\dimen@=\ht0
1606   \advance\dimen@1ex
1607   \dimen@.45\dimen@
1608   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1609   \advance\dimen@ii.5ex
1610   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
1611 \def\DDJ@{%
1612   \setbox0\hbox{D}\dimen@=.55\ht0
1613   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1614   \advance\dimen@ii.15ex % correction for the dash position
1615   \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
1616   \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
1617   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
1618 %
1619 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
1620 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}
```

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

```
1621 \ProvideTextCommandDefault{\dj}{%
1622   \UseTextSymbol{OT1}{\dj}}
1623 \ProvideTextCommandDefault{\DJ}{%
1624   \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.

```
1625 \DeclareTextCommand{\SS}{OT1}{SS}
1626 \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 1627 \ProvideTextCommandDefault{\glq}{%
1628   \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}
```

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



```

1629 \ProvideTextCommand{\grq}{T1}{%
1630   \textormath{\kern\z@\textquoteleft}{\mbox{\textquoteleft}}}
1631 \ProvideTextCommand{\grq}{TU}{%
1632   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
1633 \ProvideTextCommand{\grq}{OT1}{%
1634   \save@sf@q{\kern-.0125em
1635     \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
1636     \kern.07em\relax}}
1637 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}

```

`\glqq` The ‘german’ double quotes.

```

\grqq 1638 \ProvideTextCommandDefault{\glqq}{%
1639   \textormath{\quotedblbase}{\mbox{\quotedblbase}}}

```

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

```

1640 \ProvideTextCommand{\grqq}{T1}{%
1641   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1642 \ProvideTextCommand{\grqq}{TU}{%
1643   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1644 \ProvideTextCommand{\grqq}{OT1}{%
1645   \save@sf@q{\kern-.07em
1646     \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
1647     \kern.07em\relax}}
1648 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}

```

`\flq` The ‘french’ single guillemets.

```

\frq 1649 \ProvideTextCommandDefault{\flq}{%
1650   \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
1651 \ProvideTextCommandDefault{\frq}{%
1652   \textormath{\guilsinglright}{\mbox{\guilsinglright}}}

```

`\flqq` The ‘french’ double guillemets.

```

\frqq 1653 \ProvideTextCommandDefault{\flqq}{%
1654   \textormath{\guillemotleft}{\mbox{\guillemotleft}}}
1655 \ProvideTextCommandDefault{\frqq}{%
1656   \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).

```

1657 \def\umlauthigh{%
1658   \def\bb1@umlauta##1{\leavevmode\bggroup%
1659     \expandafter\accent\csname\fontencoding dqpos\endcsname
1660     ##1\bb1@allowhyphens\egroup}%
1661   \let\bb1@umlaute\bb1@umlauta}
1662 \def\umlautlow{%
1663   \def\bb1@umlauta{\protect\lower@umlaut}}
1664 \def\umlautelow{%
1665   \def\bb1@umlaute{\protect\lower@umlaut}}
1666 \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.

```
1667 \expandafter\ifx\csname U@D\endcsname\relax
1668   \csname newdimen\endcsname\U@D
1669 \fi
```

The following code fools T<sub>E</sub>X's `make_accent` procedure about the current x-height of the font to force another placement of the umlaut character. First we have to save the current x-height of the font, because we'll change this font dimension and this is always done globally.

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

```
1670 \def\lower@umlaut#1{%
1671   \leavevmode\bgroup
1672     \U@D 1ex%
1673     {\setbox\z@\hbox{%
1674       \expandafter\char\csname\fontencoding dqpos\endcsname}%
1675       \dimen@ -.45ex\advance\dimen@\ht\z@
1676       \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
1677     \expandafter\accent\csname\fontencoding dqpos\endcsname
1678     \fontdimen5\font\U@D #1%
1679   \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).

```
1680 \AtBeginDocument{%
1681   \DeclareTextCompositeCommand{\"}{OT1}{a}{\bbl@umlauta{a}}%
1682   \DeclareTextCompositeCommand{\"}{OT1}{e}{\bbl@umlaute{e}}%
1683   \DeclareTextCompositeCommand{\"}{OT1}{i}{\bbl@umlaute{~i}}%
1684   \DeclareTextCompositeCommand{\"}{OT1}{~i}{\bbl@umlaute{~i}}%
1685   \DeclareTextCompositeCommand{\"}{OT1}{o}{\bbl@umlauta{o}}%
1686   \DeclareTextCompositeCommand{\"}{OT1}{u}{\bbl@umlauta{u}}%
1687   \DeclareTextCompositeCommand{\"}{OT1}{A}{\bbl@umlauta{A}}%
1688   \DeclareTextCompositeCommand{\"}{OT1}{E}{\bbl@umlaute{E}}%
1689   \DeclareTextCompositeCommand{\"}{OT1}{I}{\bbl@umlaute{I}}%
1690   \DeclareTextCompositeCommand{\"}{OT1}{O}{\bbl@umlauta{O}}%
1691   \DeclareTextCompositeCommand{\"}{OT1}{U}{\bbl@umlauta{U}}%
1692 }
```

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

```
1693 \ifx\l@english\undefined
1694   \chardef\l@english\z@
1695 \fi
1696 \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.

```

1697 \bbl@trace{Bidi layout}
1698 \providecommand\IfBabelLayout[3]{#3}%
1699 \newcommand\BabelPatchSection[1]{%
1700   \@ifundefined{#1}{}{%
1701     \bbl@exp{\let\<bbl@ss@#1>\<#1>}%
1702     \@namedef{#1}{%
1703       \ifstar{\bbl@presec@s{#1}}%
1704       {\@dblarg{\bbl@presec@x{#1}}}}}%
1705 \def\bbl@presec@x#1[#2]#3{%
1706   \bbl@exp{%
1707     \\\select@language@x{\bbl@main@language}%
1708     \\\@nameuse{bbl@sspre@#1}%
1709     \\\@nameuse{bbl@ss@#1}%
1710     [\\foreignlanguage{\languagename}{\unexpanded{#2}}}%
1711     {\\foreignlanguage{\languagename}{\unexpanded{#3}}}%
1712     \\\select@language@x{\languagename}}}%
1713 \def\bbl@presec@s#1#2{%
1714   \bbl@exp{%
1715     \\\select@language@x{\bbl@main@language}%
1716     \\\@nameuse{bbl@sspre@#1}%
1717     \\\@nameuse{bbl@ss@#1}*%
1718     {\\foreignlanguage{\languagename}{\unexpanded{#2}}}%
1719     \\\select@language@x{\languagename}}}%
1720 \IfBabelLayout{sectioning}%
1721   {\BabelPatchSection{part}%
1722     \BabelPatchSection{chapter}%
1723     \BabelPatchSection{section}%
1724     \BabelPatchSection{subsection}%
1725     \BabelPatchSection{subsubsection}%
1726     \BabelPatchSection{paragraph}%
1727     \BabelPatchSection{subparagraph}}%
1728   \def\babel@toc#1{%
1729     \select@language@x{\bbl@main@language}}}%
1730 \IfBabelLayout{captions}%
1731   {\BabelPatchSection{caption}}}%

```

### 9.13 Load engine specific macros

```

1732 \bbl@trace{Input engine specific macros}
1733 \ifcase\bbl@engine
1734   \input txtbabel.def
1735 \or
1736   \input luababel.def
1737 \or
1738   \input xebabel.def
1739 \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.

```

1740 \bbl@trace{Creating languages and reading ini files}
1741 \newcommand\babelprovide[2][{}]{%
1742   \let\bbl@savelangname\languagename
1743   \edef\bbl@savlocaleid{\the\localeid}%

```

```

1744 % Set name and locale id
1745 \def\languagename{#2}%
1746 \bbl@id@assign
1747 \chardef\localeid\@nameuse{\bbl@id@\languagename}%
1748 \let\bbl@KVP@captions\@nil
1749 \let\bbl@KVP@import\@nil
1750 \let\bbl@KVP@main\@nil
1751 \let\bbl@KVP@script\@nil
1752 \let\bbl@KVP@language\@nil
1753 \let\bbl@KVP@dir\@nil
1754 \let\bbl@KVP@hyphenrules\@nil
1755 \let\bbl@KVP@mapfont\@nil
1756 \let\bbl@KVP@maparabic\@nil
1757 \let\bbl@KVP@mapdigits\@nil
1758 \let\bbl@KVP@intraspace\@nil
1759 \let\bbl@KVP@intrapenalty\@nil
1760 \bbl@forkv{#1}{% TODO - error handling
1761   \in@{/}{##1}%
1762   \ifin@
1763     \bbl@renewinikey##1\@{##2}%
1764   \else
1765     \bbl@csarg\def{KVP@##1}{##2}%
1766   \fi}%
1767 \ifx\bbl@KVP@import\@nil\else
1768   \bbl@exp{\bbl@ifblank{\bbl@KVP@import}}%
1769   {\begingroup
1770     \def\BabelBeforeIni##1##2{\gdef\bbl@KVP@import{##1}\endinput}%
1771     \InputIfFileExists{babel-#2.tex}{}}%
1772   \endgroup}%
1773   {}%
1774 \fi
1775 \ifx\bbl@KVP@captions\@nil
1776   \let\bbl@KVP@captions\bbl@KVP@import
1777 \fi
1778 % Load ini
1779 \bbl@ifunset{date#2}%
1780   {\bbl@provide@new{#2}}%
1781   {\bbl@ifblank{#1}%
1782     {\bbl@error
1783       {If you want to modify `#2' you must tell how in\\
1784         the optional argument. See the manual for the\\
1785         available options.}%
1786       {Use this macro as documented}}%
1787     {\bbl@provide@renew{#2}}}%
1788 % Post tasks
1789 \bbl@exp{\bbl@babelensure[exclude=\\today]{#2}}%
1790 \bbl@ifunset{\bbl@ensure@\languagename}%
1791   {\bbl@exp{%
1792     \\DeclareRobustCommand\<\bbl@ensure@\languagename>[1]{%
1793       \\foreignlanguage{\languagename}%
1794       {###1}}}%
1795   }%
1796 % At this point all parameters are defined if 'import'. Now we
1797 % execute some code depending on them. But what about if nothing was
1798 % imported? We just load the very basic parameters: ids and a few
1799 % more.
1800 \bbl@ifunset{\bbl@lname@#2}%
1801   {\def\BabelBeforeIni##1##2{%
1802     \begingroup

```

```

1803 \catcode\ [=12 \catcode\]=12 \catcode\==12 %
1804 \let\bb1@ini@captions@aux\@gobbletwo
1805 \def\bb1@inidate ####1.####2.####3.####4\relax ####5####6{}%
1806 \bb1@read@ini{##1}%
1807 \bb1@exportkey{chrng}{characters.ranges}{}%
1808 \bb1@exportkey{dgnat}{numbers.digits.native}{}%
1809 \endgroup% boxed, to avoid extra spaces:
1810 {\setbox\z@\hbox{\InputIfFileExists{babel-#2.tex}{}}}%
1811 {}%
1812 % -
1813 % Override script and language names with script= and language=
1814 \ifx\bb1@KVP@script\@nil\else
1815 \bb1@csarg\edef{sname@#2}{\bb1@KVP@script}%
1816 \fi
1817 \ifx\bb1@KVP@language\@nil\else
1818 \bb1@csarg\edef{lname@#2}{\bb1@KVP@language}%
1819 \fi
1820 % For bidi texts, to switch the language based on direction
1821 \ifx\bb1@KVP@mapfont\@nil\else
1822 \bb1@ifsamestring{\bb1@KVP@mapfont}{direction}{}%
1823 {\bb1@error{Option '\bb1@KVP@mapfont' unknown for\\
1824 mapfont. Use 'direction'.%
1825 {See the manual for details.}}}%
1826 \bb1@ifunset{\bb1@lsys@\language}\bb1@provide@lsys{\language}{}%
1827 \bb1@ifunset{\bb1@wdir@\language}\bb1@provide@dirs{\language}{}%
1828 \ifx\bb1@mapselect\@undefined
1829 \AtBeginDocument{%
1830 \expandafter\bb1@add\csname selectfont \endcsname{\bb1@mapselect}%
1831 {\selectfont}}%
1832 \def\bb1@mapselect{%
1833 \let\bb1@mapselect\relax
1834 \edef\bb1@prefontid{\fontid\font}%
1835 \def\bb1@mapdir##1{%
1836 {\def\language{##1}%
1837 \let\bb1@ifrestoring\@firstoftwo % avoid font warning
1838 \bb1@switchfont
1839 \directlua{Babel.fontmap
1840 [\the\csname bbl@wdir@##1\endcsname]%
1841 [\bb1@prefontid]=\fontid\font}}%
1842 \fi
1843 \bb1@exp{\bb1@add\bb1@mapselect{\bb1@mapdir{\language}}}%
1844 \fi
1845 % For East Asian, Southeast Asian, if interspace in ini - TODO: as hook?
1846 \ifx\bb1@KVP@intraspace\@nil\else % We may override the ini
1847 \bb1@csarg\edef{intsp@#2}{\bb1@KVP@intraspace}%
1848 \fi
1849 \ifcase\bb1@engine\or
1850 \bb1@ifunset{\bb1@intsp@\language}{%
1851 {\expandafter\ifx\csname bbl@intsp@\language\endcsname\@empty\else
1852 \bb1@xin@{\bb1@cs{sbcpr@\language}}{Hant,Hans,Jpan,Kore,Kana}%
1853 \fin@
1854 \bb1@cjk@intraspace
1855 \directlua{
1856 Babel = Babel or {}
1857 Babel.locale_props = Babel.locale_props or {}
1858 Babel.locale_props[\the\localeid].linebreak = 'c'
1859 }%
1860 \bb1@exp{\bb1@intraspace\bb1@cs{intsp@\language}}\@{}%
1861 \ifx\bb1@KVP@intrapenalty\@nil

```

```

1862         \bbl@intrapenalty0\@@
1863     \fi
1864 \else
1865     \bbl@seaintraspace
1866     \bbl@exp{\bbl@intraspace\bbl@cs{intsp@\language}\@@}%
1867     \directlua{
1868         Babel = Babel or {}
1869         Babel.sea_ranges = Babel.sea_ranges or {}
1870         Babel.set_chranges('\bbl@cs{sbcpr@\language}',
1871             '\bbl@cs{chrng@\language}')
1872     }%
1873     \ifx\bbl@KVP@intrapenalty\@nil
1874         \bbl@intrapenalty0\@@
1875     \fi
1876 \fi
1877 \fi
1878 \ifx\bbl@KVP@intrapenalty\@nil\else
1879     \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@@
1880 \fi}%
1881 \or
1882     \bbl@xin@\bbl@cs{sbcpr@\language}}{Thai,Lao,Khmr}%
1883 \ifin@
1884     \bbl@ifunset{\bbl@intsp@\language}}{%
1885         {\expandafter\ifx\csname\bbl@intsp@\language\endcsname\@empty\else
1886             \ifx\bbl@KVP@intraspace\@nil
1887                 \bbl@exp{%
1888                     \bbl@intraspace\bbl@cs{intsp@\language}\@@}%
1889             \fi
1890             \ifx\bbl@KVP@intrapenalty\@nil
1891                 \bbl@intrapenalty0\@@
1892             \fi
1893             \fi
1894             \ifx\bbl@KVP@intraspace\@nil\else % We may override the ini
1895                 \expandafter\bbl@intraspace\bbl@KVP@intraspace\@@
1896             \fi
1897             \ifx\bbl@KVP@intrapenalty\@nil\else
1898                 \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@@
1899             \fi
1900             \ifx\bbl@ispace\@undefined
1901                 \AtBeginDocument{%
1902                     \expandafter\bbl@add
1903                     \csname selectfont\endcsname{\bbl@ispace}%
1904                     \def\bbl@ispace{\bbl@cs{xeisp@\bbl@cs{sbcpr@\language}}}%
1905                 \fi}%
1906         \fi
1907 \fi
1908 % Native digits, if provided in ini (TeX level, xe and lua)
1909 \ifcase\bbl@engine\else
1910     \bbl@ifunset{\bbl@dgnat@\language}}{%
1911         {\expandafter\ifx\csname\bbl@dgnat@\language\endcsname\@empty\else
1912             \expandafter\expandafter\expandafter
1913             \bbl@setdigits\csname\bbl@dgnat@\language\endcsname
1914             \ifx\bbl@KVP@maparabic\@nil\else
1915                 \ifx\bbl@latinarabic\@undefined
1916                     \expandafter\let\expandafter\@arabic
1917                     \csname\bbl@counter@\language\endcsname
1918                 \else % ie, if layout=counters, which redefines \@arabic
1919                     \expandafter\let\expandafter\bbl@latinarabic
1920                     \csname\bbl@counter@\language\endcsname

```

```

1921     \fi
1922     \fi
1923     \fi}%
1924 \fi
1925 % Native digits (lua level).
1926 \ifodd\bbl@engine
1927     \ifx\bbl@KVP@mapdigits\@nil\else
1928         \bbl@ifunset{bbl@dgnat@language}{}%
1929         {\RequirePackage{luatexbase}%
1930         \bbl@activate@preotf
1931         \directlua{
1932             Babel = Babel or {} %% -> presets in luababel
1933             Babel.digits_mapped = true
1934             Babel.digits = Babel.digits or {}
1935             Babel.digits[\the\localeid] =
1936                 table.pack(string.utfvalue('\bbl@cs{dgnat@language}'))
1937             if not Babel.numbers then
1938                 function Babel.numbers(head)
1939                     local LOCALE = luatexbase.registernumber'bbl@attr@locale'
1940                     local GLYPH = node.id'glyph'
1941                     local inmath = false
1942                     for item in node.traverse(head) do
1943                         if not inmath and item.id == GLYPH then
1944                             local temp = node.get_attribute(item, LOCALE)
1945                             if Babel.digits[temp] then
1946                                 local chr = item.char
1947                                 if chr > 47 and chr < 58 then
1948                                     item.char = Babel.digits[temp][chr-47]
1949                                 end
1950                             end
1951                             elseif item.id == node.id'math' then
1952                                 inmath = (item.subtype == 0)
1953                             end
1954                         end
1955                     return head
1956                 end
1957             end
1958         }}
1959     \fi
1960 \fi
1961 % To load or reload the babel-*.tex, if require.babel in ini
1962 \bbl@ifunset{bbl@rqtex@language}{}%
1963     {\expandafter\ifx\csname bbl@rqtex@language\endcsname\@empty\else
1964         \let\BabelBeforeIni\@gobbletwo
1965         \chardef\atcatcode=\catcode`\@
1966         \catcode`\@=11\relax
1967         \InputIfFileExists{babel-\bbl@cs{rqtex@language}.tex}{\}%
1968         \catcode`\@=\atcatcode
1969         \let\atcatcode\relax
1970     \fi}%
1971 \ifx\bbl@KVP@main\@nil % Restore only if not 'main'
1972     \let\language\bbl@savelangname
1973     \chardef\localeid\bbl@savelocaleid\relax
1974 \fi}

```

A tool to define the macros for native digits from the list provided in the ini file.  
Somewhat convoluted because there are 10 digits, but only 9 arguments in T<sub>E</sub>X.

```

1975 \def\bbl@setdigits#1#2#3#4#5{%
1976     \bbl@exp{%

```

[illegible]

—



```

2033 \bbl@exp{%
2034   \def\<#1hyphenmins>{%
2035     {\bbl@ifunset{\bbl@lfthm@#1}{2}{\@nameuse{\bbl@lfthm@#1}}}%
2036     {\bbl@ifunset{\bbl@rgthm@#1}{3}{\@nameuse{\bbl@rgthm@#1}}}}}%
2037 \bbl@provide@hyphens{#1}%
2038 \ifx\bbl@KVP@main\@nil\else
2039   \expandafter\main@language\expandafter{#1}%
2040 \fi}
2041 \def\bbl@provide@renew#1{%
2042   \ifx\bbl@KVP@captions\@nil\else
2043     \StartBabelCommands*{#1}{captions}%
2044     \bbl@read@ini{\bbl@KVP@captions}%   Here all letters cat = 11
2045     \bbl@after@ini
2046     \bbl@savestrings
2047     \EndBabelCommands
2048   \fi
2049   \ifx\bbl@KVP@import\@nil\else
2050     \StartBabelCommands*{#1}{date}%
2051     \bbl@savetoday
2052     \bbl@savestate
2053     \EndBabelCommands
2054   \fi
2055   \bbl@provide@hyphens{#1}}

```

The hyphenrules option is handled with an auxiliary macro.

```

2056 \def\bbl@provide@hyphens#1{%
2057   \let\bbl@tempa\relax
2058   \ifx\bbl@KVP@hyphenrules\@nil\else
2059     \bbl@replace\bbl@KVP@hyphenrules{ }{,}%
2060     \bbl@foreach\bbl@KVP@hyphenrules{%
2061       \ifx\bbl@tempa\relax    % if not yet found
2062         \bbl@ifsamestring{##1}{+}%
2063         {\bbl@exp{\addlanguage\<l@##1>}}}%
2064       {}%
2065       \bbl@ifunset{l@##1}%
2066       {}%
2067       {\bbl@exp{\let\bbl@tempa\<l@##1>}}}%
2068     \fi}%
2069   \fi
2070   \ifx\bbl@tempa\relax %      if no opt or no language in opt found
2071     \ifx\bbl@KVP@import\@nil\else % if importing
2072       \bbl@exp{%              and hyphenrules is not empty
2073         \bbl@ifblank{\@nameuse{\bbl@hyphr@#1}}}%
2074         {}%
2075         {\let\bbl@tempa\<l@\@nameuse{\bbl@hyphr@\language}\>}}}%
2076     \fi
2077   \fi
2078   \bbl@ifunset{\bbl@tempa}%    ie, relax or undefined
2079   {\bbl@ifunset{l@#1}%         no hyphenrules found - fallback
2080     {\bbl@exp{\adddialect\<l@#1>\language}}}%
2081     {}%                         so, l@<lang> is ok - nothing to do
2082     {\bbl@exp{\adddialect\<l@#1>\bbl@tempa}}}% found in opt list or ini
2083   \bbl@ifunset{\bbl@prehc@\language}%
2084   {}%   TODO - XeTeX, based on \babelfont and HyphenChar?
2085   {\ifodd\bbl@engine\bbl@exp{%
2086     \bbl@ifblank{\@nameuse{\bbl@prehc@#1}}}%
2087     {}%
2088     {\AddBabelHook[\language]{babel-prehc-\language}{patterns}%
2089     {\prehyphenchar=\@nameuse{\bbl@prehc@\language}\relax}}}%

```

```
2090     \fi}}
```

The reader of ini files. There are 3 possible cases: a section name (in the form [ . . . ]), a comment (starting with ;) and a key/value pair. *TODO - Work in progress.*

```
2091 \def\bbl@read@ini#1{%
2092   \openin1=babel-#1.ini           % FIXME - number must not be hardcoded
2093   \ifeof1
2094     \bbl@error
2095     {There is no ini file for the requested language\\%
2096      (#1). Perhaps you misspelled it or your installation\\%
2097      is not complete.}%
2098     {Fix the name or reinstall babel.}%
2099   \else
2100     \let\bbl@section\@empty
2101     \let\bbl@savestrings\@empty
2102     \let\bbl@savetoday\@empty
2103     \let\bbl@savestate\@empty
2104     \def\bbl@inipreread##1=##2\@{%
2105       \bbl@trim@def\bbl@tempa{##1}% Redundant below !!
2106       % Move trims here ??
2107       \bbl@ifunset{bbl@KVP@\bbl@section/\bbl@tempa}%
2108       {\expandafter\bbl@inireader\bbl@tempa=##2\@}%
2109       {}}%
2110     \let\bbl@inireader\bbl@iniskip
2111     \bbl@info{Importing data from babel-#1.ini for \language}%
2112     \loop
2113     \if T\ifeof1F\fi T\relax % Trick, because inside \loop
2114       \endlinechar\m@ne
2115       \read1 to \bbl@line
2116       \endlinechar\^^M
2117       \ifx\bbl@line\@empty\else
2118         \expandafter\bbl@iniline\bbl@line\bbl@iniline
2119       \fi
2120     \repeat
2121   \fi}
2122 \def\bbl@iniline#1\bbl@iniline{%
2123   \@ifnextchar[\bbl@inisec{\@ifnextchar;\bbl@iniskip\bbl@inipreread}#1\@}% ]
```

The special cases for comment lines and sections are handled by the two following commands. In sections, we provide the possibility to take extra actions at the end or at the start (TODO - but note the last section is not ended). By default, key=val pairs are ignored.

```
2124 \def\bbl@iniskip#1\@{%           if starts with ;
2125 \def\bbl@inisec[#1]#2\@{%       if starts with opening bracket
2126   \def\bbl@elt##1##2{\bbl@inireader##1=##2\@}%
2127   \@nameuse{bbl@renew@\bbl@section}%
2128   \global\bbl@csarg\let{renew@\bbl@section}\relax
2129   \@nameuse{bbl@secpost@\bbl@section}% ends previous section
2130   \def\bbl@section{#1}%
2131   \def\bbl@elt##1##2{%
2132     \@namedef{bbl@KVP@#1/##1}{}}%
2133   \@nameuse{bbl@renew@#1}%
2134   \@nameuse{bbl@secpre@#1}% starts current section
2135   \bbl@ifunset{bbl@inikv@#1}%
2136   {\let\bbl@inireader\bbl@iniskip}%
2137   {\bbl@exp{\let\\bbl@inireader<bbl@inikv@#1>}}}
```

Reads a key=val line and stores the trimmed val in \bbl@kv@<section>.<key>.

```
2138 \def\bbl@inikv#1=##2\@{%       key=value
2139   \bbl@trim@def\bbl@tempa{#1}%
```

```

2140 \bbl@trim\toks@{#2}%
2141 \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.

```

2142 \def\bbl@exportkey#1#2#3{%
2143 \bbl@ifunset{bbl@kv@#2}%
2144   {\bbl@csarg\gdef{#1@\language}\{#3}}%
2145   {\expandafter\ifx\csname bbl@kv@#2\endcsname\empty
2146     \bbl@csarg\gdef{#1@\language}\{#3}}%
2147   \else
2148     \bbl@exp{\global\let\<bbl@#1@\language>\<bbl@kv@#2>}}%
2149   \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.

```

2150 \let\bbl@inikv@identification\bbl@inikv
2151 \def\bbl@secpost@identification{%
2152 \bbl@exportkey{lname}{identification.name.english}{}}%
2153 \bbl@exportkey{lbcpr}{identification.tag.bcp47}{}}%
2154 \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
2155 \bbl@exportkey{sname}{identification.script.name}{}}%
2156 \bbl@exportkey{sbcpr}{identification.script.tag.bcp47}{}}%
2157 \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}}
2158 \let\bbl@inikv@typography\bbl@inikv
2159 \let\bbl@inikv@characters\bbl@inikv
2160 \let\bbl@inikv@numbers\bbl@inikv
2161 \def\bbl@after@ini{%
2162 \bbl@exportkey{lftm}{typography.lefthyphenmin}{2}}%
2163 \bbl@exportkey{rgtm}{typography.righthyphenmin}{3}}%
2164 \bbl@exportkey{prehc}{typography.prehyphenchar}{}}%
2165 \bbl@exportkey{hyphr}{typography.hyphenrules}{}}%
2166 \bbl@exportkey{intsp}{typography.intraspace}{}}%
2167 \bbl@exportkey{jstfy}{typography.justify}{w}}%
2168 \bbl@exportkey{chrng}{characters.ranges}{}}%
2169 \bbl@exportkey{dgnat}{numbers.digits.native}{}}%
2170 \bbl@exportkey{rqtex}{identification.require.babel}{}}%
2171 \bbl@xin@{0.5}{\@nameuse{bbl@kv@identification.version}}%
2172 \ifin@
2173 \bbl@warning{%
2174   There are neither captions nor date in '\language'.\%
2175   It may not be suitable for proper typesetting, and it\%
2176   could change. Reported}%
2177 \fi
2178 \bbl@xin@{0.9}{\@nameuse{bbl@kv@identification.version}}%
2179 \ifin@
2180 \bbl@warning{%
2181   The '\language' date format may not be suitable\%
2182   for proper typesetting, and therefore it very likely will\%
2183   change in a future release. Reported}%
2184 \fi
2185 \bbl@toglobal\bbl@savetoday
2186 \bbl@toglobal\bbl@savetoday

```

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.

```

2187 \ifcase\bbl@engine
2188 \bbl@csarg\def{inikv@captions.licr}#1=#2\@@{%

```

```

2189 \bbl@ini@captions@aux{#1}{#2}}
2190 \else
2191 \def\bbl@inikv@captions#1=#2\@@{%
2192 \bbl@ini@captions@aux{#1}{#2}}
2193 \fi

```

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

```

2194 \def\bbl@ini@captions@aux#1#2{%
2195 \bbl@trim@def\bbl@tempa{#1}%
2196 \bbl@ifblank{#2}%
2197 {\bbl@exp{%
2198 \toks@{\bbl@nocaption{\bbl@tempa}{\language\language\bbl@tempa name}}}%
2199 {\bbl@trim\toks@{#2}}}%
2200 \bbl@exp{%
2201 \bbl@add\bbl@savestrings{%
2202 \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.

```

2203 \bbl@csarg\def{inikv@date.gregorian}#1=#2\@@{% for defaults
2204 \bbl@inidate#1...\relax{#2}{}}
2205 \bbl@csarg\def{inikv@date.islamic}#1=#2\@@{%
2206 \bbl@inidate#1...\relax{#2}{islamic}}
2207 \bbl@csarg\def{inikv@date.hebrew}#1=#2\@@{%
2208 \bbl@inidate#1...\relax{#2}{hebrew}}
2209 \bbl@csarg\def{inikv@date.persian}#1=#2\@@{%
2210 \bbl@inidate#1...\relax{#2}{persian}}
2211 \bbl@csarg\def{inikv@date.indian}#1=#2\@@{%
2212 \bbl@inidate#1...\relax{#2}{indian}}
2213 \ifcase\bbl@engine
2214 \bbl@csarg\def{inikv@date.gregorian.licr}#1=#2\@@{% override
2215 \bbl@inidate#1...\relax{#2}{}}
2216 \bbl@csarg\def{secpre@date.gregorian.licr}{% discard uni
2217 \ifcase\bbl@engine\let\bbl@savestate\empty\fi}
2218 \fi
2219 % eg: 1=months, 2=wide, 3=1, 4=dummy
2220 \def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
2221 \bbl@trim@def\bbl@tempa{#1.#2}%
2222 \bbl@ifsamestring{\bbl@tempa}{months.wide}% to savedate
2223 {\bbl@trim@def\bbl@tempa{#3}%
2224 \bbl@trim\toks@{#5}%
2225 \bbl@exp{%
2226 \bbl@add\bbl@savestate{%
2227 \SetString\<month\romannumeral\bbl@tempa#6name>{\the\toks@}}}%
2228 {\bbl@ifsamestring{\bbl@tempa}{date.long}% defined now
2229 {\bbl@trim@def\bbl@toreplace{#5}%
2230 \bbl@TG@date
2231 \global\bbl@csarg\let{date@\language}\bbl@toreplace
2232 \bbl@exp{%
2233 \gdef\<\language date>{\protect\<\language date >}%
2234 \gdef\<\language date >####1####2####3{%
2235 \bbl@usedategroupttrue
2236 \<\bbl@ensure@\language>{%
2237 \<\bbl@date@\language>{####1}{####2}{####3}}}%
2238 \bbl@add\bbl@savetoday{%
2239 \SetString\<\today{%
2240 \<\language date>{\the\year}{\the\month}{\the\day}}}%

```

2241       {}}

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.

```

2242 \let\bbl@calendar\@empty
2243 \newcommand\BabelDateSpace{\nobreakspace}
2244 \newcommand\BabelDateDot{.\@}
2245 \newcommand\BabelDated[1]{\number#1}
2246 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
2247 \newcommand\BabelDateM[1]{\number#1}
2248 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}
2249 \newcommand\BabelDateMMMM[1]{%
2250   \csname month\romannumeral#1\bbl@calendar name\endcsname}%
2251 \newcommand\BabelDatey[1]{\number#1}%
2252 \newcommand\BabelDateyy[1]{%
2253   \ifnum#1<10 0\number#1 %
2254   \else\ifnum#1<100 \number#1 %
2255   \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
2256   \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
2257   \else
2258     \bbl@error
2259     {Currently two-digit years are restricted to the\
2260      range 0-9999.}%
2261     {There is little you can do. Sorry.}%
2262     \fi\fi\fi\fi}
2263 \newcommand\BabelDateyyyy[1]{\number#1} % FIXME - add leading 0
2264 \def\bbl@replace@finish@iii#1{%
2265   \bbl@exp{\def\#1####1####2####3{\the\toks@}}
2266 \def\bbl@TG@date{%
2267   \bbl@replace\bbl@toreplace{[ ]}{\BabelDateSpace{}}%
2268   \bbl@replace\bbl@toreplace{[. ]}{\BabelDateDot{}}%
2269   \bbl@replace\bbl@toreplace{[d]}{\BabelDated{####3}}%
2270   \bbl@replace\bbl@toreplace{[dd]}{\BabelDatedd{####3}}%
2271   \bbl@replace\bbl@toreplace{[M]}{\BabelDateM{####2}}%
2272   \bbl@replace\bbl@toreplace{[MM]}{\BabelDateMM{####2}}%
2273   \bbl@replace\bbl@toreplace{[MMMM]}{\BabelDateMMMM{####2}}%
2274   \bbl@replace\bbl@toreplace{[y]}{\BabelDatey{####1}}%
2275   \bbl@replace\bbl@toreplace{[yy]}{\BabelDateyy{####1}}%
2276   \bbl@replace\bbl@toreplace{[yyyy]}{\BabelDateyyyy{####1}}%
2277 % Note after \bbl@replace \toks@ contains the resulting string.
2278 % TODO - Using this implicit behavior doesn't seem a good idea.
2279   \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.

```

2280 \def\bbl@provide@lsys#1{%
2281   \bbl@ifunset\bbl@lname@#1{%
2282     {\bbl@ini@ids{#1}}%
2283     {}%
2284   \bbl@csarg\let{lsys@#1}\@empty
2285   \bbl@ifunset\bbl@sname@#1{\bbl@csarg\gdef{sname@#1}{Default}}}%
2286   \bbl@ifunset\bbl@sotf@#1{\bbl@csarg\gdef{sotf@#1}{DFLT}}}%
2287   \bbl@csarg\bbl@add@list{lsys@#1}{Script=\bbl@cs{sname@#1}}%
2288   \bbl@ifunset\bbl@lname@#1{%
2289     {\bbl@csarg\bbl@add@list{lsys@#1}{Language=\bbl@cs{lname@#1}}}%
2290   \bbl@csarg\bbl@toGLOBAL{lsys@#1}}

```

The following ini reader ignores everything but the identification section. It is called when a font is defined (ie, when the language is first selected) to know which

script/language must be enabled. This means we must make sure a few characters are not active. The ini is not read directly, but with a proxy tex file named as the language (which means any code in it must be skipped, too).

```

2291 \def\bbl@ini@ids#1{%
2292   \def\BabelBeforeIni##1##2{%
2293     \begingroup
2294       \bbl@add\bbl@secpost@identification{\closein1 }%
2295       \catcode`\[=12 \catcode`\]=12 \catcode`\==12 %
2296       \bbl@read@ini{##1}%
2297       \endinput
2298     \endgroup}%
2299   {\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`.

```

2300 {\def\format{plain}
2301 \ifx\fmtname\format
2302 \else
2303   \def\format{LaTeX2e}
2304   \ifx\fmtname\format
2305   \else
2306     \aftergroup\endinput
2307   \fi
2308 \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.

```

2309 %\bbl@redefine\newlabel#1#2{%
2310 %  \@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.

```
2311 <<*More package options>> ≡
2312 \DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
2313 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
2314 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
2315 <</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.

```
2316 \bbl@trace{Cross referencing macros}
2317 \ifx\bbl@opt@safe\@empty\else
2318   \def\@newl@bel#1#2#3{%
2319     {\@safe@activestrue
2320       \bbl@ifunset{#1@#2}%
2321         \relax
2322         {\gdef\@multiplelabels{%
2323           \@latex@warning@no@line{There were multiply-defined labels}}%
2324           \@latex@warning@no@line{Label `#2' multiply defined}}%
2325       \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.

```
2326 \CheckCommand*\@testdef[3]{%
2327   \def\reserved@a{#3}%
2328   \expandafter\ifx\cname#1@#2\endcsname\reserved@a
2329   \else
2330     \@tempswatrue
2331   \fi}
```

Now that we made sure that `\@testdef` still has the same definition we can rewrite it. First we make the shorthands ‘safe’.

```
2332 \def\@testdef#1#2#3{%
2333   \@safe@activestrue
```

Then we use `\bbl@tempa` as an ‘alias’ for the macro that contains the label which is being checked.

```
2334   \expandafter\let\expandafter\bbl@tempa\cname #1@#2\endcsname
```

Then we define `\bbl@tempb` just as `\@newl@bel` does it.

```
2335   \def\bbl@tempb{#3}%
2336   \@safe@activesfalse
```

When the label is defined we replace the definition of `\bbl@tempa` by its meaning.

```
2337   \ifx\bbl@tempa\relax
2338   \else
2339     \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
2340   \fi
```

We do the same for `\bbl@tempb`.

```
2341   \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
```

If the label didn't change, `\bbl@tempa` and `\bbl@tempb` should be identical macros.

```
2342 \ifx\bbl@tempa\bbl@tempb
2343 \else
2344 \@tempswattrue
2345 \fi}
2346 \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.

```
2347 \bbl@xin@{R}\bbl@opt@safe
2348 \ifin@
2349 \bbl@redefineroobust\ref#1{%
2350 \@safe@activestruer\org@ref{#1}\@safe@activestruerfalse}
2351 \bbl@redefineroobust\pageref#1{%
2352 \@safe@activestruer\org@pageref{#1}\@safe@activestruerfalse}
2353 \else
2354 \let\org@ref\ref
2355 \let\org@pageref\pageref
2356 \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.

```
2357 \bbl@xin@{B}\bbl@opt@safe
2358 \ifin@
2359 \bbl@redefine\@citex[#1]#2{%
2360 \@safe@activestruer\edef\@tempa{#2}\@safe@activestruerfalse
2361 \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.

```
2362 \AtBeginDocument{%
2363 \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.)

```
2364 \def\@citex[#1][#2]#3{%
2365 \@safe@activestruer\edef\@tempa{#3}\@safe@activestruerfalse
2366 \org@@citex[#1][#2]{\@tempa}}%
2367 }{}}
```

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

```
2368 \AtBeginDocument{%
2369 \ifpackageloaded{cite}{%
2370 \def\@citex[#1]#2{%
2371 \@safe@activestruer\org@@citex[#1][#2]\@safe@activestruerfalse}%
2372 }{}}
```



`\nocite` The macro `\nocite` which is used to instruct BiB<sub>T</sub><sub>E</sub>X to extract uncited references from the database.

```

2373 \bbl@redefine\nocite#1{%
2374   \@safe@activetrue\org@nocite{#1}\@safe@activesfalse}

```

`\bibcite` The macro that is used in the .aux file to define citation labels. When packages such as natbib or cite are not loaded its second argument is used to typeset the citation label. In that case, this second argument can contain active characters but is used in an environment where `\@safe@activetrue` is in effect. This switch needs to be reset inside the `\hbox` which contains the citation label. In order to determine during .aux file processing which definition of `\bibcite` is needed we define `\bibcite` in such a way that it redefines itself with the proper definition. We call `\bbl@cite@choice` to select the proper definition for `\bibcite`. This new definition is then activated.

```

2375 \bbl@redefine\bibcite{%
2376   \bbl@cite@choice
2377   \bibcite}

```

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

```

2378 \def\bbl@bibcite#1#2{%
2379   \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.

```

2380 \def\bbl@cite@choice{%
2381   \global\let\bibcite\bbl@bibcite

```

Then, when natbib is loaded we restore the original definition of `\bibcite`. For cite we do the same.

```

2382   \@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{}%
2383   \@ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{}%

```

Make sure this only happens once.

```

2384   \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.

```

2385   \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.

```

2386 \bbl@redefine\@bibitem#1{%
2387   \@safe@activetrue\org@bibitem{#1}\@safe@activesfalse}
2388 \else
2389   \let\org@nocite\nocite
2390   \let\org@@citex\citex
2391   \let\org@bibcite\bibcite
2392   \let\org@@bibitem\@bibitem
2393 \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.

```

2394 \bbl@trace{Marks}
2395 \IfBabelLayout{sectioning}
2396   {\ifx\bbl@opt@headfoot\@nnil
2397     \g@addto@macro\@resetactivechars{%
2398       \set@typeset@protect
2399       \expandafter\select@language@\x\expandafter{\bbl@main@language}%
2400       \let\protect\noexpand
2401       \edef\thepage{%
2402         \noexpand\babelsublr{\unexpanded\expandafter{\thepage}}}%
2403     \fi}
2404   {\ifbbl@single\else
2405     \bbl@ifunset{markright } \bbl@redefine\bbl@redefineroobust
2406     \markright#1{%
2407       \bbl@ifblank{#1}%
2408       {\org@markright{}}}%
2409     {\toks@{#1}%
2410      \bbl@exp{%
2411        \\org@markright{\\protect\\foreignlanguage{\language}%
2412          {\protect\\bbl@restore@actives\the\toks@}}}%

```

`\markboth`    The definition of `\markboth` is equivalent to that of `\markright`, except that we need two token registers. The documentclasses report and book define and set the headings for the page. While doing so they also store a copy of `\markboth` in `\@mkboth`. Therefore we need to check whether `\@mkboth` has already been set. If so we need to do that again with the new definition of `\markboth`. (As of Oct 2019,  $\text{\LaTeX}$  stores the definition in an intermediate macros, so it's not necessary anymore, but it's preserved for older versions.)

```

2413   \ifx\@mkboth\markboth
2414     \def\bbl@tempc{\let\@mkboth\markboth}
2415   \else
2416     \def\bbl@tempc{}
2417   \fi
2418   \bbl@ifunset{markboth } \bbl@redefine\bbl@redefineroobust
2419   \markboth#1#2{%
2420     \protected@edef\bbl@tempb##1{%
2421       \protect\foreignlanguage
2422       {\language}{\protect\bbl@restore@actives##1}}%
2423     \bbl@ifblank{#1}%
2424     {\toks@{}}%
2425     {\toks@\expandafter{\bbl@tempb{#1}}}%
2426     \bbl@ifblank{#2}%
2427     {\@temptokena{}}%
2428     {\@temptokena\expandafter{\bbl@tempb{#2}}}%
2429     \bbl@exp{\\org@markboth{\the\toks@}{\the\@temptokena}}
2430     \bbl@tempc
2431   \fi} % end ifbbl@single, end \IfBabelLayout

```

## 10.4 Preventing clashes with other packages

### 10.4.1 ifthen

`\ifthenelse` Sometimes a document writer wants to create a special effect depending on the page a certain fragment of text appears on. This can be achieved by the following piece of code:

```
\ifthenelse{\isodd{\pageref{some:label}}}{%
  {code for odd pages}%
  {code for even pages}%}
```

In order for this to work the argument of `\isodd` needs to be fully expandable. With the above redefinition of `\pageref` it is not in the case of this example. To overcome that, we add some code to the definition of `\ifthenelse` to make things work.

The first thing we need to do is check if the package `ifthen` is loaded. This should be done at `\begin{document}` time.

```
2432 \bbl@trace{Preventing clashes with other packages}%
2433 \bbl@xin@{R}\bbl@opt@safe
2434 \ifin@
2435   \AtBeginDocument{%
2436     \@ifpackageloaded{ifthen}{%
```

Then we can redefine `\ifthenelse`:

```
2437       \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.

```
2438       \let\bbl@temp@pref\pageref
2439       \let\pageref\org@pageref
2440       \let\bbl@temp@ref\ref
2441       \let\ref\org@ref
```

Then we can set the `\@safe@actives` switch and call the original `\ifthenelse`. In order to be able to use shorthands in the second and third arguments of `\ifthenelse` the resetting of the switch *and* the definition of `\pageref` happens inside those arguments. When the package wasn't loaded we do nothing.

```
2442       \@safe@activestrue
2443       \org@ifthenelse{#1}%
2444       {\let\pageref\bbl@temp@pref
2445        \let\ref\bbl@temp@ref
2446        \@safe@activesfalse
2447        #2}%
2448       {\let\pageref\bbl@temp@pref
2449        \let\ref\bbl@temp@ref
2450        \@safe@activesfalse
2451        #3}%
2452     }%
2453   }{}%
2454 }
```

### 10.4.2 varioref

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

```
2455   \AtBeginDocument{%
2456     \@ifpackageloaded{varioref}{%
```

```

2457 \bbl@redefine\@@vpageref#1[#2]#3{%
2458 \@safe@activestru
2459 \org@@@vpageref{#1}[#2]{#3}%
2460 \@safe@activesfalse}%
2461 \bbl@redefine\hrefpagenum#1#2{%
2462 \@safe@activestru
2463 \org@hrefpagenum{#1}{#2}%
2464 \@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.

```

2465 \expandafter\def\csname Ref \endcsname#1{%
2466 \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
2467 }{}%
2468 }
2469 \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.

```

2470 \AtEndOfPackage{%
2471 \AtBeginDocument{%
2472 \ifpackageloaded{hhline}%

```

Then we check whether the expansion of `\normal@char:` is not equal to `\relax`.

```

2473 {\expandafter\ifx\csname normal@char\string\endcsname\relax
2474 \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.

```

2475 \makeatletter
2476 \def\@currname{hhline}\input{hhline.sty}\makeatother
2477 \fi}%
2478 {}}}

```

### 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.

```

2479 \AtBeginDocument{%
2480 \ifx\pdfstringdefDisableCommands\undefined\else
2481 \pdfstringdefDisableCommands{\languageshortands{system}}%
2482 \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.

```
2483 \DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
2484   \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.

```
2485 \def\substitutefontfamily#1#2#3{%
2486   \lowercase{\immediate\openout15=#1#2.fd\relax}%
2487   \immediate\write15{%
2488     \string\ProvidesFile{#1#2.fd}%
2489     [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
2490     \space generated font description file]^{}
2491     \string\DeclareFontFamily{#1}{#2}{}^{}
2492     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{}^{}
2493     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{}^{}
2494     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{}^{}
2495     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{}^{}
2496     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{}^{}
2497     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{}^{}
2498     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{}^{}
2499     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{}^{}
2500   }%
2501   \closeout15
2502 }
```

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

```
2503 \@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>enc.def`. If a non-ASCII has been loaded, we define versions of `\TeX` and `\LaTeX` for them using `\ensureascii`. The default ASCII encoding is `set`, too (in reverse order): the “main” encoding (when the document begins), the last loaded, or `OT1`.

`\ensureascii`

```
2504 \bbl@trace{Encoding and fonts}
2505 \newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU,PU,PD1}
2506 \newcommand\BabelNonText{TS1,T3,TS3}
2507 \let\org@TeX\TeX
2508 \let\org@LaTeX\LaTeX
2509 \let\ensureascii\@firstofone
2510 \AtBeginDocument{%
2511   \in@false
2512   \bbl@foreach\BabelNonASCII{% is there a text non-ascii enc?
2513     \ifin@false
2514       \lowercase{\bbl@xin@{, #1enc.def,}{, \@filelist,}}%
2515     \fi}%
2516   \ifin@ % if a text non-ascii has been loaded
```

```

2517 \def\ensureascii#1{\fontencoding{OT1}\selectfont#1}}%
2518 \DeclareTextCommandDefault{\TeX}{\org@TeX}%
2519 \DeclareTextCommandDefault{\LaTeX}{\org@LaTeX}%
2520 \def\bbl@tempb#1\@{\uppercase{\bbl@tempc#1}ENC.DEF\@empty\@}%
2521 \def\bbl@tempc#1ENC.DEF#2\@{\%
2522   \ifx\@empty#2\else
2523     \bbl@ifunset{T#1}%
2524     {}%
2525     {\bbl@xin@{, #1, }{\, \BabelNonASCII, \BabelNonText, }%
2526     \ifin@
2527       \DeclareTextCommand{\TeX}{#1}{\ensureascii{\org@TeX}}%
2528       \DeclareTextCommand{\LaTeX}{#1}{\ensureascii{\org@LaTeX}}%
2529     \else
2530       \def\ensureascii##1{\fontencoding{#1}\selectfont##1}}%
2531     \fi}%
2532   \fi}%
2533 \bbl@foreach\@filelist{\bbl@tempb#1\@}% TODO - \@ de mas??
2534 \bbl@xin@{, \cf@encoding, }{\, \BabelNonASCII, \BabelNonText, }%
2535 \ifin@\else
2536   \edef\ensureascii#1{\%
2537     \noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}}%
2538   \fi
2539 \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.

```

2540 \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.

```

2541 \AtBeginDocument{%
2542   \@ifpackageloaded{fontspec}%
2543   {\xdef\latinencoding{%
2544     \ifx\UTFencname\@undefined
2545       EU\ifcase\bbl@engine\or2\or1\fi
2546     \else
2547       \UTFencname
2548     \fi}}%
2549   {\gdef\latinencoding{OT1}%
2550     \ifx\cf@encoding\bbl@t@one
2551       \xdef\latinencoding{\bbl@t@one}%
2552     \else
2553       \@ifl@aded{def}{t1enc}{\xdef\latinencoding{\bbl@t@one}}}%
2554     \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.

```

2555 \DeclareRobustCommand{\latintext}{%
2556   \fontencoding{\latinencoding}\selectfont
2557   \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.

```
2558 \ifx\@undefined\DeclareTextFontCommand
2559 \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
2560 \else
2561 \DeclareTextFontCommand{\textlatin}{\latintext}
2562 \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>>).

```
2563 \bbl@trace{Basic (internal) bidi support}
2564 \def\bbl@alscripts{,Arabic,Syriac,Thaana,}
2565 \def\bbl@rscripts{%
2566   ,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
2567   Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaeen,%
2568   Manichaeen,Meroitic Cursive,Meroitic,Old North Arabian,%
2569   Nabataean,N'Ko,Orkhon,Palmyrene,Inscriptional Pahlavi,%
2570   Psalter Pahlavi,Phoenician,Inscriptional Parthian,Samaritan,%
2571   Old South Arabian,}%
2572 \def\bbl@provide@dirs#1{%
2573   \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts\bbl@rscripts}%
2574   \ifin@
2575     \global\bbl@csarg\chardef{wdir@#1}\@ne
2576     \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts}%
2577     \ifin@
2578       \global\bbl@csarg\chardef{wdir@#1}\two  % useless in xetex
2579     \fi
2580   \else
2581     \global\bbl@csarg\chardef{wdir@#1}\z@
2582   \fi
2583   \ifodd\bbl@engine
```

```

2584 \bbl@csarg\ifcase{wdir@#1}%
2585 \directlua{ Babel.locale_props[\the\localeid].texdir = 'l' }%
2586 \or
2587 \directlua{ Babel.locale_props[\the\localeid].texdir = 'r' }%
2588 \or
2589 \directlua{ Babel.locale_props[\the\localeid].texdir = 'al' }%
2590 \fi
2591 \fi}
2592 \def\bbl@switchdir{%
2593 \bbl@ifunset{bbl@sys@\language}{\bbl@provide@sys{\language}}{}%
2594 \bbl@ifunset{bbl@wdir@\language}{\bbl@provide@dirs{\language}}{}%
2595 \bbl@exp{\bbl@setdirs\bbl@cs{wdir@\language}}%
2596 \def\bbl@setdirs#1{% TODO - math
2597 \ifcase\bbl@select@type % TODO - strictly, not the right test
2598 \bbl@bodydir{#1}%
2599 \bbl@pardir{#1}%
2600 \fi
2601 \bbl@texdir{#1}}
2602 \ifodd\bbl@engine % luatex=1
2603 \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
2604 \DisableBabelHook{babel-bidi}
2605 \chardef\bbl@thetexdir\z@
2606 \chardef\bbl@thepardir\z@
2607 \def\bbl@getluadir#1{%
2608 \directlua{
2609 if tex.#1dir == 'TLT' then
2610 tex.sprint('0')
2611 elseif tex.#1dir == 'TRT' then
2612 tex.sprint('1')
2613 end}}
2614 \def\bbl@setluadir#1#2#3{% 1=text/par.. 2=\texdir.. 3=0 lr/1 r1
2615 \ifcase#3\relax
2616 \ifcase\bbl@getluadir{#1}\relax\else
2617 #2 TLT\relax
2618 \fi
2619 \else
2620 \ifcase\bbl@getluadir{#1}\relax
2621 #2 TRT\relax
2622 \fi
2623 \fi}
2624 \def\bbl@texdir#1{%
2625 \bbl@setluadir{tex}\texdir{#1}%
2626 \chardef\bbl@thetexdir#1\relax
2627 \setattribute\bbl@attr@dir{\numexpr\bbl@thepardir*3+#1}}
2628 \def\bbl@pardir#1{%
2629 \bbl@setluadir{par}\pardir{#1}%
2630 \chardef\bbl@thepardir#1\relax}
2631 \def\bbl@bodydir{\bbl@setluadir{body}\bodydir}
2632 \def\bbl@pagedir{\bbl@setluadir{page}\pagedir}
2633 \def\bbl@dirparastext{\pardir\the\texdir\relax}% %%%
2634 % Sadly, we have to deal with boxes in math with basic.
2635 % Activated every math with the package option bidi=:
2636 \def\bbl@mathboxdir{%
2637 \ifcase\bbl@thetexdir\relax
2638 \everyhbox{\texdir TLT\relax}%
2639 \else
2640 \everyhbox{\texdir TRT\relax}%
2641 \fi}
2642 \else % pdftex=0, xetex=2

```



```

2643 \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
2644 \DisableBabelHook{babel-bidi}
2645 \newcount\bbl@dirlevel
2646 \chardef\bbl@thetextdir\z@
2647 \chardef\bbl@thepardir\z@
2648 \def\bbl@textdir#1{%
2649   \ifcase#1\relax
2650     \chardef\bbl@thetextdir\z@
2651     \bbl@textdir@i\beginL\endL
2652   \else
2653     \chardef\bbl@thetextdir\@ne
2654     \bbl@textdir@i\beginR\endR
2655   \fi}
2656 \def\bbl@textdir@i#1#2{%
2657   \ifhmode
2658     \ifnum\currentgrouplevel>\z@
2659       \ifnum\currentgrouplevel=\bbl@dirlevel
2660         \bbl@error{Multiple bidi settings inside a group}%
2661         {I'll insert a new group, but expect wrong results.}%
2662         \bgroup\aftergroup#2\aftergroup\egroup
2663       \else
2664         \ifcase\currentgrouptype\or % 0 bottom
2665           \aftergroup#2% 1 simple {}
2666         \or
2667           \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
2668         \or
2669           \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
2670         \or\or % vbox vtop align
2671         \or
2672           \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
2673         \or\or\or\or\or\or % output math disc insert vcent mathchoice
2674         \or
2675           \aftergroup#2% 14 \begingroup
2676         \else
2677           \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
2678         \fi
2679       \fi
2680       \bbl@dirlevel\currentgrouplevel
2681     \fi
2682     #1%
2683   \fi}
2684 \def\bbl@pardir#1{\chardef\bbl@thepardir#1\relax}
2685 \let\bbl@bodydir\@gobble
2686 \let\bbl@pagedir\@gobble
2687 \def\bbl@dirparastext{\chardef\bbl@thepardir\bbl@thetextdir}

```

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

```

2688 \def\bbl@xebidipar{%
2689   \let\bbl@xebidipar\relax
2690   \TeXeTstate\@ne
2691   \def\bbl@xeverypar{%
2692     \ifcase\bbl@thepardir
2693       \ifcase\bbl@thetextdir\else\beginR\fi
2694     \else
2695       {\setbox\z@\lastbox\beginR\box\z@}%
2696     \fi}%
2697   \let\bbl@severypar\everypar

```

```

2698 \newtoks\everypar
2699 \everypar=\bbl@severypar
2700 \bbl@severypar{\bbl@xeverypar\the\everypar}}
2701 \@ifpackagewith{babel}{bidi=bidi}%
2702 {\let\bbl@textdir@i@gobbletwo
2703 \let\bbl@xebidipar@empty
2704 \AddBabelHook{bidi}{foreign}{%
2705 \def\bbl@tempa{\def\BabelText###1}%
2706 \ifcase\bbl@thetextdir
2707 \expandafter\bbl@tempa\expandafter{\BabelText{\LR{##1}}}%
2708 \else
2709 \expandafter\bbl@tempa\expandafter{\BabelText{\RL{##1}}}%
2710 \fi}
2711 \def\bbl@pardir#1{\ifcase#1\relax\setLR\else\setRL\fi}}
2712 {}%
2713 \fi

```

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

```

2714 \DeclareRobustCommand\babelsublr[1]{\leavevmode{\bbl@textdir\z@#1}}
2715 \AtBeginDocument{%
2716 \ifx\pdfstringdefDisableCommands@undefined\else
2717 \ifx\pdfstringdefDisableCommands\relax\else
2718 \pdfstringdefDisableCommands{\let\babelsublr@firstofone}%
2719 \fi
2720 \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 `nor sk.cfg` will be loaded when the language definition file `nor sk.ldf` is loaded. For plain-based formats we don't want to override the definition of `\loadlocalcfg` from `plain.def`.

```

2721 \bbl@trace{Local Language Configuration}
2722 \ifx\loadlocalcfg@undefined
2723 \@ifpackagewith{babel}{noconfigs}%
2724 {\let\loadlocalcfg@gobble}%
2725 {\def\loadlocalcfg#1{%
2726 \InputIfFileExists{#1.cfg}%
2727 {\typeout{*****^J%
2728 * Local config file #1.cfg used^^J%
2729 *}}}%
2730 \@empty}}
2731 \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:

```

2732 \ifx\@unexpandable@protect@undefined
2733 \def\@unexpandable@protect{\noexpand\protect\noexpand}
2734 \long\def\protected@write#1#2#3{%
2735 \begingroup
2736 \let\thepage\relax
2737 #2%
2738 \let\protect\@unexpandable@protect
2739 \edef\reserved@a{\write#1{#3}}%
2740 \reserved@a
2741 \endgroup
2742 \if@nobreak\ifvmode\nobreak\fi\fi}

```

```

2743 \fi
2744 </core>
2745 <*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.

```

2746 <<Make sure ProvidesFile is defined>>
2747 \ProvidesFile{switch.def}[\<date>] \<version> Babel switching mechanism]
2748 <<Load macros for plain if not LaTeX>>
2749 <<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.

```

2750 \def\bbl@version{\<version>}
2751 \def\bbl@date{\<date>}
2752 \def\adddialect#1#2{%
2753   \global\chardef#1#2\relax
2754   \bbl@usehooks{adddialect}{\#1}{\#2}}%
2755   \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.

```

2756 \def\bbl@fixname#1{%
2757   \begingroup
2758     \def\bbl@tempe{l@}%
2759     \edef\bbl@tempd{\noexpand\@ifundefined{\noexpand\bbl@tempe#1}}%
2760     \bbl@tempd
2761     {\lowercase\expandafter{\bbl@tempd}}%
2762     {\uppercase\expandafter{\bbl@tempd}}%
2763     \@empty
2764     {\edef\bbl@tempd{\def\noexpand#1{\#1}}%
2765      \uppercase\expandafter{\bbl@tempd}}}%
2766     {\edef\bbl@tempd{\def\noexpand#1{\#1}}%
2767      \lowercase\expandafter{\bbl@tempd}}}%
2768     \@empty
2769     \edef\bbl@tempd{\endgroup\def\noexpand#1{\#1}}%
2770   \bbl@tempd}
2771 \def\bbl@iflanguage#1{%
2772   \@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.

```

2773 \def\iflanguage#1{%
2774   \bbl@iflanguage{\#1}{%

```

```

2775 \ifnum\csname l@#1\endcsname=\language
2776 \expandafter\@firstoftwo
2777 \else
2778 \expandafter\@secondoftwo
2779 \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.

```

2780 \let\bbl@select@type\z@
2781 \edef\selectlanguage{%
2782 \noexpand\protect
2783 \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`.

```

2784 \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.

```

2785 \ifx\documentclass\@undefined
2786 \def\xstring{\string\string\string}
2787 \else
2788 \let\xstring\string
2789 \fi

```

Since version 3.5 babel writes entries to the auxiliary files in order to typeset table of contents etc. in the correct language environment.

`\bbl@pop@language` But when the language change happens *inside* a group the end of the group doesn't write anything to the auxiliary files. Therefore we need  $\TeX$ 's `aftergroup` mechanism to help us. The command `\aftergroup` stores the token immediately following it to be executed when the current group is closed. So we define a temporary control sequence `\bbl@pop@language` to be executed at the end of the group. It calls `\bbl@set@language` with the name of the current language as its argument.

`\bbl@language@stack` The previous solution works for one level of nesting groups, but as soon as more levels are used it is no longer adequate. For that case we need to keep track of the nested languages using a stack mechanism. This stack is called `\bbl@language@stack` and initially empty.

```
2790 \def\bbl@language@stack{}
```

When using a stack we need a mechanism to push an element on the stack and to retrieve the information afterwards.

`\bbl@push@language` The stack is simply a list of languagenames, separated with a ‘+’ sign; the push function can be simple:  
`\bbl@pop@language`

```
2791 \def\bbl@push@language{%
2792   \xdef\bbl@language@stack{\language+\bbl@language@stack}}
```

Retrieving information from the stack is a little bit less simple, as we need to remove the element from the stack while storing it in the macro `\language`. For this we first define a helper function.

`\bbl@pop@lang` This macro stores its first element (which is delimited by the ‘+’-sign) in `\language` and stores the rest of the string (delimited by ‘-’) in its third argument.

```
2793 \def\bbl@pop@lang#1+#2-#3{%
2794   \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.

```
2795 \let\bbl@ifrestoring\@secondoftwo
2796 \def\bbl@pop@language{%
2797   \expandafter\bbl@pop@lang\bbl@language@stack-\bbl@language@stack
2798   \let\bbl@ifrestoring\@firstoftwo
2799   \expandafter\bbl@set@language\expandafter{\language}%
2800   \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.

```
2801 \chardef\localeid\z@
2802 \def\bbl@id@last{0} % No real need for a new counter
2803 \def\bbl@id@assign{%
2804   \bbl@ifunset{\bbl@id@@\language}%
2805   {\count@\bbl@id@last\relax
2806    \advance\count@\@ne
2807    \bbl@csarg\chardef{id@@\language}\count@
2808    \edef\bbl@id@last{\the\count@}%
2809    \ifcase\bbl@engine\or
2810      \directlua{
2811        Babel = Babel or {}
2812        Babel.locale_props = Babel.locale_props or {}
2813        Babel.locale_props[\bbl@id@last] = {}
2814      }%
2815    \fi}%
2816   {}}
```

The unprotected part of \selectlanguage.

```
2817 \expandafter\def\csname selectlanguage \endcsname#1{%
2818   \ifnum\babel@hymapsel=\@cclv\let\babel@hymapsel\tw\fi
2819   \babel@push@language
2820   \aftergroup\babel@pop@language
2821   \babel@set@language{#1}}
```

\babel@set@language The macro \babel@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.

```
2822 \def\BabelContentsFiles{toc,lof,lot}
2823 \def\babel@set@language#1{% from selectlanguage, pop@
2824   \edef\language{%
2825     \ifnum\escapechar=\expandafter`\string#1\@empty
2826     \else\string#1\@empty\fi}%
2827   \select@language{\language}%
2828   % write to auxs
2829   \expandafter\ifx\csname date\language\endcsname\relax\else
2830     \if@files
2831       \ifx\babel@aux\@gobbletwo\else % Set if single in the first, redundant
2832         \protected@write\@auxout{}\string\babel@aux{\language}{}}%
2833       \fi
2834       \babel@usehooks{write}{}%
2835     \fi
2836   \fi}
2837 \def\select@language#1{% from set@, babel@aux
2838   % set hmap
2839   \ifnum\babel@hymapsel=\@cclv\chardef\babel@hymapsel4\relax\fi
2840   % set name
2841   \edef\language{#1}%
2842   \babel@fixname\language
2843   \babel@iflanguage\language{%
2844     \expandafter\ifx\csname date\language\endcsname\relax
2845       \babel@error
2846       {Unknown language `#1'. Either you have\\%
2847        misspelled its name, it has not been installed,\\%
2848        or you requested it in a previous run. Fix its name,\\%
2849        install it or just rerun the file, respectively. In\\%
2850        some cases, you may need to remove the aux file}%
2851       {You may proceed, but expect wrong results}%
2852     \else
2853       % set type
2854       \let\babel@select@type\z@
2855       \expandafter\babel@switch\expandafter{\language}%
2856     \fi}}
2857 \def\babel@aux#1#2{%
2858   \expandafter\ifx\csname date#1\endcsname\relax
2859     \expandafter\ifx\csname babel@auxwarn#1\endcsname\relax
2860       \@namedef{babel@auxwarn#1}{}%
2861       \babel@warning
2862       {Unknown language `#1'. Very likely you\\%
2863        requested it in a previous run. Expect some\\%
2864        wrong results in this run, which should vanish\\%
```

```

2865         in the next one. Reported}%
2866     \fi
2867 \else
2868     \select@language{#1}%
2869     \bbl@foreach\BabelContentsFiles{%
2870         \@writefile{##1}{\babel@toc{#1}{#2}}}% %% TODO - ok in plain?
2871 \fi}
2872 \def\babel@toc#1#2{%
2873     \select@language{#1}}

```

A bit of optimization. Select in heads/foots the language only if necessary. The real thing is in `babel.def`.

```

2874 \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.

```

2875 \newif\ifbbl@usedategroup
2876 \def\bbl@switch#1{% from select@, foreign@
2877     % restore
2878     \originalTeX
2879     \expandafter\def\expandafter\originalTeX\expandafter{%
2880         \csname noextras#1\endcsname
2881         \let\originalTeX\@empty
2882         \babel@beginsave}%
2883 \bbl@usehooks{afterreset}{}%
2884 \languageshorthands{none}%
2885 % set the locale id
2886 \bbl@id@assign
2887 \chardef\localeid\@nameuse{bbl@id@\@language}%
2888 % switch captions, date
2889 \ifcase\bbl@select@type
2890     \ifhmode
2891         \hskip\z@skip % trick to ignore spaces
2892         \csname captions#1\endcsname\relax
2893         \csname date#1\endcsname\relax
2894         \loop\ifdim\lastskip>\z@\@unskip\repeat\unskip
2895     \else
2896         \csname captions#1\endcsname\relax
2897         \csname date#1\endcsname\relax
2898     \fi
2899 \else
2900     \ifbbl@usedategroup % if \foreign... within \<lang>date
2901         \bbl@usedategroupfalse
2902         \ifhmode
2903             \hskip\z@skip % trick to ignore spaces

```

```

2904      \csname date#1\endcsname\relax
2905      \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
2906      \else
2907      \csname date#1\endcsname\relax
2908      \fi
2909    \fi
2910  \fi
2911  % switch extras
2912  \bbl@usehooks{beforeextras}{}%
2913  \csname extras#1\endcsname\relax
2914  \bbl@usehooks{afterextras}{}%
2915  % > babel-ensure
2916  % > babel-sh-<short>
2917  % > babel-bidi
2918  % > babel-fontspec
2919  % hyphenation - case mapping
2920  \ifcase\bbl@opt@hyphenmap\or
2921    \def\BabelLower##1##2{\lccode##1=##2\relax}%
2922    \ifnum\bbl@hymapsel>4\else
2923      \csname\language\name @bbl@hyphenmap\endcsname
2924      \fi
2925      \chardef\bbl@opt@hyphenmap\z@
2926  \else
2927    \ifnum\bbl@hymapsel>\bbl@opt@hyphenmap\else
2928      \csname\language\name @bbl@hyphenmap\endcsname
2929      \fi
2930  \fi
2931  \global\let\bbl@hymapsel@cclv
2932  % hyphenation - patterns
2933  \bbl@patterns{#1}%
2934  % hyphenation - mins
2935  \babel@savevariable\lefthyphenmin
2936  \babel@savevariable\righthyphenmin
2937  \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2938    \set@hyphenmins\tw@\thr@@\relax
2939  \else
2940    \expandafter\expandafter\expandafter\set@hyphenmins
2941    \csname #1hyphenmins\endcsname\relax
2942  \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.

```

2943 \long\def\otherlanguage#1{%
2944   \ifnum\bbl@hymapsel=\cclv\let\bbl@hymapsel\thr@@\fi
2945   \csname selectlanguage \endcsname{#1}%
2946   \ignorespaces}

```

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

```

2947 \long\def\endotherlanguage{%
2948   \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`.

```
2949 \expandafter\def\csname otherlanguage*\endcsname#1{%
2950   \ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi
2951   \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”.

```
2952 \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`.

```
2953 \providecommand\bbl@beforeforeign{}
2954 \edef\foreignlanguage{%
2955   \noexpand\protect
2956   \expandafter\noexpand\csname foreignlanguage \endcsname}
2957 \expandafter\def\csname foreignlanguage \endcsname{%
2958   \@ifstar\bbl@foreign@s\bbl@foreign@x}
2959 \def\bbl@foreign@x#1#2{%
2960   \begingroup
2961     \let\BabelText\@firstofone
2962     \bbl@beforeforeign
2963     \foreign@language{#1}%
2964     \bbl@usehooks{foreign}{}%
2965     \BabelText{#2}% Now in horizontal mode!
2966   \endgroup}
2967 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \@setpar, ?\@par
2968   \begingroup
2969     {\par}%
2970     \let\BabelText\@firstofone
2971     \foreign@language{#1}%
2972     \bbl@usehooks{foreign*}{}%
2973     \bbl@dirparastext
2974     \BabelText{#2}% Still in vertical mode!
2975     {\par}%
2976   \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`.

```

2977 \def\foreign@language#1{%
2978   % set name
2979   \edef\language#1}%
2980   \bbl@fixname\language
2981   \bbl@iflanguage\language{%
2982     \expandafter\ifx\csname date\language\endcsname\relax
2983       \bbl@warning % TODO - why a warning, not an error?
2984       {Unknown language `#1'. Either you have\\%
2985         misspelled its name, it has not been installed,\\%
2986         or you requested it in a previous run. Fix its name,\\%
2987         install it or just rerun the file, respectively. In\\%
2988         some cases, you may need to remove the aux file.\\%
2989         I'll proceed, but expect wrong results.\\%
2990         Reported}%
2991     \fi
2992     % set type
2993     \let\bbl@select@type\@ne
2994     \expandafter\bbl@switch\expandafter{\language}}

```

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

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

```

2995 \let\bbl@hyphlist\@empty
2996 \let\bbl@hyphenation@\relax
2997 \let\bbl@pttnlist\@empty
2998 \let\bbl@patterns@\relax
2999 \let\bbl@hymapsel=\@cclv
3000 \def\bbl@patterns#1{%
3001   \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
3002     \csname l@#1\endcsname
3003     \edef\bbl@tempa{#1}%
3004   \else
3005     \csname l@#1:\f@encoding\endcsname
3006     \edef\bbl@tempa{#1:\f@encoding}%
3007   \fi
3008   \@expandtwoargs\bbl@usehooks{patterns}{#1}{\bbl@tempa}%
3009   % > luatex
3010   \@ifundefined{bbl@hyphenation@}{% Can be \relax!
3011     \begingroup
3012       \bbl@xin@{,\number\language,}{,\bbl@hyphlist}%
3013     \ifin@else
3014       \@expandtwoargs\bbl@usehooks{hyphenation}{#1}{\bbl@tempa}%
3015       \hyphenation{%
3016         \bbl@hyphenation@
3017         \@ifundefined{bbl@hyphenation@#1}%
3018         \@empty
3019         {\space\csname bbl@hyphenation@#1\endcsname}}%
3020     \xdef\bbl@hyphlist{\bbl@hyphlist\number\language,}%

```

```

3021     \fi
3022 \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*`.

```

3023 \def\hyphenrules#1{%
3024   \edef\bbl@tempf{#1}%
3025   \bbl@fixname\bbl@tempf
3026   \bbl@iflanguage\bbl@tempf{%
3027     \expandafter\bbl@patterns\expandafter{\bbl@tempf}%
3028     \languageshortands{none}%
3029     \expandafter\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
3030       \set@hyphenmins\tw@\thr@@\relax
3031     \else
3032       \expandafter\expandafter\expandafter\set@hyphenmins
3033       \csname\bbl@tempf hyphenmins\endcsname\relax
3034     \fi}}
3035 \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.

```

3036 \def\providehyphenmins#1#2{%
3037   \expandafter\ifx\csname #1hyphenmins\endcsname\relax
3038     \@namedef{#1hyphenmins}{#2}%
3039   \fi}

```

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

```

3040 \def\set@hyphenmins#1#2{%
3041   \lefthyphenmin#1\relax
3042   \righthyphenmin#2\relax}

```

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

```

3043 \ifx\ProvidesFile\@undefined
3044   \def\ProvidesLanguage#1[#2 #3 #4]{%
3045     \wlog{Language: #1 #4 #3 <#2>}%
3046   }
3047 \else
3048   \def\ProvidesLanguage#1{%
3049     \begingroup
3050     \catcode`\ 10 %
3051     \@makeother\%
3052     \ifnextchar[%]
3053       {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}
3054   \def\@provideslanguage#1[#2]{%
3055     \wlog{Language: #1 #2}%
3056     \expandafter\xdef\csname ver@#1.1df\endcsname{#2}%
3057   \endgroup}
3058 \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`

```
3059 \def\LdfInit{%
3060   \chardef\atcatcode=\catcode` \@
3061   \catcode`\@=11\relax
3062   \input babel.def\relax
3063   \catcode`\@=\atcatcode \let\atcatcode\relax
3064   \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`.

```
3065 \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.

```
3066 \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’:

```
3067 \providecommand\setlocale{%
3068   \bbl@error
3069   {Not yet available}%
3070   {Find an armchair, sit down and wait}}
3071 \let\uselocale\setlocale
3072 \let\locale\setlocale
3073 \let\selectlocale\setlocale
3074 \let\textlocale\setlocale
3075 \let\textlanguage\setlocale
3076 \let\languagegetext\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’.

```
3077 \edef\bbl@nulllanguage{\string\language=0}
3078 \ifx\PackageError\undefined
3079   \def\bbl@error#1#2{%
3080     \begingroup
3081     \newlinechar=`^^J
3082     \def\{^^J(babel) }%
3083     \errhelp{#2}\errmessage{\{#1}%
3084     \endgroup}
3085   \def\bbl@warning#1{%
```

```

3086 \begingroup
3087 \newlinechar=`^^J
3088 \def\{^^J(babel) }%
3089 \message{\#1}%
3090 \endgroup}
3091 \def\bbl@info#1{%
3092 \begingroup
3093 \newlinechar=`^^J
3094 \def\{^^J}%
3095 \wlog{#1}%
3096 \endgroup}
3097 \else
3098 \def\bbl@error#1#2{%
3099 \begingroup
3100 \def\{MessageBreak}%
3101 \PackageError{babel}{#1}{#2}%
3102 \endgroup}
3103 \def\bbl@warning#1{%
3104 \begingroup
3105 \def\{MessageBreak}%
3106 \PackageWarning{babel}{#1}%
3107 \endgroup}
3108 \def\bbl@info#1{%
3109 \begingroup
3110 \def\{MessageBreak}%
3111 \PackageInfo{babel}{#1}%
3112 \endgroup}
3113 \fi
3114 \@ifpackagewith{babel}{silent}
3115 {\let\bbl@info@gobble
3116 \let\bbl@warning@gobble}
3117 {}
3118 \def\bbl@nocaption{\protect\bbl@nocaption@i}
3119 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
3120 \global\@namedef{#2}{\textbf{?#1?}}%
3121 \@nameuse{#2}%
3122 \bbl@warning{%
3123 \@backslashchar#2 not set. Please, define\\%
3124 it in the preamble with something like:\\%
3125 \string\renewcommand\@backslashchar#2{..}\\%
3126 Reported}}
3127 \def\bbl@tentative{\protect\bbl@tentative@i}
3128 \def\bbl@tentative@i#1{%
3129 \bbl@warning{%
3130 Some functions for '#1' are tentative.\\%
3131 They might not work as expected and their behavior\\%
3132 could change in the future.\\%
3133 Reported}}
3134 \def\@nolanerr#1{%
3135 \bbl@error
3136 {You haven't defined the language #1\space yet}%
3137 {Your command will be ignored, type <return> to proceed}}
3138 \def\@nopatterns#1{%
3139 \bbl@warning
3140 {No hyphenation patterns were preloaded for\\%
3141 the language '#1' into the format.\\%
3142 Please, configure your TeX system to add them and\\%
3143 rebuild the format. Now I will use the patterns\\%
3144 preloaded for \bbl@nulllanguage\space instead}}

```

```

3145 \let\bbl@usehooks\@gobbletwo
3146 \</kernel>
3147 \<*patterns>

```

## 12 Loading hyphenation patterns

The following code is meant to be read by  $\text{\texttt{iniT\TeX}}$  because it should instruct  $\text{\texttt{T\TeX}}$  to read hyphenation patterns. To this end the `docstrip` option `patterns` can be used to include this code in the file `hyphen.cfg`. Code is written with lower level macros. We want to add a message to the message  $\text{\texttt{L\TeX}}$  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  $\text{\texttt{L\TeX}}$  fills the register. There are some problems with this approach though.

- When someone wants to use several hyphenation patterns with  $\text{\texttt{SL\TeX}}$  the above scheme won't work. The reason is that  $\text{\texttt{SL\TeX}}$  overwrites the contents of the `\everyjob` register with its own message.
- Plain  $\text{\texttt{T\TeX}}$  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  $\text{\texttt{L\TeX}}$  2.09 executes the `\@begindocumenthook` we would want to alter `\begin{document}`, but as this done too often already, we add the new code at the front of `\@preamblecmds`. But we can only do that after it has been defined, so we add this piece of code to `\dump`.

This new definition starts by adding an instruction to write a message on the terminal and in the transcript file to inform the user of the preloaded hyphenation patterns. Then everything is restored to the old situation and the format is dumped.

```

3148 <<Make sure ProvidesFile is defined>>
3149 \ProvidesFile{hyphen.cfg}[\<date>] [\<version>] Babel hyphens]
3150 \xdef\bbl@format{\jobname}
3151 \ifx\AtBeginDocument\@undefined
3152   \def\@empty{}
3153   \let\orig@dump\dump
3154   \def\dump{%
3155     \ifx\@ztryfc\@undefined
3156     \else
3157       \toks0=\expandafter{\@preamblecmds}%
3158       \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
3159       \def\@begindocumenthook{}%
3160     \fi
3161     \let\dump\orig@dump\let\orig@dump\@undefined\dump}
3162 \fi
3163 <<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.

```

3164 \def\process@line#1#2 #3 #4 {%
3165   \ifx=#1%
3166     \process@synonym{#2}%
3167   \else
3168     \process@language{#1#2}{#3}{#4}%
3169   \fi
3170   \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.

```

3171 \toks@{}
3172 \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.

```

3173 \def\process@synonym#1{%
3174   \ifnum\last@language=\m@ne
3175     \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
3176   \else
3177     \expandafter\chardef\csname l@#1\endcsname\last@language
3178     \wlog{\string\l@#1=\string\language\the\last@language}%
3179     \expandafter\let\csname #1hyphenmins\endcsname\expandafter\endcsname
3180     \csname\language\endcsname hyphenmins\endcsname
3181     \let\bbl@elt\relax
3182     \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}}}%
3183   \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 `\langhyphenmins` macro. When no assignments were made we provide a default setting.

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

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

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

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

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

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

```

3184 \def\process@language#1#2#3{%
3185   \expandafter\addlanguage\csname l@#1\endcsname
3186   \expandafter\language\csname l@#1\endcsname
3187   \edef\language#1{%
3188     \bbl@hook@everylanguage{#1}%
3189     % > luatex
3190     \bbl@get@enc#1::\@@@
3191   \begingroup
3192     \lefthyphenmin\m@ne
3193     \bbl@hook@loadpatterns{#2}%
3194     % > luatex
3195     \ifnum\lefthyphenmin=\m@ne
3196     \else
3197       \expandafter\xdef\csname #1hyphenmins\endcsname{%
3198         \the\lefthyphenmin\the\righthyphenmin}%
3199     \fi
3200   \endgroup
3201   \def\bbl@tempa{#3}%
3202   \ifx\bbl@tempa\@empty\else
3203     \bbl@hook@loadexceptions{#3}%
3204     % > luatex
3205   \fi
3206   \let\bbl@elt\relax
3207   \edef\bbl@languages{%
3208     \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
3209   \ifnum\the\language=\z@
3210     \expandafter\ifx\csname #1hyphenmins\endcsname\relax
3211       \set@hyphenmins\tw@\thr@@\relax
3212     \else
3213       \expandafter\expandafter\expandafter\set@hyphenmins
3214       \csname #1hyphenmins\endcsname
3215     \fi
3216     \the\toks@
3217     \toks@{}%
3218   \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.

```

3219 \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.

```

3220 \def\bbl@hook@everylanguage#1{}
3221 \def\bbl@hook@loadpatterns#1{\input #1\relax}
3222 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
3223 \let\bbl@hook@loadkernel\bbl@hook@loadpatterns
3224 \begingroup
3225   \def\AddBabelHook#1#2{%

```



```

3226 \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
3227 \def\next{\toks1}%
3228 \else
3229 \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname####1}%
3230 \fi
3231 \next}
3232 \ifx\directlua\undefined
3233 \ifx\XeTeXinputencoding\undefined\else
3234 \input xebabel.def
3235 \fi
3236 \else
3237 \input luababel.def
3238 \fi
3239 \openin1 = babel-\bbl@format.cfg
3240 \ifeof1
3241 \else
3242 \input babel-\bbl@format.cfg\relax
3243 \fi
3244 \closein1
3245 \endgroup
3246 \bbl@hook@loadkernel{switch.def}

```

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

```

3247 \openin1 = language.dat

```

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

```

3248 \def\language{english}%
3249 \ifeof1
3250 \message{I couldn't find the file language.dat,\space
3251         I will try the file hyphen.tex}
3252 \input hyphen.tex\relax
3253 \chardef\l@english\z@
3254 \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.

```

3255 \last@language@m@ne

```

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

```

3256 \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.

```

3257 \endlinechar@m@ne
3258 \read1 to \bbl@line
3259 \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.

```

3260 \if T\ifeof1F\fi T\relax
3261 \ifx\bbl@line\@empty\else
3262 \edef\bbl@line{\bbl@line\space\space\space}%
3263 \expandafter\process@line\bbl@line\relax
3264 \fi
3265 \repeat

```

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

```

3266 \begingroup
3267   \def\bbl@elt#1#2#3#4{%
3268     \global\language=#2\relax
3269     \gdef\language{#1}%
3270     \def\bbl@elt##1##2##3##4{}}%
3271   \bbl@languages
3272 \endgroup
3273 \fi

```

and close the configuration file.

```

3274 \closein1

```

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

```

3275 \if/\the\toks@/\else
3276   \errhelp{language.dat loads no language, only synonyms}
3277   \errmessage{Orphan language synonym}
3278 \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.

```

3279 \let\bbl@line\@undefined
3280 \let\process@line\@undefined
3281 \let\process@synonym\@undefined
3282 \let\process@language\@undefined
3283 \let\bbl@get@enc\@undefined
3284 \let\bbl@hyph@enc\@undefined
3285 \let\bbl@tempa\@undefined
3286 \let\bbl@hook@loadkernel\@undefined
3287 \let\bbl@hook@everylanguage\@undefined
3288 \let\bbl@hook@loadpatterns\@undefined
3289 \let\bbl@hook@loadexceptions\@undefined
3290 \</patterns>

```

Here the code for `iniTEX` ends.

## 13 Font handling with fontspec

Add the bidi handler just before `luaotfload`, which is loaded by default by LaTeX. Just in case, consider the possibility it has not been loaded. First, a couple of definitions related to bidi [misplaced].

```

3291 <<{*More package options}>> ≡
3292 \ifodd\bbl@engine
3293   \DeclareOption{bidi=basic-r}%
3294   {\ExecuteOptions{bidi=basic}}
3295   \DeclareOption{bidi=basic}%
3296   {\let\bbl@beforeforeign\leavevmode
3297     % TODO - to locale_props, not as separate attribute
3298     \newattribute\bbl@attr@dir
3299     % I don't like it, hackish:
3300     \frozen@everymath\expandafter{%
3301       \expandafter\bbl@mathboxdir\the\frozen@everymath}%
3302     \frozen@everydisplay\expandafter{%
3303       \expandafter\bbl@mathboxdir\the\frozen@everydisplay}%

```

```

3304 \bbl@exp{\output{\bodydir\pagedir\the\output}}%
3305 \AtEndOfPackage{\EnableBabelHook{babel-bidi}}%
3306 \else
3307 \DeclareOption{bidi=basic-r}%
3308 {\ExecuteOptions{bidi=basic}}
3309 \DeclareOption{bidi=basic}%
3310 {\bbl@error
3311 {The bidi method 'basic' is available only in\%
3312 luatex. I'll continue with 'bidi=default', so\%
3313 expect wrong results}%
3314 {See the manual for further details.}%
3315 \let\bbl@beforeforeign\leavevmode
3316 \AtEndOfPackage{%
3317 \EnableBabelHook{babel-bidi}%
3318 \bbl@xebidipar}}
3319 \def\bbl@loadxebidi#1{%
3320 \ifx\RTLfootnotetext\@undefined
3321 \AtEndOfPackage{%
3322 \EnableBabelHook{babel-bidi}%
3323 \ifx\fontspec\@undefined
3324 \usepackage{fontspec}% bidi needs fontspec
3325 \fi
3326 \usepackage#1{bidi}}%
3327 \fi}
3328 \DeclareOption{bidi=bidi}%
3329 {\bbl@tentative{bidi=bidi}%
3330 \bbl@loadxebidi{}}
3331 \DeclareOption{bidi=bidi-r}%
3332 {\bbl@tentative{bidi=bidi-r}%
3333 \bbl@loadxebidi{[rldocument]}}
3334 \DeclareOption{bidi=bidi-l}%
3335 {\bbl@tentative{bidi=bidi-l}%
3336 \bbl@loadxebidi{}}
3337 \fi
3338 \DeclareOption{bidi=default}%
3339 {\let\bbl@beforeforeign\leavevmode
3340 \ifodd\bbl@engine
3341 \newattribute\bbl@attr@dir
3342 \bbl@exp{\output{\bodydir\pagedir\the\output}}%
3343 \fi
3344 \AtEndOfPackage{%
3345 \EnableBabelHook{babel-bidi}%
3346 \ifodd\bbl@engine\else
3347 \bbl@xebidipar
3348 \fi}}
3349 <</More package options>>

```

With explicit languages, we could define the font at once, but we don't. Just wait and see if the language is actually activated. `bbl@font` replaces hardcoded font names inside `\. . family` by the corresponding macro `\. . default`.

```

3350 <<{*Font selection}>> ≡
3351 \bbl@trace{Font handling with fontspec}
3352 \@onlypreamble\babelfont
3353 \newcommand\babelfont[2][]{% 1=langs/scripts 2=fam
3354 \edef\bbl@tempa{#1}%
3355 \def\bbl@tempb{#2}% Used by \bbl@bblfont
3356 \ifx\fontspec\@undefined
3357 \usepackage{fontspec}%
3358 \fi

```

```

3359 \EnableBabelHook{babel-fontspec}% Just calls \bbl@switchfont
3360 \bbl@bblfont}
3361 \newcommand\bbl@bblfont[2][]{% 1=features 2=fontname, @font=rm|sf|tt
3362 \bbl@ifunset{\bbl@tempb family}%
3363 {\bbl@providfam{\bbl@tempb}}%
3364 {\bbl@exp{%
3365 \\\bbl@sreplace\<\bbl@tempb family >%
3366 {\@nameuse{\bbl@tempb default}}{\<\bbl@tempb default>}}}%
3367 % For the default font, just in case:
3368 \bbl@ifunset{\bbl@lsys@\language}\bbl@provide@lsys{\language}}{%
3369 \expandafter\bbl@ifblank\expandafter{\bbl@tempa}%
3370 {\bbl@csarg\edef{\bbl@tempb dflt@}{<{#1}{#2}}% save bbl@rmdflt@
3371 \bbl@exp{%
3372 \let\<bbl@\bbl@tempb dflt@\language>\<bbl@\bbl@tempb dflt@>%
3373 \\\bbl@font@set\<bbl@\bbl@tempb dflt@\language>%
3374 \<\bbl@tempb default>\<\bbl@tempb family>}}%
3375 {\bbl@foreach\bbl@tempa{% ie bbl@rmdflt@lang / *scrt
3376 \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:

```

3377 \def\bbl@providfam#1{%
3378 \bbl@exp{%
3379 \\\newcommand\<#1default>{}% Just define it
3380 \\\bbl@add@list\\bbl@font@fams{#1}%
3381 \\\DeclareRobustCommand\<#1family>%
3382 \\\not@math@alphabet\<#1family>\relax
3383 \\\fontfamily\<#1default>\selectfont}%
3384 \\\DeclareTextFontCommand{\<text#1>}{\<#1family>}}

```

The following macro is activated when the hook babel-fontspec is enabled. But before we define a macro for a warning, which sets a flag to avoid duplicate them.

```

3385 \def\bbl@nostdfont#1{%
3386 \bbl@ifunset{\bbl@WFF@\f@family}%
3387 {\bbl@csarg\gdef{\WFF@\f@family}}{% Flag, to avoid dupl warns
3388 \bbl@warning{The current font is not a babel standard family:\%
3389 #1%
3390 \fontname\font\%
3391 There is nothing intrinsically wrong with this warning, and\%
3392 you can ignore it altogether if you do not need these\%
3393 families. But if they are used in the document, you should be\%
3394 aware 'babel' will no set Script and Language for them, so\%
3395 you may consider defining a new family with \string\babelfont.\%
3396 See the manual for further details about \string\babelfont.\%
3397 Reported}}
3398 {}}%
3399 \gdef\bbl@switchfont{%
3400 \bbl@ifunset{\bbl@lsys@\language}\bbl@provide@lsys{\language}}{%
3401 \bbl@exp{% eg Arabic -> arabic
3402 \lowercase{\edef\\bbl@tempa{\bbl@cs{sname@\language}}}}%
3403 \bbl@foreach\bbl@font@fams{%
3404 \bbl@ifunset{\bbl@##1dflt@\language}% (1) language?
3405 {\bbl@ifunset{\bbl@##1dflt@*\bbl@tempa}% (2) from script?
3406 {\bbl@ifunset{\bbl@##1dflt@}% 2=F - (3) from generic?
3407 {}% 123=F - nothing!
3408 {\bbl@exp{% 3=T - from generic
3409 \global\let\<bbl@##1dflt@\language>%
3410 \<bbl@##1dflt@>}}}%
3411 {\bbl@exp{% 2=T - from script
3412 \global\let\<bbl@##1dflt@\language>%

```

```

3413 \<bbl@##1dflt@*\bbl@tempa>}}}%
3414 {}}% 1=T - language, already defined
3415 \def\bbl@tempa{\bbl@nostdfont{}}}%
3416 \bbl@foreach\bbl@font@fams{% don't gather with prev for
3417 \bbl@ifunset{\bbl@##1dflt@\language}%
3418 {\bbl@cs{famrst@##1}%
3419 \global\bbl@csarg\let{famrst@##1}\relax}%
3420 {\bbl@exp{% order is relevant
3421 \\\bbl@add\\\originalTeX{%
3422 \\\bbl@font@rst{\bbl@cs{##1dflt@\language}}}%
3423 \<##1default>\<##1family>{##1}}}%
3424 \\\bbl@font@set\bbl@##1dflt@\language>% the main part!
3425 \<##1default>\<##1family>}}}%
3426 \bbl@ifrestoring{}\bbl@tempa}}}%

```

The following is executed at the beginning of the aux file or the document to warn about fonts not defined with `\babelfont`.

```

3427 \ifx\f@family\undefined\else % if latex
3428 \ifcase\bbl@engine % if pdftex
3429 \let\bbl@ckeckstdfonts\relax
3430 \else
3431 \def\bbl@ckeckstdfonts{%
3432 \begingroup
3433 \global\let\bbl@ckeckstdfonts\relax
3434 \let\bbl@tempa\@empty
3435 \bbl@foreach\bbl@font@fams{%
3436 \bbl@ifunset{\bbl@##1dflt@}%
3437 {\@nameuse{##1family}%
3438 \bbl@csarg\gdef{WFF@f@family}}}% Flag
3439 \bbl@exp{\\\bbl@add\\\bbl@tempa{* \<##1family> / \f@family\\\%
3440 \space\space\fontname\font\\\}}}%
3441 \bbl@csarg\xdef{##1dflt@}{\f@family}%
3442 \expandafter\xdef\csname ##1default\endcsname{\f@family}}}%
3443 {}}%
3444 \ifx\bbl@tempa\@empty\else
3445 \bbl@warning{The following fonts are not babel standard families:\\%
3446 \bbl@tempa
3447 There is nothing intrinsically wrong with it, but\\%
3448 'babel' will no set Script and Language. Consider\\%
3449 defining a new family with \string\babelfont.\\%
3450 Reported}%
3451 \fi
3452 \endgroup}
3453 \fi
3454 \fi

```

Now the macros defining the font with `fontspec`.

When there are repeated keys in `fontspec`, the last value wins. So, we just place the ini settings at the beginning, and user settings will take precedence. We must deactivate temporarily `\bbl@mapselect` because `\selectfont` is called internally when a font is defined.

```

3455 \def\bbl@font@set#1#2#3{% eg \bbl@rmdflt@lang \rmdefault \rmfamily
3456 \bbl@xin@{<>}{#1}%
3457 \ifin@
3458 \bbl@exp{\\\bbl@fontspec@set\\#1\expandafter\@gobbletwo#1\\#3}%
3459 \fi
3460 \bbl@exp{%
3461 \def\\#2{#1}% eg, \rmdefault{\bbl@rmdflt@lang}
3462 \\\bbl@ifsamestring{#2}{\f@family}{\\#3\let\\bbl@tempa\relax}}}%

```

```

3463%      TODO - next should be global?, but even local does its job. I'm
3464%      still not sure -- must investigate:
3465\def\bb1@fontspec@set#1#2#3#4{% eg \bb1@rmdflt@lang fnt-opt fnt-nme \xxfamily
3466  \let\bb1@tempe\bb1@mapselect
3467  \let\bb1@mapselect\relax
3468  \let\bb1@temp@fam#4%      eg, '\rmfamily', to be restored below
3469  \let#4\relax      % So that can be used with \newfontfamily
3470  \bb1@exp{%
3471    \let\\bb1@temp@pfam\<\bb1@stripslash#4\space>% eg, '\rmfamily '
3472    \<keys_if_exist:nnF>{\fontspec-opentype}%
3473      {Script/\bb1@cs{sname@\language}}}%
3474      {\newfontscript{\bb1@cs{sname@\language}}}%
3475      {\bb1@cs{sotf@\language}}}%
3476    \<keys_if_exist:nnF>{\fontspec-opentype}%
3477      {Language/\bb1@cs{lname@\language}}}%
3478      {\newfontlanguage{\bb1@cs{lname@\language}}}%
3479      {\bb1@cs{lotf@\language}}}%
3480    \newfontfamily\\#4%
3481    [\bb1@cs{lsys@\language},#2]{#3}% ie \bb1@exp{.}{#3}
3482  \begingroup
3483    #4%
3484    \xdef#1{\f@family}%      eg, \bb1@rmdflt@lang{FreeSerif(0)}
3485  \endgroup
3486  \let#4\bb1@temp@fam
3487  \bb1@exp{\let\<\bb1@stripslash#4\space>}\bb1@temp@pfam
3488  \let\bb1@mapselect\bb1@tempe}%

```

font@rst and famrst are only used when there is no global settings, to save and restore de previous families. Not really necessary, but done for optimization.

```

3489\def\bb1@font@rst#1#2#3#4{%
3490  \bb1@csarg\def{famrst@#4}{\bb1@font@set{#1}#2#3}}

```

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

```

3491\def\bb1@font@fams{rm,sf,tt}

```

The old tentative way. Short and preverved for compatibility, but deprecated. Note there is no direct alternative for \babelFSfeatures. The reason in explained in the user guide, but essentially – that was not the way to go :-).

```

3492\newcommand\babelFSstore[2][{%
3493  \bb1@ifblank{#1}%
3494    {\bb1@csarg\def{sname@#2}{Latin}}%
3495    {\bb1@csarg\def{sname@#2}{#1}}%
3496  \bb1@provide@dirs{#2}%
3497  \bb1@csarg\ifnum{wdir@#2}>\z@
3498    \let\bb1@beforeforeign\leavevmode
3499    \EnableBabelHook{babel-bidi}%
3500  \fi
3501  \bb1@foreach{#2}{%
3502    \bb1@FSstore{##1}{rm}\rmdefault\bb1@save@rmdefault
3503    \bb1@FSstore{##1}{sf}\sfdefault\bb1@save@sfdefault
3504    \bb1@FSstore{##1}{tt}\ttdefault\bb1@save@ttdefault}}
3505\def\bb1@FSstore#1#2#3#4{%
3506  \bb1@csarg\edef{#2default#1}{#3}%
3507  \expandafter\addto\csname extras#1\endcsname{%
3508    \let#4#3%
3509    \ifx#3\f@family
3510      \edef#3{\csname \bb1@#2default#1\endcsname}%
3511      \fontfamily{#3}\selectfont

```

```

3512 \else
3513 \edef#3{\csname bbl@#2default#1\endcsname}%
3514 \fi}%
3515 \expandafter\addto\csname noextras#1\endcsname{%
3516 \ifx#3\fontfamily
3517 \fontfamily{#4}\selectfont
3518 \fi
3519 \let#3#4}}
3520 \let\bbl@langfeatures\@empty
3521 \def\babelFSfeatures{% make sure \fontspec is redefined once
3522 \let\bbl@ori@fontspec\fontspec
3523 \renewcommand\fontspec[1][{}]{%
3524 \bbl@ori@fontspec[\bbl@langfeatures##1]}
3525 \let\babelFSfeatures\bbl@FSfeatures
3526 \babelFSfeatures}
3527 \def\bbl@FSfeatures#1#2{%
3528 \expandafter\addto\csname extras#1\endcsname{%
3529 \babel@save\bbl@langfeatures
3530 \edef\bbl@langfeatures{#2,}}
3531 <</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.

$\text{\LaTeX}$  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  $\text{\LaTeX}$ . Anyway, for consistency `LuaTeX` also resets the catcodes.

```

3532 <<{*Restore Unicode catcodes before loading patterns}>> ≡
3533 \begingroup
3534 % Reset chars "80-"C0 to category "other", no case mapping:
3535 \catcode\@=11 \count@=128
3536 \loop\ifnum\count@<192
3537 \global\uccode\count@=0 \global\lccode\count@=0
3538 \global\catcode\count@=12 \global\sffcode\count@=1000
3539 \advance\count@ by 1 \repeat
3540 % Other:
3541 \def\O ##1 {%
3542 \global\uccode"##1=0 \global\lccode"##1=0
3543 \global\catcode"##1=12 \global\sffcode"##1=1000 }%
3544 % Letter:
3545 \def\L ##1 ##2 ##3 {\global\catcode"##1=11
3546 \global\uccode"##1="##2
3547 \global\lccode"##1="##3
3548 % Uppercase letters have sffcode=999:
3549 \ifnum"##1="##3 \else \global\sffcode"##1=999 \fi }%
3550 % Letter without case mappings:
3551 \def\l ##1 {\L ##1 ##1 ##1 }%
3552 \l 00AA
3553 \L 00B5 039C 00B5
3554 \l 00BA
3555 \O 00D7
3556 \l 00DF
3557 \O 00F7

```

```

3558 \L 00FF 0178 00FF
3559 \endgroup
3560 \input #1\relax
3561 <</Restore Unicode catcodes before loading patterns>>

```

Some more common code.

```

3562 <<(*Footnote changes)>> ≡
3563 \bbl@trace{Bidi footnotes}
3564 \ifx\bbl@beforeforeign\leavevmode
3565 \def\bbl@footnote#1#2#3{%
3566 \ifnextchar[%
3567 {\bbl@footnote@o{#1}{#2}{#3}}%
3568 {\bbl@footnote@x{#1}{#2}{#3}}}
3569 \def\bbl@footnote@x#1#2#3#4{%
3570 \bgroup
3571 \select@language@x{\bbl@main@language}%
3572 \bbl@fn@footnote{#2#1{\ignorespaces#4}#3}%
3573 \egroup}
3574 \def\bbl@footnote@o#1#2#3[#4]#5{%
3575 \bgroup
3576 \select@language@x{\bbl@main@language}%
3577 \bbl@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}%
3578 \egroup}
3579 \def\bbl@footnotetext#1#2#3{%
3580 \ifnextchar[%
3581 {\bbl@footnotetext@o{#1}{#2}{#3}}%
3582 {\bbl@footnotetext@x{#1}{#2}{#3}}}
3583 \def\bbl@footnotetext@x#1#2#3#4{%
3584 \bgroup
3585 \select@language@x{\bbl@main@language}%
3586 \bbl@fn@footnotetext{#2#1{\ignorespaces#4}#3}%
3587 \egroup}
3588 \def\bbl@footnotetext@o#1#2#3[#4]#5{%
3589 \bgroup
3590 \select@language@x{\bbl@main@language}%
3591 \bbl@fn@footnotetext[#4]{#2#1{\ignorespaces#5}#3}%
3592 \egroup}
3593 \def\BabelFootnote#1#2#3#4{%
3594 \ifx\bbl@fn@footnote\@undefined
3595 \let\bbl@fn@footnote\footnote
3596 \fi
3597 \ifx\bbl@fn@footnotetext\@undefined
3598 \let\bbl@fn@footnotetext\footnotetext
3599 \fi
3600 \bbl@ifblank{#2}%
3601 {\def#1{\bbl@footnote{\@firstofone}{#3}{#4}}
3602 \@namedef{\bbl@stripslash#1text}%
3603 {\bbl@footnotetext{\@firstofone}{#3}{#4}}}%
3604 {\def#1{\bbl@exp{\bbl@footnote{\bbl@foreignlanguage{#2}}}{#3}{#4}}%
3605 \@namedef{\bbl@stripslash#1text}%
3606 {\bbl@exp{\bbl@footnotetext{\bbl@foreignlanguage{#2}}}{#3}{#4}}}%
3607 \fi
3608 <</Footnote changes>>

```

Now, the code.

```

3609 (*xetex)
3610 \def\BabelStringsDefault{unicode}
3611 \let\xebbl@stop\relax
3612 \AddBabelHook{xetex}{encodedcommands}{%

```



```

3613 \def\bbl@tempa{#1}%
3614 \ifx\bbl@tempa\@empty
3615 \XeTeXinputencoding"bytes"%
3616 \else
3617 \XeTeXinputencoding"#1"%
3618 \fi
3619 \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
3620 \AddBabelHook{xetex}{stopcommands}{%
3621 \xebbl@stop
3622 \let\xebbl@stop\relax}
3623 \def\bbl@intraspace#1 #2 #3\@{%%
3624 \bbl@csarg\gdef{\xeisp@\bbl@cs{sbcpr@\languagename}}}%
3625 {\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}}
3626 \def\bbl@intrapenalty#1\@{%%
3627 \bbl@csarg\gdef{\xeipn@\bbl@cs{sbcpr@\languagename}}}%
3628 {\XeTeXlinebreakpenalty #1\relax}}
3629 \AddBabelHook{xetex}{loadkernel}{%
3630 <<Restore Unicode catcodes before loading patterns>>}}
3631 \ifx\DisableBabelHook\undefined\endinput\fi
3632 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
3633 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ccheckstdfonts}
3634 \DisableBabelHook{babel-fontspec}
3635 <<Font selection>>
3636 \input txtbabel.def
3637 </xetex>

```

## 14.2 Layout

*In progress.*

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

\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.

```

3638 (*texxet)
3639 \bbl@trace{Redefinitions for bidi layout}
3640 \def\bbl@sspre@caption{%
3641 \bbl@exp{\everyhbox{\bbl@textdir\bbl@cs{wdir@\bbl@main@language}}}}
3642 \ifx\bbl@opt@layout\@nnil\endinput\fi % No layout
3643 \def\bbl@startskip{\ifcase\bbl@thepardir\leftskip\else\rightskip\fi}
3644 \def\bbl@endskip{\ifcase\bbl@thepardir\rightskip\else\leftskip\fi}
3645 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
3646 \def\@hangfrom#1{%
3647 \setbox\@tempboxa\hbox{#1}}%
3648 \hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
3649 \noindent\box\@tempboxa}
3650 \def\raggedright{%
3651 \let\@centercr
3652 \bbl@startskip\z@skip
3653 \@rightskip\@flushglue
3654 \bbl@endskip\@rightskip
3655 \parindent\z@
3656 \parfillskip\bbl@startskip}
3657 \def\raggedleft{%
3658 \let\@centercr

```

```

3659 \bbl@startskip\@flushglue
3660 \bbl@endskip\z@skip
3661 \parindent\z@
3662 \parfillskip\bbl@endskip}
3663 \fi
3664 \IfBabelLayout{lists}
3665 {\bbl@sreplace\list
3666 {\@totalleftmargin\leftmargin}{\@totalleftmargin\bbl@listleftmargin}%
3667 \def\bbl@listleftmargin{%
3668 \ifcase\bbl@thepardir\leftmargin\else\rightmargin\fi}%
3669 \ifcase\bbl@engine
3670 \def\labelenumii{}\theenumii{}\pdfTeX doesn't reverse ()
3671 \def\p@enumiii{\p@enumii}\theenumii{}\fi
3672 \fi
3673 \bbl@sreplace\@verbatim
3674 {\leftskip\@totalleftmargin}%
3675 {\bbl@startskip\textwidth
3676 \advance\bbl@startskip-\linewidth}%
3677 \bbl@sreplace\@verbatim
3678 {\rightskip\z@skip}%
3679 {\bbl@endskip\z@skip}}%
3680 {}
3681 \IfBabelLayout{contents}
3682 {\bbl@sreplace\@dottedtocline{\leftskip}{\bbl@startskip}%
3683 \bbl@sreplace\@dottedtocline{\rightskip}{\bbl@endskip}}
3684 {}
3685 \IfBabelLayout{columns}
3686 {\bbl@sreplace\@outputdblcol{\hb@xt@\textwidth}{\bbl@outputbox}%
3687 \def\bbl@outputbox#1{%
3688 \hb@xt@\textwidth{%
3689 \hskip\columnwidth
3690 \hfil
3691 {\normalcolor\vrule \@width\columnseprule}%
3692 \hfil
3693 \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
3694 \hskip-\textwidth
3695 \hb@xt@\columnwidth{\box\@outputbox \hss}%
3696 \hskip\columnsep
3697 \hskip\columnwidth}}}%
3698 {}
3699 <<Footnote changes>>
3700 \IfBabelLayout{footnotes}%
3701 {\BabelFootnote\footnote\language\{}}%
3702 \BabelFootnote\localfootnote\language\{}}%
3703 \BabelFootnote\mainfootnote\{}}%
3704 {}

```

Implicitly reverses sectioning labels in bidi=basic, because the full stop is not in contact with L numbers any more. I think there must be a better way.

```

3705 \IfBabelLayout{counters}%
3706 {\let\bbl@latinarabic=\@arabic
3707 \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
3708 \let\bbl@asciroman=\@roman
3709 \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
3710 \let\bbl@asciiRoman=\@Roman
3711 \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}}%
3712 </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.

```
3713 (*luatex)
3714 \ifx\AddBabelHook\undefined
3715 \bbl@trace{Read language.dat}
3716 \begingroup
3717 \toks@{}
3718 \count@ \z@ % 0=start, 1=0th, 2=normal
3719 \def\bbl@process@line#1#2 #3 #4 {%
3720   \ifx=#1%
3721     \bbl@process@synonym{#2}%
3722   \else
3723     \bbl@process@language{#1#2}{#3}{#4}%
3724   \fi
3725   \ignorespaces}
3726 \def\bbl@manylang{%
3727   \ifnum\bbl@last>\@ne
3728     \bbl@info{Non-standard hyphenation setup}%
3729   \fi
3730   \let\bbl@manylang\relax}
3731 \def\bbl@process@language#1#2#3{%
3732   \ifcase\count@
3733     \@ifundefined{zth#1}{\count@\tw@}{\count@\@ne}%
3734   \or
3735     \count@\tw@
3736   \fi
```

```

3737 \ifnum\count@=\tw@
3738 \expandafter\addlanguage\csname l@#1\endcsname
3739 \language\allocationnumber
3740 \chardef\bbl@last\allocationnumber
3741 \bbl@manylang
3742 \let\bbl@elt\relax
3743 \xdef\bbl@languages{%
3744 \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
3745 \fi
3746 \the\toks@
3747 \toks@{}}
3748 \def\bbl@process@synonym@aux#1#2{%
3749 \global\expandafter\chardef\csname l@#1\endcsname#2\relax
3750 \let\bbl@elt\relax
3751 \xdef\bbl@languages{%
3752 \bbl@languages\bbl@elt{#1}{#2}{}}}%
3753 \def\bbl@process@synonym#1{%
3754 \ifcase\count@
3755 \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
3756 \or
3757 \@ifundefined{zth@#1}{\bbl@process@synonym@aux{#1}{0}}}%
3758 \else
3759 \bbl@process@synonym@aux{#1}{\the\bbl@last}%
3760 \fi}
3761 \ifx\bbl@languages\@undefined % Just a (sensible?) guess
3762 \chardef\l@english\z@
3763 \chardef\l@USenglish\z@
3764 \chardef\bbl@last\z@
3765 \global\@namedef\bbl@hyphendata@0{{hyphen.tex}}
3766 \gdef\bbl@languages{%
3767 \bbl@elt{english}{0}{hyphen.tex}}%
3768 \bbl@elt{USenglish}{0}{}}
3769 \else
3770 \global\let\bbl@languages@format\bbl@languages
3771 \def\bbl@elt#1#2#3#4{% Remove all except language 0
3772 \ifnum#2>\z@\else
3773 \noexpand\bbl@elt{#1}{#2}{#3}{#4}%
3774 \fi}%
3775 \xdef\bbl@languages{\bbl@languages}%
3776 \fi
3777 \def\bbl@elt#1#2#3#4{\@namedef{zth@#1}} % Define flags
3778 \bbl@languages
3779 \openin1=language.dat
3780 \ifeof1
3781 \bbl@warning{I couldn't find language.dat. No additional\\%
3782 patterns loaded. Reported}%
3783 \else
3784 \loop
3785 \endlinechar\m@ne
3786 \read1 to \bbl@line
3787 \endlinechar\^^M
3788 \if T\ifeof1F\fi T\relax
3789 \ifx\bbl@line\empty\else
3790 \edef\bbl@line{\bbl@line\space\space\space}%
3791 \expandafter\bbl@process@line\bbl@line\relax
3792 \fi
3793 \repeat
3794 \fi
3795 \endgroup

```

```

3796 \bbl@trace{Macros for reading patterns files}
3797 \def\bbl@get@enc#1:#2:#3\@@{\def\bbl@hyph@enc{#2}}
3798 \ifx\babelcatcodetablenum\undefined
3799 \def\babelcatcodetablenum{5211}
3800 \fi
3801 \def\bbl@luapatterns#1#2{%
3802 \bbl@get@enc#1::\@@@
3803 \setbox\z@\hbox\bgroup
3804 \begingroup
3805 \ifx\catcodetable\undefined
3806 \let\savecatcodetable\luatexsavecatcodetable
3807 \let\initcatcodetable\luatexinitcatcodetable
3808 \let\catcodetable\luatexcatcodetable
3809 \fi
3810 \savecatcodetable\babelcatcodetablenum\relax
3811 \initcatcodetable\numexpr\babelcatcodetablenum+1\relax
3812 \catcodetable\numexpr\babelcatcodetablenum+1\relax
3813 \catcode`\# =6 \catcode`\$ =3 \catcode`\& =4 \catcode`\^ =7
3814 \catcode`\_ =8 \catcode`\{ =1 \catcode`\} =2 \catcode`\~ =13
3815 \catcode`\@ =11 \catcode`\^^I =10 \catcode`\^^J =12
3816 \catcode`\< =12 \catcode`\> =12 \catcode`\* =12 \catcode`\.=12
3817 \catcode`\- =12 \catcode`\/=12 \catcode`\[ =12 \catcode`\] =12
3818 \catcode`\` =12 \catcode`\' =12 \catcode`\" =12
3819 \input #1\relax
3820 \catcodetable\babelcatcodetablenum\relax
3821 \endgroup
3822 \def\bbl@tempa{#2}%
3823 \ifx\bbl@tempa\@empty\else
3824 \input #2\relax
3825 \fi
3826 \egroup}%
3827 \def\bbl@patterns@lua#1{%
3828 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
3829 \csname l@#1\endcsname
3830 \edef\bbl@tempa{#1}%
3831 \else
3832 \csname l@#1:\f@encoding\endcsname
3833 \edef\bbl@tempa{#1:\f@encoding}%
3834 \fi\relax
3835 \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
3836 \@ifundefined{bbl@hyphendata@the\language}%
3837 {\def\bbl@elt##1##2##3##4{%
3838 \ifnum##2=\csname l@bbl@tempa\endcsname % #2=spanish, dutch:OT1...
3839 \def\bbl@tempb{##3}%
3840 \ifx\bbl@tempb\@empty\else % if not a synonymous
3841 \def\bbl@tempc{##3}{##4}%
3842 \fi
3843 \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
3844 \fi}%
3845 \bbl@languages
3846 \@ifundefined{bbl@hyphendata@the\language}%
3847 {\bbl@info{No hyphenation patterns were set for\%
3848 language '\bbl@tempa'. Reported}}%
3849 {\expandafter\expandafter\expandafter\bbl@luapatterns
3850 \csname bbl@hyphendata@the\language\endcsname}}}%
3851 \endinput\fi
3852 \begingroup
3853 \catcode`\% =12
3854 \catcode`\' =12

```

```

3855 \catcode`\:=12
3856 \catcode`\:=12
3857 \directlua{
3858   Babel = Babel or {}
3859   function Babel.bytes(line)
3860     return line:gsub(".",
3861       function (chr) return unicode.utf8.char(string.byte(chr)) end)
3862   end
3863   function Babel.begin_process_input()
3864     if luatexbase and luatexbase.add_to_callback then
3865       luatexbase.add_to_callback('process_input_buffer',
3866         Babel.bytes, 'Babel.bytes')
3867     else
3868       Babel.callback = callback.find('process_input_buffer')
3869       callback.register('process_input_buffer', Babel.bytes)
3870     end
3871   end
3872   function Babel.end_process_input ()
3873     if luatexbase and luatexbase.remove_from_callback then
3874       luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
3875     else
3876       callback.register('process_input_buffer', Babel.callback)
3877     end
3878   end
3879   function Babel.addpatterns(pp, lg)
3880     local lg = lang.new(lg)
3881     local pats = lang.patterns(lg) or ''
3882     lang.clear_patterns(lg)
3883     for p in pp:gmatch('[^%s]+') do
3884       ss = ''
3885       for i in string.utfcharacters(p:gsub('%d', '')) do
3886         ss = ss .. '%d?' .. i
3887       end
3888       ss = ss:gsub('^%%d%?%', '%%.') .. '%d?'
3889       ss = ss:gsub('%.%d%?$', '%%.')
3890       pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
3891       if n == 0 then
3892         tex.sprint(
3893           [[\string\csname\space bbl@info\endcsname{New pattern: }]]
3894           .. p .. [[{}]])
3895       pats = pats .. ' ' .. p
3896     else
3897       tex.sprint(
3898         [[\string\csname\space bbl@info\endcsname{Renew pattern: }]]
3899         .. p .. [[{}]])
3900     end
3901   end
3902   lang.patterns(lg, pats)
3903 end
3904 }
3905 \endgroup
3906 \ifx\newattribute\@undefined\else
3907   \newattribute\bbl@attr@locale
3908   \AddBabelHook{luatex}{beforeextras}{%
3909     \setattribute\bbl@attr@locale\localeid}
3910 \fi
3911 \def\BabelStringsDefault{unicode}
3912 \let\luabbbl@stop\relax
3913 \AddBabelHook{luatex}{encodedcommands}{%

```

```

3914 \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
3915 \ifx\bbl@tempa\bbl@tempb\else
3916   \directlua{Babel.begin_process_input()}%
3917   \def\luabbl@stop{%
3918     \directlua{Babel.end_process_input()}}%
3919   \fi}%
3920 \AddBabelHook{luatex}{stopcommands}{%
3921   \luabbl@stop
3922   \let\luabbl@stop\relax}
3923 \AddBabelHook{luatex}{patterns}{%
3924   \@ifundefined{bbl@hyphendata@the\language}%
3925     {\def\bbl@elt##1##2##3##4{%
3926       \ifnum##2=\csname l@##2\endcsname % #2=spanish, dutch:OT1...
3927       \def\bbl@tempb{##3}%
3928       \ifx\bbl@tempb\@empty\else % if not a synonymous
3929         \def\bbl@tempc{##3}{##4}}%
3930       \fi
3931       \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
3932     \fi}%
3933   \bbl@languages
3934   \@ifundefined{bbl@hyphendata@the\language}%
3935     {\bbl@info{No hyphenation patterns were set for\%
3936       language '#2'. Reported}}%
3937     {\expandafter\expandafter\expandafter\bbl@luapatterns
3938       \csname bbl@hyphendata@the\language\endcsname}}}%
3939   \@ifundefined{bbl@patterns@}{}%
3940   \begingroup
3941     \bbl@xin@{,\number\language,}{,\bbl@pttnlist}%
3942     \ifin@ \else
3943       \ifx\bbl@patterns@\@empty\else
3944         \directlua{ Babel.addpatterns(
3945           [[\bbl@patterns@]], \number\language) }%
3946         \fi
3947         \@ifundefined{bbl@patterns@#1}%
3948         \@empty
3949         {\directlua{ Babel.addpatterns(
3950           [[\space\csname bbl@patterns@#1\endcsname]],
3951           \number\language) }}%
3952         \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
3953       \fi
3954   \endgroup}}
3955 \AddBabelHook{luatex}{everylanguage}{%
3956   \def\process@language##1##2##3{%
3957     \def\process@line####1####2 ####3 ####4 {}}
3958 \AddBabelHook{luatex}{loadpatterns}{%
3959   \input #1\relax
3960   \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
3961     {{#1}}}%
3962 \AddBabelHook{luatex}{loadexceptions}{%
3963   \input #1\relax
3964   \def\bbl@tempb##1##2{{##1}{##1}}%
3965   \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
3966     {\expandafter\expandafter\expandafter\bbl@tempb
3967       \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.

```

3968 \@onlypreamble\babelpatterns
3969 \AtEndOfPackage{%
3970   \newcommand\babelpatterns[2][\@empty]{%
3971     \ifx\bbbl@patterns\relax
3972       \let\bbbl@patterns@\@empty
3973     \fi
3974     \ifx\bbbl@pttnlist\@empty\else
3975       \bbbl@warning{%
3976         You must not intermingle \string\selectlanguage\space and\%
3977         \string\babelpatterns\space or some patterns will not\%
3978         be taken into account. Reported}%
3979     \fi
3980     \ifx\@empty#1%
3981       \protected@edef\bbbl@patterns@\bbbl@patterns@\space#2}%
3982     \else
3983       \edef\bbbl@tempb{\zap@space#1 \@empty}%
3984       \bbbl@for\bbbl@tempa\bbbl@tempb{%
3985         \bbbl@fixname\bbbl@tempa
3986         \bbbl@iflanguage\bbbl@tempa{%
3987           \bbbl@csarg\protected@edef{patterns@\bbbl@tempa}{%
3988             \@ifundefined{bbbl@patterns@\bbbl@tempa}%
3989             \@empty
3990             {\csname bbbl@patterns@\bbbl@tempa\endcsname\space}%
3991             #2}}}%
3992     \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).

```

3993 \def\bbbl@intraspace#1 #2 #3\@@{%
3994   \directlua{
3995     Babel = Babel or {}
3996     Babel.intraspaces = Babel.intraspaces or {}
3997     Babel.intraspaces['\csname bbbl@sbcpr@\language\endcsname'] = %
3998       {b = #1, p = #2, m = #3}
3999     Babel.locale_props[\the\localeid].intraspace = %
4000       {b = #1, p = #2, m = #3}
4001   }}
4002 \def\bbbl@intrapenalty#1\@@{%
4003   \directlua{
4004     Babel = Babel or {}
4005     Babel.intrapenalties = Babel.intrapenalties or {}
4006     Babel.intrapenalties['\csname bbbl@sbcpr@\language\endcsname'] = #1
4007     Babel.locale_props[\the\localeid].intrapenalty = #1
4008   }}
4009 \begingroup
4010 \catcode`\%=12
4011 \catcode`\^=14
4012 \catcode`\'=12
4013 \catcode`\~=12
4014 \gdef\bbbl@seaintraspace{^
4015   \let\bbbl@seaintraspace\relax
4016   \directlua{
4017     Babel = Babel or {}
4018     Babel.sea_enabled = true

```



```

4019 Babel.sea_ranges = Babel.sea_ranges or {}
4020 function Babel.set_chranges (script, chrng)
4021     local c = 0
4022     for s, e in string.gmatch(chrng..' ', '(.-%.(-)%s') do
4023         Babel.sea_ranges[script..c]={tonumber(s,16), tonumber(e,16)}
4024         c = c + 1
4025     end
4026 end
4027 function Babel.sea_disc_to_space (head)
4028     local sea_ranges = Babel.sea_ranges
4029     local last_char = nil
4030     local quad = 655360      ^^ 10 pt = 655360 = 10 * 65536
4031     for item in node.traverse(head) do
4032         local i = item.id
4033         if i == node.id'glyph' then
4034             last_char = item
4035         elseif i == 7 and item.subtype == 3 and last_char
4036             and last_char.char > 0x0C99 then
4037             quad = font.getfont(last_char.font).size
4038             for lg, rg in pairs(sea_ranges) do
4039                 if last_char.char > rg[1] and last_char.char < rg[2] then
4040                     lg = lg:sub(1, 4)
4041                     local intraspace = Babel.intraspaces[lg]
4042                     local intrapenalty = Babel.intrapenalties[lg]
4043                     local n
4044                     if intrapenalty ~= 0 then
4045                         n = node.new(14, 0)      ^^ penalty
4046                         n.penalty = intrapenalty
4047                         node.insert_before(head, item, n)
4048                     end
4049                     n = node.new(12, 13)      ^^ (glue, spaceskip)
4050                     node.setglue(n, intraspace.b * quad,
4051                                 intraspace.p * quad,
4052                                 intraspace.m * quad)
4053                     node.insert_before(head, item, n)
4054                     node.remove(head, item)
4055                 end
4056             end
4057         end
4058     end
4059 end
4060 }^^
4061 \bbl@luahyphenate}
4062 \catcode`\%=14
4063 \gdef\bbl@cjkintraspaces{%
4064 \let\bbl@cjkintraspaces\relax
4065 \directlua{
4066     Babel = Babel or {}
4067     require'babel-data-cjk.lua'
4068     Babel.cjk_enabled = true
4069     function Babel.cjk_linebreak(head)
4070         local GLYPH = node.id'glyph'
4071         local last_char = nil
4072         local quad = 655360      % 10 pt = 655360 = 10 * 65536
4073         local last_class = nil
4074         local last_lang = nil
4075
4076         for item in node.traverse(head) do
4077             if item.id == GLYPH then

```

```

4078
4079     local lang = item.lang
4080
4081     local LOCALE = node.get_attribute(item,
4082         luatexbase.registernumber'bbl@attr@locale')
4083     local props = Babel.locale_props[LOCALE]
4084
4085     class = Babel.cjk_class[item.char].c
4086
4087     if class == 'cp' then class = 'cl' end % ]) as CL
4088     if class == 'id' then class = 'I' end
4089
4090     if class and last_class and Babel.cjk_breaks[last_class][class] then
4091         br = Babel.cjk_breaks[last_class][class]
4092     else
4093         br = 0
4094     end
4095
4096     if br == 1 and props.linebreak == 'c' and
4097         lang ~= \the\l@nohyphenation\space and
4098         last_lang ~= \the\l@nohyphenation then
4099         local intrapenalty = props.intrapenalty
4100         if intrapenalty ~= 0 then
4101             local n = node.new(14, 0) % penalty
4102             n.penalty = intrapenalty
4103             node.insert_before(head, item, n)
4104         end
4105         local intraspace = props.intraspace
4106         local n = node.new(12, 13) % (glue, spaceskip)
4107         node.setglue(n, intraspace.b * quad,
4108             intraspace.p * quad,
4109             intraspace.m * quad)
4110         node.insert_before(head, item, n)
4111     end
4112
4113     quad = font.getfont(item.font).size
4114     last_class = class
4115     last_lang = lang
4116     else % if penalty, glue or anything else
4117         last_class = nil
4118     end
4119 end
4120 lang.hyphenate(head)
4121 end
4122 }%
4123 \bbl@luahyphenate}
4124 \gdef\bbl@luahyphenate{%
4125 \let\bbl@luahyphenate\relax
4126 \directlua{
4127     luatexbase.add_to_callback('hyphenate',
4128         function (head, tail)
4129             if Babel.cjk_enabled then
4130                 Babel.cjk_linebreak(head)
4131             end
4132             lang.hyphenate(head)
4133             if Babel.sea_enabled then
4134                 Babel.sea_disc_to_space(head)
4135             end
4136         end,

```

```

4137 'Babel.hyphenate')
4138 }
4139 }
4140 \endgroup

```

## 14.5 CJK line breaking

Minimal line breaking for CJK scripts, mainly intended for simple documents and short texts as a secondary language. Only line breaking, with a little stretching for justification, without any attempt to adjust the spacing. It is based on (but does not strictly follow) the Unicode algorithm.

We first need a little table with the corresponding line breaking properties. A few characters have an additional key for the width (fullwidth vs. halfwidth), not yet used. There is a separate file, defined below.

*Work in progress.*

Common stuff.

```

4141 \AddBabelHook{luatex}{loadkernel}{%
4142 <<Restore Unicode catcodes before loading patterns>>}
4143 \ifx\DisableBabelHook\undefined\endinput\fi
4144 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
4145 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@cckstdfonts}
4146 \DisableBabelHook{babel-fontspec}
4147 <<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.

```

4148 \def\bbl@luafixboxdir{%
4149   \setbox\z@\hbox{\textdir TLT}%
4150   \directlua{
4151     function Babel.first_dir(head)
4152       for item in node.traverse_id(node.id'dir', head) do
4153         return item
4154       end
4155       return nil
4156     end
4157     if Babel.first_dir(tex.box[0].head) then
4158       function Babel.fixboxdirs(head)
4159         local fd = Babel.first_dir(head)
4160         if fd and fd.dir:sub(1,1) == '-' then
4161           head = node.remove(head, fd)
4162         end
4163         return head
4164       end
4165     end
4166   }}
4167 \AtBeginDocument{\bbl@luafixboxdir}

```

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

```

4168 \newcommand\babelcharproperty[1]{%
4169   \count@=#1\relax
4170   \ifvmode
4171     \expandafter\bbl@chprop
4172   \else
4173     \bbl@error{\string\babelcharproperty\space can be used only in\\%
4174               vertical mode (preamble or between paragraphs)}%
4175     {See the manual for futher info}%

```

```

4176 \fi}
4177 \newcommand\bbbl@chprop[3][\the\count@]{%
4178 \@tempcnta=#1\relax
4179 \bbbl@ifunset{\bbbl@chprop@#2}%
4180 {\bbbl@error{No property named '#2'. Allowed values are\%
4181 direction (bc), mirror (bmg), and linebreak (lb)}%
4182 {See the manual for futher info}}%
4183 }%
4184 \loop
4185 \@nameuse{\bbbl@chprop@#2}{#3}%
4186 \ifnum\count@<\@tempcnta
4187 \advance\count@\@ne
4188 \repeat}
4189 \def\bbbl@chprop@direction#1{%
4190 \directlua{
4191 Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
4192 Babel.characters[\the\count@]['d'] = '#1'
4193 }}
4194 \let\bbbl@chprop@bc\bbbl@chprop@direction
4195 \def\bbbl@chprop@mirror#1{%
4196 \directlua{
4197 Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
4198 Babel.characters[\the\count@]['m'] = '\number#1'
4199 }}
4200 \let\bbbl@chprop@bmg\bbbl@chprop@mirror
4201 \def\bbbl@chprop@linebreak#1{%
4202 \directlua{
4203 Babel.Babel.cjk_characters[\the\count@] = Babel.Babel.cjk_characters[\the\count@] or {}
4204 Babel.Babel.cjk_characters[\the\count@]['c'] = '#1'
4205 }}
4206 \let\bbbl@chprop@lb\bbbl@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.) with `bidibasic`, without having to patch almost any macro where text direction is relevant.

`\@hangfrom` is useful in many contexts and it is redefined always with the layout option. There are, however, a number of issues when the text direction is not the same as the box direction (as set by `\bodydir`), and when `\parbox` and `\hangindent` are involved.

Fortunately, latest releases of luatex simplify a lot the solution with `\shapemode`.

With the issue #15 I realized commands are best patched, instead of redefined. With a few lines, a modification could be applied to several classes and packages. Now, `tabular` seems to work (at least in simple cases) with `array`, `tabularx`, `hhline`, `colortbl`, `longtable`, `booktabs`, etc. However, `dcolumn` still fails.

```

4207 \bbbl@trace{Redefinitions for bidi layout}
4208 \ifx\@eqnnum\undefined\else
4209 \ifx\bbbl@attr@dir\undefined\else
4210 \edef\@eqnnum{%
4211 \unexpanded{\ifcase\bbbl@attr@dir\else\bbbl@textdir\@ne\fi}%
4212 \unexpanded\expandafter{\@eqnnum}}
4213 \fi
4214 \fi
4215 \ifx\bbbl@opt@layout\@nnil\endinput\fi % if no layout
4216 \ifx\bbbl@beforeforeign\leavevmode % A poor test for bidi=

```

```

4217 \def\bb1@nextfake#1{% non-local changes, use always inside a group!
4218   \bb1@exp{%
4219     \mathdir\the\bodydir
4220     #1%           Once entered in math, set boxes to restore values
4221     \<ifmode>%
4222       \everyvbox{%
4223         \the\everyvbox
4224         \bodydir\the\bodydir
4225         \mathdir\the\mathdir
4226         \everyhbox{\the\everyhbox}%
4227         \everyvbox{\the\everyvbox}}%
4228       \everyhbox{%
4229         \the\everyhbox
4230         \bodydir\the\bodydir
4231         \mathdir\the\mathdir
4232         \everyhbox{\the\everyhbox}%
4233         \everyvbox{\the\everyvbox}}%
4234     \<fi>}}%
4235 \def\@hangfrom#1{%
4236   \setbox\@tempboxa\hbox{{#1}}%
4237   \hangindent\wd\@tempboxa
4238   \ifnum\bb1@getluadir{page}=\bb1@getluadir{par}\else
4239     \shapemode\@ne
4240   \fi
4241   \noindent\box\@tempboxa}
4242 \fi
4243 \IfBabelLayout{tabular}
4244   {\let\bb1@OL@tabular\@tabular
4245     \bb1@replace\@tabular{$}{\bb1@nextfake$}%
4246     \let\bb1@tabular\@tabular
4247     \AtBeginDocument{%
4248       \ifx\bb1@tabular\@tabular\else
4249         \bb1@replace\@tabular{$}{\bb1@nextfake$}%
4250       \fi}}
4251   {}
4252 \IfBabelLayout{lists}
4253   {\let\bb1@OL@list\list
4254     \bb1@sreplace\list{\parshape}{\bb1@listparshape}%
4255     \def\bb1@listparshape#1#2#3{%
4256       \parshape #1 #2 #3 %
4257       \ifnum\bb1@getluadir{page}=\bb1@getluadir{par}\else
4258         \shapemode\tw@
4259       \fi}}
4260   {}
4261 \IfBabelLayout{graphics}
4262   {\let\bb1@pictresetdir\relax
4263     \def\bb1@pictsetdir{%
4264       \ifcase\bb1@thetextdir
4265       \let\bb1@pictresetdir\relax
4266       \else
4267         \textdir TLT\relax
4268         \def\bb1@pictresetdir{\textdir TRT\relax}%
4269       \fi}%
4270     \let\bb1@OL@picture\@picture
4271     \let\bb1@OL@put\put
4272     \bb1@sreplace\@picture{\hskip-}{\bb1@pictsetdir\hskip-}%
4273     \def\put(#1,#2)#3{% Not easy to patch. Better redefine.
4274       \@killglue
4275       \raise#2\unitlength

```



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.

```
4319 (*basic-r)
4320 Babel = Babel or {}
4321
4322 Babel.bidi_enabled = true
4323
4324 require('babel-data-bidi.lua')
4325
4326 local characters = Babel.characters
4327 local ranges = Babel.ranges
4328
4329 local DIR = node.id("dir")
4330
4331 local function dir_mark(head, from, to, outer)
4332   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
4333   local d = node.new(DIR)
4334   d.dir = '+' .. dir
4335   node.insert_before(head, from, d)
4336   d = node.new(DIR)
4337   d.dir = '-' .. dir
4338   node.insert_after(head, to, d)
4339 end
4340
4341 function Babel.bidi(head, ispar)
4342   local first_n, last_n      -- first and last char with nums
4343   local last_es              -- an auxiliary 'last' used with nums
4344   local first_d, last_d      -- first and last char in L/R block
4345   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):

```
4346 local strong = ('TRT' == tex.pardir) and 'r' or 'l'
4347 local strong_lr = (strong == 'l') and 'l' or 'r'
```

```

4348 local outer = strong
4349
4350 local new_dir = false
4351 local first_dir = false
4352 local inmath = false
4353
4354 local last_lr
4355
4356 local type_n = ''
4357
4358 for item in node.traverse(head) do
4359
4360   -- three cases: glyph, dir, otherwise
4361   if item.id == node.id'glyph'
4362     or (item.id == 7 and item.subtype == 2) then
4363
4364     local itemchar
4365     if item.id == 7 and item.subtype == 2 then
4366       itemchar = item.replace.char
4367     else
4368       itemchar = item.char
4369     end
4370     local chardata = characters[itemchar]
4371     dir = chardata and chardata.d or nil
4372     if not dir then
4373       for nn, et in ipairs(ranges) do
4374         if itemchar < et[1] then
4375           break
4376         elseif itemchar <= et[2] then
4377           dir = et[3]
4378           break
4379         end
4380       end
4381     end
4382     dir = dir or 'l'
4383     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).

```

4384   if new_dir then
4385     attr_dir = 0
4386     for at in node.traverse(item.attr) do
4387       if at.number == luatexbase.registernumber'bbl@attr@dir' then
4388         attr_dir = at.value % 3
4389       end
4390     end
4391     if attr_dir == 1 then
4392       strong = 'r'
4393     elseif attr_dir == 2 then
4394       strong = 'al'
4395     else
4396       strong = 'l'
4397     end
4398     strong_lr = (strong == 'l') and 'l' or 'r'
4399     outer = strong_lr
4400     new_dir = false

```



```

4401     end
4402
4403     if dir == 'nsm' then dir = strong end          -- W1

```

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

```

4404     dir_real = dir          -- We need dir_real to set strong below
4405     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:

```

4406     if strong == 'al' then
4407         if dir == 'en' then dir = 'an' end          -- W2
4408         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
4409         strong_lr = 'r'                             -- W3
4410     end

```

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

```

4411     elseif item.id == node.id'dir' and not inmath then
4412         new_dir = true
4413         dir = nil
4414     elseif item.id == node.id'math' then
4415         inmath = (item.subtype == 0)
4416     else
4417         dir = nil          -- Not a char
4418     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>.

```

4419     if dir == 'en' or dir == 'an' or dir == 'et' then
4420         if dir ~= 'et' then
4421             type_n = dir
4422         end
4423         first_n = first_n or item
4424         last_n = last_es or item
4425         last_es = nil
4426     elseif dir == 'es' and last_n then -- W3+W6
4427         last_es = item
4428     elseif dir == 'cs' then          -- it's right - do nothing
4429     elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
4430         if strong_lr == 'r' and type_n ~= '' then
4431             dir_mark(head, first_n, last_n, 'r')
4432         elseif strong_lr == 'l' and first_d and type_n == 'an' then
4433             dir_mark(head, first_n, last_n, 'r')
4434             dir_mark(head, first_d, last_d, outer)
4435             first_d, last_d = nil, nil
4436         elseif strong_lr == 'l' and type_n ~= '' then
4437             last_d = last_n
4438         end
4439         type_n = ''
4440         first_n, last_n = nil, nil
4441     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:

```
4442   if dir == 'l' or dir == 'r' then
4443       if dir ~= outer then
4444           first_d = first_d or item
4445           last_d = item
4446       elseif first_d and dir ~= strong_lr then
4447           dir_mark(head, first_d, last_d, outer)
4448           first_d, last_d = nil, nil
4449       end
4450   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.

```
4451   if dir and not last_lr and dir ~= 'l' and outer == 'r' then
4452       item.char = characters[item.char] and
4453           characters[item.char].m or item.char
4454   elseif (dir or new_dir) and last_lr ~= item then
4455       local mir = outer .. strong_lr .. (dir or outer)
4456       if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
4457           for ch in node.traverse(node.next(last_lr)) do
4458               if ch == item then break end
4459               if ch.id == node.id'glyph' and characters[ch.char] then
4460                   ch.char = characters[ch.char].m or ch.char
4461               end
4462           end
4463       end
4464   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).

```
4465   if dir == 'l' or dir == 'r' then
4466       last_lr = item
4467       strong = dir_real          -- Don't search back - best save now
4468       strong_lr = (strong == 'l') and 'l' or 'r'
4469   elseif new_dir then
4470       last_lr = nil
4471   end
4472   end
```

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

```
4473   if last_lr and outer == 'r' then
4474       for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
4475           if characters[ch.char] then
4476               ch.char = characters[ch.char].m or ch.char
4477           end
4478       end
4479   end
4480   if first_n then
4481       dir_mark(head, first_n, last_n, outer)
4482   end
4483   if first_d then
4484       dir_mark(head, first_d, last_d, outer)
4485   end
```

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

```
4486 return node.prev(head) or head
4487 end
4488 (/basic-r)
```

And here the Lua code for bidi=basic:

```
4489 (*basic)
4490 Babel = Babel or {}
4491
4492 -- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
4493
4494 Babel.fontmap = Babel.fontmap or {}
4495 Babel.fontmap[0] = {}      -- l
4496 Babel.fontmap[1] = {}      -- r
4497 Babel.fontmap[2] = {}      -- al/an
4498
4499 Babel.bidi_enabled = true
4500 Babel.mirroring_enabled = true
4501
4502 -- Temporary:
4503
4504 if harf then
4505   Babel.mirroring_enabled = false
4506 end
4507
4508 require('babel-data-bidi.lua')
4509
4510 local characters = Babel.characters
4511 local ranges = Babel.ranges
4512
4513 local DIR = node.id('dir')
4514 local GLYPH = node.id('glyph')
4515
4516 local function insert_implicit(head, state, outer)
4517   local new_state = state
4518   if state.sim and state.eim and state.sim ~= state.eim then
4519     dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
4520     local d = node.new(DIR)
4521     d.dir = '+' .. dir
4522     node.insert_before(head, state.sim, d)
4523     local d = node.new(DIR)
4524     d.dir = '-' .. dir
4525     node.insert_after(head, state.eim, d)
4526   end
4527   new_state.sim, new_state.eim = nil, nil
4528   return head, new_state
4529 end
4530
4531 local function insert_numeric(head, state)
4532   local new
4533   local new_state = state
4534   if state.san and state.ean and state.san ~= state.ean then
4535     local d = node.new(DIR)
4536     d.dir = '+TLT'
4537     _, new = node.insert_before(head, state.san, d)
4538     if state.san == state.sim then state.sim = new end
4539     local d = node.new(DIR)
```

```

4540     d.dir = '-TLT'
4541     _, new = node.insert_after(head, state.ean, d)
4542     if state.ean == state.eim then state.eim = new end
4543 end
4544 new_state.san, new_state.ean = nil, nil
4545 return head, new_state
4546 end
4547
4548 -- TODO - \hbox with an explicit dir can lead to wrong results
4549 -- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>. A small attempt
4550 -- was s made to improve the situation, but the problem is the 3-dir
4551 -- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
4552 -- well.
4553
4554 function Babel.bidi(head, ispar, hdir)
4555     local d -- d is used mainly for computations in a loop
4556     local prev_d = ''
4557     local new_d = false
4558
4559     local nodes = {}
4560     local outer_first = nil
4561     local inmath = false
4562
4563     local glue_d = nil
4564     local glue_i = nil
4565
4566     local has_en = false
4567     local first_et = nil
4568
4569     local ATDIR = luatexbase.registernumber'bbl@attr@dir'
4570
4571     local save_outer
4572     local temp = node.get_attribute(head, ATDIR)
4573     if temp then
4574         temp = temp % 3
4575         save_outer = (temp == 0 and 'l') or
4576                     (temp == 1 and 'r') or
4577                     (temp == 2 and 'al')
4578     elseif ispar then -- Or error? Shouldn't happen
4579         save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
4580     else -- Or error? Shouldn't happen
4581         save_outer = ('TRT' == hdir) and 'r' or 'l'
4582     end
4583     -- when the callback is called, we are just _after_ the box,
4584     -- and the textdir is that of the surrounding text
4585     -- if not ispar and hdir ~= tex.textdir then
4586     --     save_outer = ('TRT' == hdir) and 'r' or 'l'
4587     -- end
4588     local outer = save_outer
4589     local last = outer
4590     -- 'al' is only taken into account in the first, current loop
4591     if save_outer == 'al' then save_outer = 'r' end
4592
4593     local fontmap = Babel.fontmap
4594
4595     for item in node.traverse(head) do
4596
4597         -- In what follows, #node is the last (previous) node, because the
4598         -- current one is not added until we start processing the neutrals.

```

```

4599
4600 -- three cases: glyph, dir, otherwise
4601 if item.id == GLYPH
4602     or (item.id == 7 and item.subtype == 2) then
4603
4604     local d_font = nil
4605     local item_r
4606     if item.id == 7 and item.subtype == 2 then
4607         item_r = item.replace -- automatic discs have just 1 glyph
4608     else
4609         item_r = item
4610     end
4611     local chardata = characters[item_r.char]
4612     d = chardata and chardata.d or nil
4613     if not d or d == 'nsm' then
4614         for nn, et in ipairs(ranges) do
4615             if item_r.char < et[1] then
4616                 break
4617             elseif item_r.char <= et[2] then
4618                 if not d then d = et[3]
4619                 elseif d == 'nsm' then d_font = et[3]
4620             end
4621             break
4622         end
4623     end
4624     end
4625     d = d or 'l'
4626
4627     -- A short 'pause' in bidi for mapfont
4628     d_font = d_font or d
4629     d_font = (d_font == 'l' and 0) or
4630             (d_font == 'nsm' and 0) or
4631             (d_font == 'r' and 1) or
4632             (d_font == 'al' and 2) or
4633             (d_font == 'an' and 2) or nil
4634     if d_font and fontmap and fontmap[d_font][item_r.font] then
4635         item_r.font = fontmap[d_font][item_r.font]
4636     end
4637
4638     if new_d then
4639         table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
4640         if inmath then
4641             attr_d = 0
4642         else
4643             attr_d = node.get_attribute(item, ATDIR)
4644             attr_d = attr_d % 3
4645         end
4646         if attr_d == 1 then
4647             outer_first = 'r'
4648             last = 'r'
4649         elseif attr_d == 2 then
4650             outer_first = 'r'
4651             last = 'al'
4652         else
4653             outer_first = 'l'
4654             last = 'l'
4655         end
4656         outer = last
4657         has_en = false

```

```

4658         first_et = nil
4659         new_d = false
4660     end
4661
4662     if glue_d then
4663         if (d == 'l' and 'l' or 'r') ~= glue_d then
4664             table.insert(nodes, {glue_i, 'on', nil})
4665         end
4666         glue_d = nil
4667         glue_i = nil
4668     end
4669
4670     elseif item.id == DIR then
4671         d = nil
4672         new_d = true
4673
4674     elseif item.id == node.id'glue' and item.subtype == 13 then
4675         glue_d = d
4676         glue_i = item
4677         d = nil
4678
4679     elseif item.id == node.id'math' then
4680         inmath = (item.subtype == 0)
4681
4682     else
4683         d = nil
4684     end
4685
4686     -- AL <= EN/ET/ES      -- W2 + W3 + W6
4687     if last == 'al' and d == 'en' then
4688         d = 'an'          -- W3
4689     elseif last == 'al' and (d == 'et' or d == 'es') then
4690         d = 'on'          -- W6
4691     end
4692
4693     -- EN + CS/ES + EN      -- W4
4694     if d == 'en' and #nodes >= 2 then
4695         if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
4696             and nodes[#nodes-1][2] == 'en' then
4697             nodes[#nodes][2] = 'en'
4698         end
4699     end
4700
4701     -- AN + CS + AN          -- W4 too, because uax9 mixes both cases
4702     if d == 'an' and #nodes >= 2 then
4703         if (nodes[#nodes][2] == 'cs')
4704             and nodes[#nodes-1][2] == 'an' then
4705             nodes[#nodes][2] = 'an'
4706         end
4707     end
4708
4709     -- ET/EN                -- W5 + W7->1 / W6->on
4710     if d == 'et' then
4711         first_et = first_et or (#nodes + 1)
4712     elseif d == 'en' then
4713         has_en = true
4714         first_et = first_et or (#nodes + 1)
4715     elseif first_et then    -- d may be nil here !
4716         if has_en then

```

```

4717         if last == 'l' then
4718             temp = 'l'      -- W7
4719         else
4720             temp = 'en'     -- W5
4721         end
4722     else
4723         temp = 'on'        -- W6
4724     end
4725     for e = first_et, #nodes do
4726         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
4727     end
4728     first_et = nil
4729     has_en = false
4730 end
4731
4732 if d then
4733     if d == 'al' then
4734         d = 'r'
4735         last = 'al'
4736     elseif d == 'l' or d == 'r' then
4737         last = d
4738     end
4739     prev_d = d
4740     table.insert(nodes, {item, d, outer_first})
4741 end
4742
4743 outer_first = nil
4744
4745 end
4746
4747 -- TODO -- repeated here in case EN/ET is the last node. Find a
4748 -- better way of doing things:
4749 if first_et then      -- dir may be nil here !
4750     if has_en then
4751         if last == 'l' then
4752             temp = 'l'      -- W7
4753         else
4754             temp = 'en'     -- W5
4755         end
4756     else
4757         temp = 'on'        -- W6
4758     end
4759     for e = first_et, #nodes do
4760         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
4761     end
4762 end
4763
4764 -- dummy node, to close things
4765 table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
4766
4767 ----- NEUTRAL -----
4768
4769 outer = save_outer
4770 last = outer
4771
4772 local first_on = nil
4773
4774 for q = 1, #nodes do
4775     local item

```

```

4776
4777     local outer_first = nodes[q][3]
4778     outer = outer_first or outer
4779     last = outer_first or last
4780
4781     local d = nodes[q][2]
4782     if d == 'an' or d == 'en' then d = 'r' end
4783     if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
4784
4785     if d == 'on' then
4786         first_on = first_on or q
4787     elseif first_on then
4788         if last == d then
4789             temp = d
4790         else
4791             temp = outer
4792         end
4793         for r = first_on, q - 1 do
4794             nodes[r][2] = temp
4795             item = nodes[r][1] -- MIRRORING
4796             if Babel.mirroring_enabled and item.id == GLYPH
4797                 and temp == 'r' and characters[item.char] then
4798                 item.char = characters[item.char].m or item.char
4799             end
4800         end
4801         first_on = nil
4802     end
4803
4804     if d == 'r' or d == 'l' then last = d end
4805 end
4806
4807 ----- IMPLICIT, REORDER -----
4808
4809 outer = save_outer
4810 last = outer
4811
4812 local state = {}
4813 state.has_r = false
4814
4815 for q = 1, #nodes do
4816
4817     local item = nodes[q][1]
4818
4819     outer = nodes[q][3] or outer
4820
4821     local d = nodes[q][2]
4822
4823     if d == 'nsm' then d = last end -- W1
4824     if d == 'en' then d = 'an' end
4825     local isdir = (d == 'r' or d == 'l')
4826
4827     if outer == 'l' and d == 'an' then
4828         state.san = state.san or item
4829         state.ean = item
4830     elseif state.san then
4831         head, state = insert_numeric(head, state)
4832     end
4833
4834     if outer == 'l' then

```



```

4835     if d == 'an' or d == 'r' then      -- im -> implicit
4836         if d == 'r' then state.has_r = true end
4837         state.sim = state.sim or item
4838         state.eim = item
4839     elseif d == 'l' and state.sim and state.has_r then
4840         head, state = insert_implicit(head, state, outer)
4841     elseif d == 'l' then
4842         state.sim, state.eim, state.has_r = nil, nil, false
4843     end
4844 else
4845     if d == 'an' or d == 'l' then
4846         if nodes[q][3] then -- nil except after an explicit dir
4847             state.sim = item -- so we move sim 'inside' the group
4848         else
4849             state.sim = state.sim or item
4850         end
4851         state.eim = item
4852     elseif d == 'r' and state.sim then
4853         head, state = insert_implicit(head, state, outer)
4854     elseif d == 'r' then
4855         state.sim, state.eim = nil, nil
4856     end
4857 end
4858
4859 if isdir then
4860     last = d      -- Don't search back - best save now
4861 elseif d == 'on' and state.san then
4862     state.san = state.san or item
4863     state.ean = item
4864 end
4865
4866 end
4867
4868 return node.prev(head) or head
4869 end
4870 </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.

```

4871 <nil>
4872 \ProvidesLanguage{nil}[<<date>>] <<version>> Nil language]
4873 \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.

```

4874 \ifx\l@nil\undefined
4875   \newlanguage\l@nil
4876   \@namedef{bbl@hyphendata@the\l@nil}{\{}}% Remove warning

```

```
4877 \fi
```

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

```
4878 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}
```

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

```
\captionnil
```

```
\datenil 4879 \let\captionnil\@empty
```

```
4880 \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.

```
4881 \ldf@finish{nil}
```

```
4882 \</nil>
```

## 17 Support for Plain $\text{\TeX}$ (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  $\text{\TeX}$ -format. When asked he responded:

That file name is “sacred”, and if anybody changes it they will cause severe upward/downward compatibility headaches.

People can have a file `localhyphen.tex` or whatever they like, but they mustn’t diddle with `hyphen.tex` (or `plain.tex` except to preload additional fonts).

The files `bplain.tex` and `blplain.tex` can be used as replacement wrappers around `plain.tex` and `lplain.tex` to achieve the desired effect, based on the `babel` package. If you load each of them with  $\text{\TeX}$ , you will get a file called either `bplain.fmt` or `blplain.fmt`, which you can use as replacements for `plain.fmt` and `lplain.fmt`. As these files are going to be read as the first thing  $\text{\TeX}$  sees, we need to set some category codes just to be able to change the definition of `\input`

```
4883 (*bplain | blplain)
```

```
4884 \catcode`\{=1 % left brace is begin-group character
```

```
4885 \catcode`\}=2 % right brace is end-group character
```

```
4886 \catcode`\#=6 % hash mark is macro parameter character
```

Now let’s see if a file called `hyphen.cfg` can be found somewhere on  $\text{\TeX}$ ’s input path by trying to open it for reading...

```
4887 \openin 0 hyphen.cfg
```

If the file wasn’t found the following test turns out true.

```
4888 \ifeof0
```

```
4889 \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).

```
4890 \let\ainput
```

Then `\input` is defined to forget about its argument and load `hyphen.cfg` instead.

```
4891 \def\input #1 {%
4892   \let\input\@
4893   \a hyphen.cfg
```

Once that's done the original meaning of `\input` can be restored and the definition of `\a` can be forgotten.

```
4894   \let\@undefined
4895 }
4896 \fi
4897 </bplain | bplain>
```

Now that we have made sure that `hyphen.cfg` will be loaded at the right moment it is time to load `plain.tex`.

```
4898 <bplain>\a plain.tex
4899 <bplain>\a lplain.tex
```

Finally we change the contents of `\fmtname` to indicate that this is *not* the plain format, but a format based on plain with the `babel` package preloaded.

```
4900 <bplain>\def\fmtname{babel-plain}
4901 <bplain>\def\fmtname{babel-lplain}
```

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

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

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

```
4902 <*plain>
4903 \def\@empty{}
4904 \def\loadlocalcfg#1{%
4905   \openin0#1.cfg
4906   \ifeof0
4907     \closein0
4908   \else
4909     \closein0
4910     {\immediate\write16{*****}%
4911      \immediate\write16{* Local config file #1.cfg used}%
4912      \immediate\write16{*}%
4913     }
4914     \input #1.cfg\relax
4915   \fi
4916   \@endofldf}
```

## 17.3 General tools

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

```
4917 \long\def\@firstofone#1{#1}
4918 \long\def\@firstoftwo#1#2{#1}
4919 \long\def\@secondoftwo#1#2{#2}
4920 \def\@nnil{\@nil}
4921 \def\@gobbletwo#1#2{}
4922 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
4923 \def\@star@or@long#1{%
4924   \@ifstar
4925   {\let\l@ngrel@x\relax#1}%
4926   {\let\l@ngrel@x\long#1}}
```

```

4927 \let\l@ngrel@x\relax
4928 \def\@car#1#2\@nil{#1}
4929 \def\@cdr#1#2\@nil{#2}
4930 \let\@typeset@protect\relax
4931 \let\protected@edef\edef
4932 \long\def\@gobble#1{}
4933 \edef\@backslashchar{\expandafter\@gobble\string\}
4934 \def\strip@prefix#1>{}
4935 \def\g@addto@macro#1#2{%
4936     \toks@\expandafter{#1#2}%
4937     \xdef#1{\the\toks@}}
4938 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
4939 \def\@nameuse#1{\csname #1\endcsname}
4940 \def\@ifundefined#1{%
4941     \expandafter\ifx\csname#1\endcsname\relax
4942     \expandafter\@firstoftwo
4943     \else
4944     \expandafter\@secondoftwo
4945     \fi}
4946 \def\@expandtwoargs#1#2#3{%
4947     \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
4948 \def\zap@space#1 #2{%
4949     #1%
4950     \ifx#2\@empty\else\expandafter\zap@space\fi
4951     #2}

```

$\text{\LaTeX} 2_{\epsilon}$  has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

4952 \ifx\@preamblecmds\@undefined
4953     \def\@preamblecmds{}
4954 \fi
4955 \def\@onlypreamble#1{%
4956     \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
4957         \@preamblecmds\do#1}}
4958 \@onlypreamble\@onlypreamble

```

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

```

4959 \def\begin{document}{%
4960     \@begin{document}hook
4961     \global\let\@begin{document}hook\@undefined
4962     \def\do##1{\global\let##1\@undefined}%
4963     \@preamblecmds
4964     \global\let\do\noexpand}

4965 \ifx\@begin{document}hook\@undefined
4966     \def\@begin{document}hook{}
4967 \fi
4968 \@onlypreamble\@begin{document}hook
4969 \def\AtBeginDocument{\g@addto@macro\@begin{document}hook}

```

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

```

4970 \def\AtEndOfPackage#1{\g@addto@macro\@endoflfd{#1}}
4971 \@onlypreamble\AtEndOfPackage
4972 \def\@endoflfd{}
4973 \@onlypreamble\@endoflfd
4974 \let\bbl@afterlang\@empty
4975 \chardef\bbl@opt@hyphenmap\z@

```

L<sup>A</sup>T<sub>E</sub>X needs to be able to switch off writing to its auxiliary files; plain doesn't have them by default.

```

4976 \ifx\if@filesw\@undefined
4977   \expandafter\let\csname if@filesw\expandafter\endcsname
4978     \csname iffalse\endcsname
4979 \fi

```

Mimick L<sup>A</sup>T<sub>E</sub>X's commands to define control sequences.

```

4980 \def\newcommand{\@star@or@long\new@command}
4981 \def\new@command#1{%
4982   \@testopt{\@newcommand#1}0}
4983 \def\@newcommand#1[#2]{%
4984   \@ifnextchar [{\@xargdef#1[#2]}%
4985     {\@argdef#1[#2]}}
4986 \long\def\@argdef#1[#2]#3{%
4987   \@yargdef#1\@ne{#2}{#3}}
4988 \long\def\@xargdef#1[#2][#3]#4{%
4989   \expandafter\def\expandafter#1\expandafter{%
4990     \expandafter\@protected@testopt\expandafter #1%
4991     \csname\string#1\expandafter\endcsname{#3}}%
4992   \expandafter\@yargdef \csname\string#1\endcsname
4993   \tw@{#2}{#4}}
4994 \long\def\@yargdef#1#2#3{%
4995   \@tempcnta#3\relax
4996   \advance \@tempcnta \@ne
4997   \let\@hash@\relax
4998   \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
4999   \@tempcntb #2%
5000   \@whilenum \@tempcntb < \@tempcnta
5001   \do{%
5002     \edef\reserved@a{\reserved@a\@hash@the\@tempcntb}%
5003     \advance \@tempcntb \@ne}%
5004   \let\@hash@###
5005   \l@ngrelx\expandafter\def\expandafter#1\reserved@a}
5006 \def\providecommand{\@star@or@long\provide@command}
5007 \def\provide@command#1{%
5008   \begingroup
5009     \escapechar\m@ne\xdef\@gtempa{\string#1}%
5010   \endgroup
5011   \expandafter\@ifundefined\@gtempa
5012     {\def\reserved@a{\new@command#1}}%
5013     {\let\reserved@a\relax
5014       \def\reserved@a{\new@command\reserved@a}}%
5015   \reserved@a}%
5016 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
5017 \def\declare@robustcommand#1{%
5018   \edef\reserved@a{\string#1}%
5019   \def\reserved@b{#1}%
5020   \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
5021   \edef#1{%
5022     \ifx\reserved@a\reserved@b
5023       \noexpand\x@protect
5024       \noexpand#1%
5025     \fi
5026     \noexpand\protect
5027     \expandafter\noexpand\csname
5028       \expandafter\@gobble\string#1 \endcsname
5029   }%

```

```

5030 \expandafter\new@command\csname
5031 \expandafter@gobble\string#1 \endcsname
5032 }
5033 \def\x@protect#1{%
5034 \ifx\protect\@typeset@protect\else
5035 \x@protect#1%
5036 \fi
5037 }
5038 \def\x@protect#1\fi#2#3{%
5039 \fi\protect#1%
5040 }

```

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`.

```

5041 \def\bbl@tempa{\csname newif\endcsname\ifin@}
5042 \ifx\in@\@undefined
5043 \def\in@#1#2{%
5044 \def\in@##1##2##3\in@{%
5045 \ifx\in@##2\in@false\else\in@true\fi}%
5046 \in@#2#1\in@\in@}
5047 \else
5048 \let\bbl@tempa\@empty
5049 \fi
5050 \bbl@tempa

```

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

```

5051 \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.

```

5052 \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.

```

5053 \ifx\@tempcnta\@undefined
5054 \csname newcount\endcsname\@tempcnta\relax
5055 \fi
5056 \ifx\@tempcntb\@undefined
5057 \csname newcount\endcsname\@tempcntb\relax
5058 \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`).

```

5059 \ifx\bye\@undefined
5060 \advance\count10 by -2\relax
5061 \fi
5062 \ifx\@ifnextchar\@undefined
5063 \def\@ifnextchar#1#2#3{%
5064 \let\reserved@d=#1%
5065 \def\reserved@a{#2}\def\reserved@b{#3}%

```

```

5066 \futurelet\@let@token\@ifnch}
5067 \def\@ifnch{%
5068 \ifx\@let@token\@sptoken
5069 \let\reserved@c\@xifnch
5070 \else
5071 \ifx\@let@token\reserved@d
5072 \let\reserved@c\reserved@a
5073 \else
5074 \let\reserved@c\reserved@b
5075 \fi
5076 \fi
5077 \reserved@c}
5078 \def\:\let\@sptoken= } \: % this makes \@sptoken a space token
5079 \def\:\@xifnch} \expandafter\def\:\{\futurelet\@let@token\@ifnch}
5080 \fi
5081 \def\@testopt#1#2{%
5082 \@ifnextchar[{\#1}{\#1[#2]}}
5083 \def\@protected@testopt#1{%
5084 \ifx\protect\@typeset@protect
5085 \expandafter\@testopt
5086 \else
5087 \@x@protect#1%
5088 \fi}
5089 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
5090 #2\relax}\fi}
5091 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
5092 \else\expandafter\@gobble\fi{#1}}

```

## 17.4 Encoding related macros

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

```

5093 \def\DeclareTextCommand{%
5094 \@dec@text@cmd\providecommand
5095 }
5096 \def\ProvideTextCommand{%
5097 \@dec@text@cmd\providecommand
5098 }
5099 \def\DeclareTextSymbol#1#2#3{%
5100 \@dec@text@cmd\chardef#1{#2}#3\relax
5101 }
5102 \def\@dec@text@cmd#1#2#3{%
5103 \expandafter\def\expandafter#2%
5104 \expandafter{%
5105 \csname#3-cmd\expandafter\endcsname
5106 \expandafter#2%
5107 \csname#3\string#2\endcsname
5108 }%
5109 % \let\@ifdefinable\@rc@ifdefinable
5110 \expandafter#1\csname#3\string#2\endcsname
5111 }
5112 \def\@current@cmd#1{%
5113 \ifx\protect\@typeset@protect\else
5114 \noexpand#1\expandafter\@gobble
5115 \fi
5116 }
5117 \def\@changed@cmd#1#2{%
5118 \ifx\protect\@typeset@protect
5119 \expandafter\ifx\csname#1\endcsname\relax

```

```

5120         \expandafter\ifx\csname ?\string#1\endcsname\relax
5121         \expandafter\def\csname ?\string#1\endcsname{%
5122             \@changed@x@err{#1}%
5123         }%
5124     \fi
5125     \global\expandafter\let
5126     \csname\cf@encoding\string#1\expandafter\endcsname
5127     \csname ?\string#1\endcsname
5128 \fi
5129 \csname\cf@encoding\string#1%
5130 \expandafter\endcsname
5131 \else
5132     \noexpand#1%
5133 \fi
5134 }
5135 \def\@changed@x@err#1{%
5136     \errhelp{Your command will be ignored, type <return> to proceed}%
5137     \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
5138 \def\DeclareTextCommandDefault#1{%
5139     \DeclareTextCommand#1?%
5140 }
5141 \def\ProvideTextCommandDefault#1{%
5142     \ProvideTextCommand#1?%
5143 }
5144 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
5145 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
5146 \def\DeclareTextAccent#1#2#3{%
5147     \DeclareTextCommand#1{#2}[1]{\accent#3 #1}
5148 }
5149 \def\DeclareTextCompositeCommand#1#2#3#4{%
5150     \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
5151     \edef\reserved@b{\string##1}%
5152     \edef\reserved@c{%
5153         \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
5154     \ifx\reserved@b\reserved@c
5155         \expandafter\expandafter\expandafter\ifx
5156             \expandafter\@car\reserved@a\relax\relax\@nil
5157             \@text@composite
5158         \else
5159             \edef\reserved@b##1{%
5160                 \def\expandafter\noexpand
5161                 \csname#2\string#1\endcsname####1{%
5162                     \noexpand\@text@composite
5163                     \expandafter\noexpand\csname#2\string#1\endcsname
5164                     ####1\noexpand\@empty\noexpand\@text@composite
5165                     {##1}%
5166                 }%
5167             }%
5168             \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
5169         \fi
5170     \expandafter\def\csname\expandafter\string\csname
5171         #2\endcsname\string#1-\string#3\endcsname{#4}
5172 \else
5173     \errhelp{Your command will be ignored, type <return> to proceed}%
5174     \errmessage{\string\DeclareTextCompositeCommand\space used on
5175         inappropriate command \protect#1}
5176 \fi
5177 }
5178 \def\@text@composite#1#2#3\@text@composite{%

```



```

5179 \expandafter\@text@composite@x
5180 \csname\string#1-\string#2\endcsname
5181 }
5182 \def\@text@composite@x#1#2{%
5183 \ifx#1\relax
5184 #2%
5185 \else
5186 #1%
5187 \fi
5188 }
5189 %
5190 \def\@strip@args#1:#2-#3\@strip@args{#2}
5191 \def\DeclareTextComposite#1#2#3#4{%
5192 \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
5193 \bgroup
5194 \lccode`\@=#4%
5195 \lowercase{%
5196 \egroup
5197 \reserved@a @%
5198 }%
5199 }
5200 %
5201 \def\UseTextSymbol#1#2{%
5202 % \let\@curr@enc\cf@encoding
5203 % \@use@text@encoding{#1}%
5204 #2%
5205 % \@use@text@encoding\@curr@enc
5206 }
5207 \def\UseTextAccent#1#2#3{%
5208 % \let\@curr@enc\cf@encoding
5209 % \@use@text@encoding{#1}%
5210 #2{\@use@text@encoding\@curr@enc\selectfont#3}%
5211 % \@use@text@encoding\@curr@enc
5212 }
5213 \def\@use@text@encoding#1{%
5214 % \edef\font@name{#1}%
5215 % \xdef\font@name{%
5216 % \csname\curr@fontshape/\font@size\endcsname
5217 % }%
5218 % \pickup@font
5219 % \font@name
5220 % \@@enc@update
5221 }
5222 \def\DeclareTextSymbolDefault#1#2{%
5223 \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
5224 }
5225 \def\DeclareTextAccentDefault#1#2{%
5226 \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
5227 }
5228 \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.

```

5229 \DeclareTextAccent{"}{OT1}{127}
5230 \DeclareTextAccent{'}{OT1}{19}
5231 \DeclareTextAccent{^}{OT1}{94}
5232 \DeclareTextAccent{\`}{OT1}{18}
5233 \DeclareTextAccent{\~}{OT1}{126}

```

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

```
5234 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
5235 \DeclareTextSymbol{\textquotedblright}{OT1}{`\'}
5236 \DeclareTextSymbol{\textquoteleft}{OT1}{`\'}
5237 \DeclareTextSymbol{\textquoteright}{OT1}{`\'}
5238 \DeclareTextSymbol{\i}{OT1}{16}
5239 \DeclareTextSymbol{\ss}{OT1}{25}
```

For a couple of languages we need the  $\TeX$ -control sequence `\scriptsize` to be available. Because plain  $\TeX$  doesn't have such a sophisticated font mechanism as  $\LaTeX$  has, we just `\let` it to `\sevenrm`.

```
5240 \ifx\scriptsize\undefined
5241   \let\scriptsize\sevenrm
5242 \fi
5243 \</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 Smitschuijzen Abifares, *Arabic Typography*, Saqi, 2001.
- [2] Donald E. Knuth, *The  $\TeX$ book*, Addison-Wesley, 1986.
- [3] Leslie Lamport,  *$\LaTeX$ , A document preparation System*, Addison-Wesley, 1986.
- [4] K.F. Treebus. *Tekstwijzer, een gids voor het grafisch verwerken van tekst*, SDU Uitgeverij ('s-Gravenhage, 1988).
- [5] Hubert Partl, *German  $\TeX$ , TUGboat* 9 (1988) #1, p. 70–72.
- [6] Leslie Lamport, in:  $\TeX$ hax Digest, Volume 89, #13, 17 February 1989.
- [7] Johannes Braams, Victor Eijkhout and Nico Poppelier, *The development of national  $\LaTeX$  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  $\LaTeX$  is ready to use*, *TUGboat* 11 (1990) #1, p. 87–90.
- [12] Apostolos Syropoulos, Antonis Tsolomititis and Nick Sofroniu, *Digital typography using  $\LaTeX$* , Springer, 2002, p. 301–373.