

# Babel

Version 3.9m  
2015/08/03

*Original author*  
Johannes L. Braams

*Current maintainer*  
Javier Bezos

The standard distribution of  $\LaTeX$  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  $\LaTeX$  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  $\TeX$  version 3 and, to some extent, 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. However, no attempt has been done to take full advantage of the features provided by the latter, which would require a completely new core (as for example polyglossia or as part of  $\LaTeX$ 3).

# Contents

|           |  |           |
|-----------|--|-----------|
| <b>I</b>  | <b>User guide</b>  | <b>4</b>  |
| <b>1</b>  | <b>The user interface</b>  | <b>4</b>  |
| 1.1       | Selecting languages . . . . .  | 5         |
| 1.2       | More on selection . . . . .  | 7         |
| 1.3       | Getting the current language name . . . . .                                      | 8         |
| 1.4       | Selecting scripts . . . . .  | 8         |
| 1.5       | Shorthands . . . . .   | 9         |
| 1.6       | Package options . . . . .  | 12        |
| 1.7       | The base option . . . . .  | 14        |
| 1.8       | Hooks . . . . .  | 14        |
| 1.9       | Hyphenation tools . . . . .  | 16        |
| 1.10      | Language attributes . . . . .  | 17        |
| 1.11      | Languages supported by babel . . . . .   | 17        |
| 1.12      | Tips, workarounds, know issues and notes . . . . .                               | 19        |
| 1.13      | Future work . . . . .  | 20        |
| <b>2</b>  | <b>Preloading languages with language.dat</b>                                    | <b>21</b> |
| <b>3</b>  | <b>The interface between the core of babel and the language definition files</b> | <b>22</b> |
| 3.1       | Basic macros . . . . .   | 23        |
| 3.2       | Skeleton . . . . .   | 24        |
| 3.3       | Support for active characters . . . . .  | 25        |
| 3.4       | Support for saving macro definitions . . . . .                                   | 26        |
| 3.5       | Support for extending macros . . . . .   | 26        |
| 3.6       | Macros common to a number of languages . . . . .                                 | 26        |
| 3.7       | Encoding-dependent strings . . . . .   | 27        |
| <b>4</b>  | <b>Compatibility and changes</b>   | <b>30</b> |
| 4.1       | Compatibility with german.sty . . . . .  | 30        |
| 4.2       | Compatibility with ngerman.sty . . . . .   | 31        |
| 4.3       | Compatibility with the french package . . . . .                                  | 31        |
| 4.4       | Changes in babel version 3.9 . . . . .   | 31        |
| 4.5       | Changes in babel version 3.7 . . . . .   | 31        |
| 4.6       | Changes in babel version 3.6 . . . . .   | 32        |
| 4.7       | Changes in babel version 3.5 . . . . .   | 33        |
| <b>II</b> | <b>The code</b>  | <b>34</b> |
| <b>5</b>  | <b>Identification and loading of required files</b>                              | <b>34</b> |
| 5.1       | Multiple languages . . . . .   | 35        |
| <b>6</b>  | <b>The Package File (L<sup>A</sup>T<sub>E</sub>X)</b>                            | <b>36</b> |
| 6.1       | base . . . . .   | 37        |
| 6.2       | key=value options and other general option . . . . .                             | 37        |
| 6.3       | Conditional loading of shorthands . . . . .                                      | 38        |
| 6.4       | Language options . . . . .   | 40        |

|           |   |            |
|-----------|---|------------|
| <b>7</b>  | <b>The kernel of Babel (common)</b>                               | <b>43</b>  |
| 7.1       | Tools . . . . .   | 43         |
| 7.2       | Hooks . . . . .   | 46         |
| 7.3       | Setting up language files . . . . .                               | 48         |
| 7.4       | Shorthands . . . . .  | 49         |
| 7.5       | Language attributes . . . . .                                     | 59         |
| 7.6       | Support for saving macro definitions . . . . .                    | 62         |
| 7.7       | Short tags . . . . .  | 63         |
| 7.8       | Hyphens . . . . .   | 63         |
| 7.9       | Multiencoding strings . . . . .                                   | 65         |
| 7.10      | Macros common to a number of languages . . . . .                  | 71         |
| 7.11      | Making glyphs available . . . . .                                 | 71         |
|           | 7.11.1 Quotation marks . . . . .                                  | 71         |
|           | 7.11.2 Letters . . . . .  | 72         |
|           | 7.11.3 Shorthands for quotation marks . . . . .                   | 73         |
|           | 7.11.4 Umlauts and tremas . . . . .                               | 75         |
| <b>8</b>  | <b>The kernel of Babel (only L<sup>A</sup>T<sub>E</sub>X)</b>     | <b>76</b>  |
| 8.1       | The redefinition of the style commands . . . . .                  | 76         |
| 8.2       | Cross referencing macros . . . . .                                | 76         |
| 8.3       | Marks . . . . .   | 80         |
| 8.4       | Preventing clashes with other packages . . . . .                  | 81         |
|           | 8.4.1 ifthen . . . . .  | 81         |
|           | 8.4.2 varioref . . . . .  | 82         |
|           | 8.4.3 hhline . . . . .  | 83         |
|           | 8.4.4 hyperref . . . . .  | 83         |
|           | 8.4.5 fancyhdr . . . . .  | 83         |
| 8.5       | Encoding issues . . . . .   | 84         |
| 8.6       | Local Language Configuration . . . . .                            | 86         |
| <b>9</b>  | <b>Internationalizing L<sup>A</sup>T<sub>E</sub>X 2.09</b>        | <b>86</b>  |
| <b>10</b> | <b>Multiple languages</b>   | <b>91</b>  |
| 10.1      | Selecting the language . . . . .                                  | 92         |
| 10.2      | Errors . . . . .  | 99         |
| <b>11</b> | <b>Loading hyphenation patterns</b>                               | <b>100</b> |
| <b>12</b> | <b>The ‘nil’ language</b>   | <b>105</b> |
| <b>13</b> | <b>Support for Plain T<sub>E</sub>X</b>                           | <b>106</b> |
| 13.1      | Not renaming hyphen.tex . . . . .                                 | 106        |
| 13.2      | Emulating some L <sup>A</sup> T <sub>E</sub> X features . . . . . | 107        |
| 13.3      | General tools . . . . .   | 107        |
| 13.4      | Encoding related macros . . . . .                                 | 111        |
| 13.5      | Babel options . . . . .   | 114        |
| <b>14</b> | <b>Tentative font handling</b>                                    | <b>115</b> |
| <b>15</b> | <b>Hooks for XeT<sub>E</sub>X and LuaT<sub>E</sub>X</b>           | <b>115</b> |
| 15.1      | XeT <sub>E</sub> X . . . . .                                      | 115        |
| 15.2      | LuaT <sub>E</sub> X . . . . .                                     | 117        |
| <b>16</b> | <b>Conclusion</b>   | <b>119</b> |



# Part I

## User guide

### 1 The user interface

The basic user interface of this package is quite simple. It consists of a set of commands that switch from one language to another, and a set of commands that deal with shorthands. It is also possible to find out what the current language is. In most cases, a single language is required, and then all you need in L<sup>A</sup>T<sub>E</sub>X is to load the package using its standard mechanism for this purpose, namely, passing that language as an optional argument. In multilingual documents, just use several option. So, in L<sup>A</sup>T<sub>E</sub>X2<sub>ε</sub> the preamble of the document:

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

would tell L<sup>A</sup>T<sub>E</sub>X that the document would be written in two languages, Dutch and English, and that English would be the first language in use, and the main one. You can also set the main language explicitly:

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

Another approach is making dutch and english global options in order to let other packages detect and use them:

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

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

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}
```

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

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

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

---

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

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. 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 'LANG'. Either you have misspelled
(babel)                its name, it has not been installed, or you requested
(babel)                it in a previous run. Fix its name, install it or just
(babel)                rerun the file, respectively
```

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

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

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

## 1.1 Selecting languages

This section describes the commands to be used in the document to switch the language in multilingual document.

The main language is selected automatically when the document environment begins. In the preamble it has *not* been selected, except hyphenation patterns and the name assigned to `\languagename` (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 following commands.

`\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. For “historical reasons”, a macro name is converted to a

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

<sup>3</sup>In former 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 will be fixed soon.

language name without the leading \; in other words, the two following declarations are equivalent:

```
\selectlanguage{german}
\selectlanguage{\german}
```

Using a macro instead of a “real” name is deprecated.

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.

This command can be used as environment, too.

`\begin{otherlanguage}` {<language>} ... `\end{otherlanguage}`

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

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

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

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

Spaces after the environment are ignored.

`\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 argument. 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).

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

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

This environment (or in some cases `otherlanguage`) may be required for intermixing left-to-right typesetting with right-to-left typesetting in engines not supporting a change in the writing direction inside a line.

`\begin{hyphenrules}` {<language>} ... `\end{hyphenrules}`

The environment `hyphenrules` can be used to select *only* the hyphenation rules to be used (it can be used as command, too). This can for instance be used to select

'nohyphenation', provided that in `language.dat` the 'language' nohyphenation is defined by loading `zerohyph.tex`. It deactivates language shorthands, too (but not user shorthands).

Except for these simple uses, `hyphenrules` is discouraged and `otherlanguage*` (the starred version) is preferred, as the former does not take into account possible changes in encodings or characters like, say, ' done by some languages (eg, `italian`, `frenchb`, `ukraineb`). To set hyphenation exceptions, use `\babelhyphenation` (see below).

## 1.2 More on selection

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

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

```
\babetags{de = german}
```

yo can write

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

and

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

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

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

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

Of course, T<sub>E</sub>X 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. 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}
```

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



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`).

### 1.3 Getting the current language name

`\language` The control sequence `\language` contains the name of the current language. However, due to some internal inconsistencies in catcodes it should *not* be used to test its value (use `iflang`, by Heiko Oberdiek).

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

If more than one language is used, it might be necessary to know which language is active at a specific time. This can be checked by a call to `\iflanguage`, but note here “language” is used in the  $\TeX$  sense, as a set of hyphenation patterns, and *not* as its babel name. This macro takes three arguments. The first argument is the name of a language; the second and third arguments are the actions to take if the result of the test is true or false respectively. The advice about `\language` also applies here – use `iflang` instead of `\iflanguage` if possible.

### 1.4 Selecting scripts

Currently babel provides no standard interface to select scripts, because they are best selected with either `\fontencoding` (low level) or a language name (high level). Even the Latin script may require different encodings (ie, sets of glyphs) depending on the language, and therefore such a switch would be in a sense incomplete.<sup>6</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>7</sup>

`\ensureascii` `{\text}`

**New 3.9i** This macro makes sure `\text` is typeset with a LICR-savvy encoding in the ASCII range. It is used to redefine `\TeX` and `\LaTeX` so that they are correctly typeset even with LGR or X2 (the complete list is stored in `\BabelNonASCII`, which by default is LGR, X2, OT2, OT3, OT6, LHE, LWN, LMA, LMC, LMS, LMU, but you can modify it). So, in some sense it fixes the bug described in the previous paragraph. If non-ASCII encodings are not loaded (or no encoding at all), it is no-op (also `\TeX` and `\LaTeX` are not redefined); otherwise, `\ensureascii` switches to the encoding at the beginning of the document if ASCII-savvy, or else the last ASCII-savvy encoding loaded. For example, if you load LY1, LGR, then it is set to LY1, but if you load LY1, T2A it is set to T2A. The symbol encodings TS1, T3, and TS3 are not taken into account, since they are not used for “ordinary” text.

<sup>6</sup>The so-called Unicode fonts does not improve the situation either. So, a font suited for Vietnamese is not necessarily suited for, say, romanization of Indic languages, and the fact it contains glyphs for Modern Greek does not mean it includes them for Classic Greek. As to directionality, it poses special challenges because it also affects individual characters and layout elements.

<sup>7</sup>But still defined for backwards compatibility.

The foregoing rules (which are applied “at begin document”) cover most of cases. Note 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.

## 1.5 Shorthands

A *shorthand* is a sequence of one or two characters that expands to arbitrary T<sub>E</sub>X code.

Shorthands can be used for different kinds of things, as for example: (1) in some languages shorthands such as "a are defined to be able to hyphenate the word if the encoding is OT1; (2) in some languages shorthands such as ! are used to insert the right amount of white space; (3) several kinds of discretionaries and breaks can be inserted easily with "-, "=", etc.

The package inputenc as well as xetex an luatex have alleviated entering non-ASCII characters, but minority languages and some kinds of text can still require characters not directly available in the keyboards (and sometimes not even as separated or precomposed Unicode characters). As to the point 2, now pdfTeX provides \knbccode. Tools of 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.

Please, 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.

A typical error when using shorthands is the following:

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

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

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

It is sometimes necessary to switch a shorthand character off temporarily, because it must be used in an entirely different way. For this purpose, the user commands \shorthandoff and \shorthandon are provided. They each take a list of characters as their arguments.

The command \shorthandoff sets the \catcode for each of the characters in its argument to other (12); the command \shorthandon sets the \catcode to active (13). Both commands only work on ‘known’ shorthand characters. If a character is not known to be a shorthand character its category code will be left unchanged.

**New 3.9a** Note 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.

`\usesshorthands` \*{<char>}

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

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

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

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

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

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

As an example of their applications, 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, behaviour of hyphens is language dependent. For example, in languages like Polish and Portugese, 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{"}{/}`. Please note the substitute character must *not* have been declared before as shorthand (in such case, `\aliasshorthands` is ignored). The following example shows how to replace a shorthand by another

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

However, 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, `^` 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>8</sup> Note that for this to work the language should have been specified as an option when loading the babel package. For example, you can use in english the shorthands defined by `ngerman` with

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

(You may also need to activate them with, for example, `\usesshorthands`.) Very often, this is a more convenient way to deactivate shorthands than `\shorthandoff`, as for example if you want to define a macro to easy typing phonetic characters with `tipa`:

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

### `\babelshorthand` `{<shorthand>}`

With this command you can use a shorthand even if (1) not activated in shorthands (in this case only shorthands for the current language are taken into account, ie, not user shorthands), (2) turned off with `\shorthandoff` or (3) deactivated with the internal `\bbl@deactivate`; for example, `\babelshorthand{"u}` or `\babelshorthand{:}`. (You can conveniently define your own macros, or even you own user shorthands provided they do not overlap.) For your records, here is a list of shorthands, but you must check them, as they may change:<sup>9</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** " ' ‘

<sup>8</sup>Actually, any name not corresponding to a language group does the same as none. However, follow this convention because it might be enforced in future releases of babel to catch possible errors.

<sup>9</sup>Thanks to Enrico Gregorio

**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>10</sup>

## 1.6 Package options

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

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

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

```
\usepackage[esperanto,frenchb,shorthands=;?!]{babel}
```

If ' is included, activeacute is set; if ' is included, activegrave is set. Active characters (like ~) should be preceded by \string (otherwise they will be expanded by L<sup>A</sup>T<sub>E</sub>X before they are passed to the package and therefore they will not be recognized); however, t is provided for the common case of ~ (as well as c for not so common case of the comma).

With shorthands=off no language shorthands are defined, As some languages use this mechanism for tools not available otherwise, a macro \babelshorthand is defined, which allows using them; see above.

**safe=** none | ref | bib

Some L<sup>A</sup>T<sub>E</sub>X macros are redefined so that using shorthands is safe. With safe=bib only \nocite, \biblecite 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. Of course, in such a case you cannot use shorthands in these macros.

<sup>10</sup>This declaration serves to nothing, but it is preserved for backward compatibility.

- math=** active | normal
- Shorthands are mainly intended for text, not for math. By setting this option with the value `normal` they are deactivated in math mode (default is `active`) and things like  $\{a'\}$  (a closing brace after a shorthand) are not a source of trouble any more.
- config=**  $\langle file \rangle$
- Load  $\langle file \rangle$ .`cfg` instead of the default config file `bblopts.cfg` (the file is loaded even with `noconfigs`).
- main=**  $\langle language \rangle$
- Sets the main language, as explained above, ie, this language is always loaded last. If it is not given as package or global option, it is added to the list of requested languages.
- headfoot=**  $\langle language \rangle$
- By default, headlines and footlines are not touched (only marks), and if they contain language dependent macros (which is not usual) there may be unexpected results. With this option you may set the language in heads and foots.
- noconfigs** Global and language default config files are not loaded, so you can make sure your document is not spoiled by an unexpected `.cfg` file. However, if the key config is set, this file is loaded.
- showlanguages** Prints to the log the list of languages loaded when the format was created: number (remember dialects can share it), name, hyphenation file and exceptions file.
- nocase** New 3.9l Language settings for uppercase and lowercase mapping (as set by `\SetCase`) are ignored. Use only if there are incompatibilities with other packages.
- silent** New 3.9l No warnings and no *infos* are written to the log file.<sup>11</sup>
- strings=** generic | unicode | encoded |  $\langle label \rangle$  |  $\langle font encoding \rangle$
- Selects the encoding of strings in languages supporting this feature. Predefined labels are `generic` (for traditional  $\TeX$ , LICR and ASCII strings), `unicode` (for engines like `xetex` and `luatex`) and `encoded` (for special cases requiring mixed encodings). Other allowed values are font encoding codes (`T1`, `T2A`, `LGR`, `L7X...`), but only in languages supporting them. Be aware with encoded captions are protected, but they work in `\MakeUppercase` and the like.
- hyphenmap=** off | main | select | other | other\*
- New 3.9g Sets the behaviour of case mapping for hyphenation, provided the language defines it.<sup>12</sup> It can take the following values:
- off** deactivates this feature and no case mapping is applied;

<sup>11</sup>You can use alternatively the package `silence`.

<sup>12</sup>Turned off in plain.

**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>13</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>14</sup>

## 1.7 The base option

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

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

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{frenchb}{...}
```

does ... at the end of `frenchb.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!`). For 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}
```

## 1.8 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` `{\langle name \rangle}{\langle event \rangle}{\langle code \rangle}`

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

<sup>13</sup>Duplicated options count as several ones.

<sup>14</sup>Providing `foreign` is pointless, because the case mapping applied is that at the end of paragraph, but if either `xetex` or `luatex` change this behaviour it might be added. On the other hand, `other` is provided even if I [JBL] think it isn't really useful, but who knows.

`\DisableBabelHook{<name>}`. Names containing the string `babel` are reserved (they are used, for example, by `\usesshortands*` to add a hook for the event `afterextras`).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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



`\BabelContentsFiles` **New 3.9a** This macro contains a list of “toc” types which require 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).

## 1.9 Hyphenation tools

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

**New 3.9a** It is customary to classify hyphens in two types: (1) *explicit* or *hard hyphens*, which in  $\TeX$  are entered as `-`, and (2) *optional* or *soft hyphens*, which are entered as `\-`. Strictly, a *soft hyphen* is not a hyphen, but just a breaking opportunity or, in  $\TeX$  terms, a “discretionary”; a *hard hyphen* is a hyphen with a breaking opportunity after it. A further type is a *non-breaking hyphen*, a hyphen without a breaking opportunity.

In  $\TeX$ , `-` and `\-` forbid further breaking opportunities in the word. This is the desired behaviour very often, but not always, and therefore many languages provide shorthands for these cases. Unfortunately, this has not been done consistently: for example, in Dutch, Portugese, Catalan or Danish, `-` is a hard hyphen, while in German, Spanish, Norwegian, Slovak or Russian, it is a soft hyphen. Furthermore, some of them even redefine `\-`, so that you cannot insert a soft hyphen without breaking opportunities in the rest of the word.

Therefore, some macros are provide with a set of basic “hyphens” which can be used by themselves, to define a user shorthand, or even in language files.

- `\babelhyphen{soft}` and `\babelhyphen{hard}` are self explanatory.
- `\babelhyphen{repeat}` inserts a hard hyphen which is repeated at the beginning of the next line, as done in languages like Polish, Portugese and Spanish.
- `\babelhyphen{nobreak}` inserts a hard hyphen without a break after it (even if a space follows).
- `\babelhyphen{empty}` inserts a break oportunity without a hyphen at all.
- `\babelhyphen{<text>}` is a hard “hyphen” using `<text>` instead. A typical case is `\babelhyphen{/}`.

With all of them hyphenation in the rest of the word is enabled. If you don’t want enabling it, there is a starred counterpart: `\babelhyphen*{soft}` (which in most cases is equivalent to the original `\-`), `\babelhyphen*{hard}`, etc.

Note `hard` is also good for isolated prefixes (eg, *anti-*) and `nobreak` for isolated suffixes (eg, *-ism*), but in both cases `\babelhyphen*{nobreak}` is usually better. There are also some differences with  $\LaTeX$ : (1) the character used is that set for the current font, while in  $\LaTeX$  it is hardwired to `-` (a typical value); (2) the hyphen to be used in fonts with a negative `\hyphenchar` is `-`, like in  $\LaTeX$ , but it can be changed to another value by redefining `\babelnullyhyphen`; (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.

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

**New 3.9m** *In luatex only*,<sup>15</sup> adds or replaces patterns for the languages given or, without the optional argument, for *all* languages. If a pattern for a certain combination already exists, it gets replaced by the new one.

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

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

## 1.10 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, `frenchb` uses `\frenchbsetup`, `magyar` (1.5) uses `\magyarOptions`; modifiers provided by `spanish` have no attribute counterparts. Macros setting options are also used (eg, `\ProsodicMarksOn` in `latin`).

## 1.11 Languages supported by babel

In the following table most of the languages supported by `babel` are listed, together with the names of the options which you can load `babel` with for each language. Note this list is open and the current options may be different.

**Afrikaans** afrikaans

<sup>15</sup>With `luatex` exceptions and patterns can be modified almost freely. However, this is very likely a task for a separate package and `babel` only provides the most basic tools.

**Bahasa** bahasa, indonesian, indon, bahasai, bahasam, malay, melayu  
**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  
**Interlingua** interlingua  
**Irish Gaelic** irish  
**Italian** italian  
**Latin** latin  
**Lower Sorbian** lowersorbian  
**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** upporsorbian  
**Welsh** welsh

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

Most of them work out of the box, but some may require extra fonts, encoding files, a preprocessor or even a complete framework (like CJK). 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 <file>`, which creates `<file>.tex`; you can then typeset the latter with  $\LaTeX$ .

## 1.12 Tips, workarounds, know issues and notes

- If you use the document class `book` and you use `\ref` inside the argument of `\chapter` (or just use `\ref` inside `\MakeUppercase`),  $\LaTeX$  will keep complaining about an undefined label. To prevent such problems, you could revert to using uppercase labels, you can use `\lowercase{\ref{foo}}` inside the argument of `\chapter`, or, if you will not use shorthands in labels, set the `safe` option to `none` or `bib`.
- Both `ltxdoc` and `babel` use `\AtBeginDocument` to change some catcodes, and `babel` reloads `hline` to make sure `:` has the right one, so if you want to change the catcode of `|` it has to be done using the same method at the proper place, with

```
\AtBeginDocument{\DeleteShortVerb{\|}}
```

before loading `babel`. This way, when the document begins the sequence is (1) make `|` active (`ltxdoc`); (2) make it unactive (your settings); (3) make `babel` shorthands active (`babel`); (4) reload `hline` (`babel`, now with the correct catcodes for `|` and `:`).

- Documents with several input encodings are not frequent, but sometimes are useful. You can set different encodings for different languages as the following example shows:

```
\addto\extrasfrench{\inputencoding{latin1}}
\addto\extrarussian{\inputencoding{koi8-r}}
```

(A recent version of `inputenc` is required.)

- For the hyphenation to work correctly, `lccodes` cannot change, because  $\TeX$  only takes into account the values when the paragraph is hyphenated, i.e., when it has been finished.<sup>16</sup> So, if you write a chunk of French text with `\foreignlanguage`, the apostrophes might not be taken into account. This is a limitation of  $\TeX$ , not of `babel`. Alternatively, you may use `\usesshorthands` to activate `'` and `\defineshorthand`, 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).

<sup>16</sup>This explains why  $\LaTeX$  assumes the lowercase mapping of T1 and does not provide a tool for multiple mappings. Unfortunately, `\savingshyphcodes` is not a solution either, because `lccodes` for hyphenation are frozen in the format and cannot be changed.

- 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. (Another issue in the ‘to do’ list, although there is a partial solution.)
- Plain luatex does not load patterns on the fly. Since this format is not based on babel but on etex.src further investigation is required. This is another task in the ‘to do’ list.

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.

### 1.13 Future work

Useful additions would be, for example, time, currency, addresses and personal names.<sup>17</sup> But that is the easy part, because they don’t require modifying the L<sup>A</sup>T<sub>E</sub>X internals.

More interesting are differences in the sentence structure or related to it. For example, in Basque the number precedes the name (including chapters), in Hungarian “from (1)” is “(1)-ból”, but “from (3)” is “(3)-ból”, in Spanish an item labelled “3.<sup>o</sup>” may be referred to as either “ítem 3.<sup>o</sup>” or “3.<sup>er</sup> ítem”, and so on. Even more interesting is right-to-left, vertical and bidi typesetting. 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).

Handling of “Unicode” fonts is also problematic. There is fontspec, but special macros are required (not only the NFSS ones) and it doesn’t provide “orthogonal axis” for features, including those related to the language (mainly language and script). A couple of tentative macros, which solve the two main cases, are provided by babel (≥3.9g) with a partial solution (only xetex and luatex, for obvious reasons), but use them at your own risk, as they might be removed in the future. For this very reason, they are described here:

- `\babelFSstore{(babel-language)}` sets the current three basic families (rm, sf, tt) as the default for the language given. In most cases, this macro will be enough.
- `\babelFSdefault{(babel-language)}{(fontspec-features)}` patches `\fontspec` so that the given features are always passed as the optional argument or added to it (not an ideal solution). Use it only if you select some fonts in the document with `\fontspec`.

<sup>17</sup>See for example POSIX, ISO 14652 and the Unicode Common Locale Data Repository (CLDR).

So, for example:

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

Note you can set any feature required for the language – not only Language, but also Script or a local .fea. This makes those macros a bit more verbose, but also more powerful.

## 2 Preloading languages with language.dat

TeX and most engines based on it (pdfTeX, xetex, e-TeX, the main exception being luatex) require hyphenation patterns to be loaded when a format is created (eg, L<sup>A</sup>TeX, XeL<sup>A</sup>TeX, pdfL<sup>A</sup>TeX). 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).

In that file the person who maintains a TeX environment has to record for which languages he has hyphenation patterns *and* in which files these are stored<sup>18</sup>. When hyphenation exceptions are stored in a separate file this can be indicated by naming that file *after* the file with the hyphenation patterns.

The file can contain empty lines and comments, as well as lines which start with an equals (=) sign. Such a line will instruct L<sup>A</sup>TeX 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>19</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
```

<sup>18</sup>This is because different operating systems sometimes use very different file-naming conventions.

<sup>19</sup>This is not a new feature, but in former versions it didn't work correctly.

```
preloaded for english instead}}
```

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

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

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

The following assumptions are made:

- Some of the language-specific definitions might be used by plain  $\text{T}_{\text{E}}\text{X}$  users, so the files have to be coded so that they can be read by both  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  and plain  $\text{T}_{\text{E}}\text{X}$ . 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`, `\langle lang \rangle captions`, `\langle lang \rangle date`, `\langle lang \rangle extras` and `\langle lang \rangle noextras` (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  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  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@⟨lang⟩` to be a dialect of `\language0` when `\l@⟨lang⟩` is undefined.
- Language names must be all lowercase. If an unknown language is selected, babel will attempt setting it after lowercasing its name.
- The semantics of modifiers is not defined (on purpose). In most cases, they will just be simple separated options (eg., `spanish`), but a language might require, say, a set of options organized as a tree with suboptions (in such a case, the recommended separator is `/`).

Some recommendations:

- The preferred shorthand is `"`, which is not used in  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  (quotes are entered as `'` and `'`). Other good choices are characters which are not used in a certain context (eg, `=` in an ancient language). Note however `=`, `<`, `>`, `:` and the like can be dangerous, because they may be used as part of the syntax of some elements (numeric expressions, key/value pairs, etc.).

- Captions should not contain shorthands or encoding dependent commands (the latter is not always possible, but should be clearly documented). They should be defined using the LICR. You may also use the new tools for encoded strings, described below.
- Avoid adding things to `\noextras⟨lang⟩` except for `umlauthigh` and friends, `\bbl@deactivate`, `\bbl@(non)frenchspacing`, and language specific macros. Use always, if possible, `\bbl@save` and `\bbl@savevariable` (except if you still want to have access to the previous value). Do not reset a macro or a setting to a hardcoded value. Never. Instead save its value in `\extras⟨lang⟩`.
- Do not switch scripts. If you want to make sure a set of glyphs is used, switch either the font encoding (low level) or the language (high level, which in turn may switch the font encoding). Usage of things like `\latintext` is deprecated.<sup>20</sup>

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 (and the corresponding PDF, if you like), as well as other files you think can be useful (eg, samples, readme).

### 3.1 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 T<sub>E</sub>X sense of set of hyphenation patterns.

`\adddialect` The macro `\adddialect` can be used when two languages can (or must) use the same hyphenation patterns. This can also be useful for languages for which no patterns are preloaded in the format. In such cases the default behaviour of the babel system is to define this language as a ‘dialect’ of the language for which the patterns were loaded as `\language0`. Here “language” is used in the T<sub>E</sub>X sense of set of hyphenation patterns.

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

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

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

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

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

`\date⟨lang⟩` The macro `\date⟨lang⟩` defines `\today`.

`\extras⟨lang⟩` The macro `\extras⟨lang⟩` contains all the extra definitions needed for a specific

<sup>20</sup>But not removed, for backward compatibility.



language. This macro, like the following, is a hook – you can add things to it, but it must not be used directly.

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

`\bbl@declare@attribute` This is a command to be used in the language definition files for declaring a language attribute. It takes three arguments: the name of the language, the attribute to be defined, and the code to be executed when the attribute is to be used.

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

`\ProvidesLanguage` The macro `\ProvidesLanguage` should be used to identify the language definition files. Its syntax is similar to the syntax of the  $\LaTeX$  command `\ProvidesPackage`.

`\LdfInit` The macro `\LdfInit` performs a couple of standard checks that must be made at the beginning of a language definition file, such as checking the category code of the @-sign, preventing the .ldf file from being processed twice, etc.

`\ldf@quit` The macro `\ldf@quit` does work needed if a .ldf file was processed earlier. This includes resetting the category code of the @-sign, preparing the language to be activated at `\begin{document}` time, and ending the input stream.

`\ldf@finish` The macro `\ldf@finish` does work needed at the end of each .ldf file. This includes resetting the category code of the @-sign, loading a local configuration file, and preparing the language to be activated at `\begin{document}` time.

`\loadlocalcfg` After processing a language definition file,  $\LaTeX$  can be instructed to load a local configuration file. This file can, for instance, be used to add strings to `\captions<lang>` to support local document classes. The user will be informed that this configuration file has been loaded. This macro is called by `\ldf@finish`.

`\substitutefontfamily` (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 .fd file will instruct  $\LaTeX$  to use a font from the second family when a font from the first family in the given encoding seems to be needed.

## 3.2 Skeleton

Here is the basic structure of an ldf file, with a language, a dialect and an attribute. Strings are best defined using the method explained in in sec. 3.7 (babel 3.9 and later).

```
\ProvidesLanguage{<language>}
  [0000/00/00 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@attribute{<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>}

```

### 3.3 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 L<sup>A</sup>T<sub>E</sub>X 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 behaviour 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.” [1, 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.

### 3.4 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>21</sup>.

`\babel@save` To save the current meaning of any control sequence, the macro `\babel@save` is provided. It takes one argument, `⟨csmame⟩`, 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.

### 3.5 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 behaviour is preserved for backward compatibility. If you are using `etoolbox`, by Philipp Lehman, consider using the tools provided by this package instead of `\addto`.

### 3.6 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 behaviour 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`.

<sup>21</sup>This mechanism was introduced by Bernd Raichle.

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

### 3.7 Encoding-dependent strings

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

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

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

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

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

The  $\langle language-list \rangle$  specifies which languages the block is intended for. A block is taken into account only if the `\CurrentOption` is listed here. Alternatively, you can define `\BabelLanguages` to a comma-separated list of languages to be defined (if undefined, `\StartBabelCommands` sets it to `\CurrentOption`). You may write `\CurrentOption` as the language, but this 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 take precedence (ie, it works much like `\providecommand`).

Encoding info is `charset=` followed by a `charset`, which if given sets how the strings should be traslated to the internal representation used by the engine, typically `utf8`, which is the only value supported currently (default is no traslations). 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 honoured (in a encoded way).

The  $\langle category \rangle$  is either captions, date or extras. You must stick to these three categories, even if no error is raised when using other name.<sup>22</sup> It may be empty, too, but in such a case using  $\backslash SetString$  is an error (but not  $\backslash SetCase$ ).

```

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

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

\EndBabelCommands

```

A real example is:

```

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

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

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

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

\StartBabelCommands{german,austrian}{date}
  \SetString\monthiiname{Februar}
  \SetString\monthiiname{M\"a}rz}
  \SetString\monthivname{April}
  \SetString\monthvname{Mai}
  \SetString\monthviname{Juni}
  \SetString\monthviiname{Juli}
  \SetString\monthviiname{August}
  \SetString\monthixname{September}
  \SetString\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  $\backslash \langle category \rangle \langle language \rangle$  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).

<sup>22</sup>In future releases further categories may be added.

**`\StartBabelCommands`** \*`{⟨language-list⟩}{⟨category⟩}[⟨selector⟩]`

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>23</sup>

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

**`\AfterBabelCommands`** `{⟨code⟩}`

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

**`\SetString`** `{⟨macro-name⟩}{⟨string⟩}`

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

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

**`\SetStringLoop`** `{⟨macro-name⟩}{⟨string-list⟩}`

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

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

`#1` is replaced by the roman numeral.

**`\SetCase`** `[⟨map-list⟩]{⟨toupper-code⟩}{⟨tolower-code⟩}`

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

For example, as T1 is the default case mapping in L<sup>A</sup>T<sub>E</sub>X, we could set for Turkish:

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

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

<sup>23</sup>This replaces in 3.9g a short-lived `\UseStrings` which has been removed because it did not work.

```

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

```

(Note the mapping for 0T1 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{\langle uccode \rangle}{\langle lccode \rangle}` 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{\langle uccode-from \rangle}{\langle uccode-to \rangle}{\langle step \rangle}{\langle lccode-from \rangle}` loops through the given uppercase codes, using the step, and assigns them the `lccode`, which is also increased (MM stands for *many-to-many*).
- `\BabelLowerM0{\langle uccode-from \rangle}{\langle uccode-to \rangle}{\langle step \rangle}{\langle lccode \rangle}` loops through the given uppercase codes, using the step, and assigns them the `lccode`, which is fixed (M0 stands for *many-to-one*).

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

```

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

```

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

## 4 Compatibility and changes

### 4.1 Compatibility with `german.sty`

The file `german.sty` has been one of the sources of inspiration for the `babel` system. Because of this I wanted to include `german.sty` in the `babel` system. To be able to do that I had to allow for one incompatibility: in the definition of the macro `\selectlanguage` in `german.sty` the argument is used as the *number* for an `\ifcase`. So in this case a call to `\selectlanguage` might look like `\selectlanguage{\german}`.

In the definition of the macro `\selectlanguage` in `babel.def` the argument is used as a part of other macronames, so a call to `\selectlanguage` now looks like `\selectlanguage{german}`. Notice the absence of the escape character. As of version 3.1a of `babel` both syntaxes are allowed.

All other features of the original `german.sty` have been copied into a new file, called `germanb.sty`<sup>24</sup>.

Although the `babel` system was developed to be used with L<sup>A</sup>T<sub>E</sub>X, some of the features implemented in the language definition files might be needed by plain T<sub>E</sub>X users. Care has been taken that all files in the system can be processed by plain T<sub>E</sub>X.

## 4.2 Compatibility with `ngerman.sty`

When used with the options `ngerman` or `naustrian`, `babel` will provide all features of the package `ngerman`. There is however one exception: The commands for special hyphenation of double consonants ("ff etc.) and ck ("ck), which are no longer required with the new German orthography, are undefined. With the `ngerman` package, however, these commands will generate appropriate warning messages only.

## 4.3 Compatibility with the french package

It has been reported to me that the package `french` by Bernard Gaulle (`gaulle@idris.fr`) works together with `babel`. On the other hand, it seems *not* to work well together with a lot of other packages. Therefore I have decided to no longer load `french.lfd` by default. Instead, when you want to use the package by Bernard Gaulle, you will have to request it specifically, by passing either `frenchle` or `frenchpro` as an option to `babel`.

## 4.4 Changes in `babel` version 3.9

Most of changes in version 3.9 are 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 behaviour for shorthands across languages). These changes are described in this manual in the correspondin place.

## 4.5 Changes in `babel` version 3.7

In `babel` version 3.7 a number of bugs that were found in version 3.6 are fixed. Also a number of changes and additions have occurred:

- Shorthands are expandable again. The disadvantage is that one has to type `'{}a` when the acute accent is used as a shorthand character. The advantage is that a number of other problems (such as the breaking of ligatures, etc.) have vanished.
- Two new commands, `\shorthandon` and `\shorthandoff` have been introduced to enable to temporarily switch off one or more shorthands.
- Support for typesetting Greek has been enhanced. Code from the `kdgreek` package (suggested by the author) was added and `\greeknumeral` has been added.
- Support for typesetting Basque is now available thanks to Juan Aguirregabiria.

---

<sup>24</sup>The 'b' is added to the name to distinguish the file from Partls' file.



- Support for typesetting Serbian with Latin script is now available thanks to Dejan Muhamedagić and Jankovic Slobodan.
- Support for typesetting Hebrew (and potential support for typesetting other right-to-left written languages) is now available thanks to Rama Porrat and Boris Lavva.
- Support for typesetting Bulgarian is now available thanks to Georgi Boshnakov.
- Support for typesetting Latin is now available, thanks to Claudio Beccari and Krzysztof Konrad Żelechowski.
- Support for typesetting North Sami is now available, thanks to Regnor Jernsletten.
- The options `canadian`, `canadien` and `acadien` have been added for Canadian English and French use.
- A language attribute has been added to the `\mark...` commands in order to make sure that a Greek header line comes out right on the last page before a language switch.
- Hyphenation pattern files are now read *inside a group*; therefore any changes a pattern file needs to make to lowercase codes, uppercase codes, and category codes are kept local to that group. If they are needed for the language, these changes will need to be repeated and stored in `\extras...`
- The concept of language attributes is introduced. It is intended to give the user some control over the features a language-definition file provides. Its first use is for the Greek language, where the user can choose the *πολυτονικό* (“Polutoniko” or multi-accented) Greek way of typesetting texts. These attributes will possibly find wider use in future releases.
- The environment `hyphenrules` is introduced.
- The syntax of the file `language.dat` has been extended to allow (optionally) specifying the font encoding to be used while processing the patterns file.
- The command `\providehyphenmins` should now be used in language definition files in order to be able to keep any settings provided by the pattern file.

## 4.6 Changes in babel version 3.6

In babel version 3.6 a number of bugs that were found in version 3.5 are fixed. Also a number of changes and additions have occurred:

- A new environment `otherlanguage*` is introduced. it only switches the ‘specials’, but leaves the ‘captions’ untouched.
- The shorthands are no longer fully expandable. Some problems could only be solved by peeking at the token following an active character. The advantage is that `'{ }a` works as expected for languages that have the ‘ active.
- Support for typesetting french texts is much enhanced; the file `français.ldf` is now replaced by `frenchb.ldf` which is maintained by Daniel Flipo.

- Support for typesetting the russian language is again available. The language definition file was originally developed by Olga Lapko from CyrTUG. The fonts needed to typeset the russian language are now part of the babel distribution. The support is not yet up to the level which is needed according to Olga, but this is a start.
- Support for typesetting greek texts is now also available. What is offered in this release is a first attempt; it will be enhanced later on by Yannis Haralambous.
- in babel 3.6j some hooks have been added for the development of support for Hebrew typesetting.
- Support for typesetting texts in Afrikaans (a variant of Dutch, spoken in South Africa) has been added to `dutch.ldf`.
- Support for typesetting Welsh texts is now available.
- A new command `\aliasshorthand` is introduced. It seems that in Poland various conventions are used to type the necessary Polish letters. It is now possible to use the character `/` as a shorthand character instead of the character `"`, by issuing the command `\aliasshorthand{"}{/}`.
- The shorthand mechanism now deals correctly with characters that are already active.
- Shorthand characters are made active at `\begin{document}`, not earlier. This is to prevent problems with other packages.
- A *preambleonly* command `\substitutefontfamily` has been added to create `.fd` files on the fly when the font families of the Latin text differ from the families used for the Cyrillic or Greek parts of the text.
- Three new commands `\LdfInit`, `\ldf@quit` and `\ldf@finish` are introduced that perform a number of standard tasks.
- In babel 3.6k the language Ukrainian has been added and the support for Russian typesetting has been adapted to the package 'cyrillic' to be released with the December 1998 release of L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>.

## 4.7 Changes in babel version 3.5

In babel version 3.5 a lot of changes have been made when compared with the previous release. Here is a list of the most important ones:

- the selection of the language is delayed until `\begin{document}`, which means you must add appropriate `\selectlanguage` commands if you include `\hyphenation` lists in the preamble of your document.
- babel now has a language environment and a new command `\foreignlanguage`;
- the way active characters are dealt with is completely changed. They are called 'shorthands'; one can have three levels of shorthands: on the user level, the language level, and on 'system level'. A consequence of the new way of handling active characters is that they are now written to auxiliary files 'verbatim';

- A language change now also writes information in the `.aux` file, as the change might also affect typesetting the table of contents. The consequence is that an `.aux` file generated by a `LaTeXformat` with `babel` preloaded gives errors when read with a `LaTeXformat` without `babel`; but I think this probably doesn't occur;
- `babel` is now compatible with the `inputenc` and `fontenc` packages;
- the language definition files now have a new extension, `ldf`;
- the syntax of the file `language.dat` is extended to be compatible with the `french` package by Bernard Gaulle;
- each language definition file looks for a configuration file which has the same name, but the extension `.cfg`. It can contain any valid `LaTeX` code.

## Part II

# The code

## 5 Identification and loading of required files

*Code documentation is still under revision.*

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

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

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

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

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

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

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

```
1 <<version=3.9m>>
2 <<date=2015/08/03>>
```

We define some basic macros which just make the code cleaner. `\bbl@add` is now used internally instead of `\addto` because of the unpredictable behaviour of the latter. Used in `babel.def` and in `babel.sty`, which means in `LaTeX` is executed twice, but we need them when defining options and `babel.def` cannot be load until options have been defined.

```
3 <<*Basic macros>> ≡
4 \def\bbl@add#1#2{%
5   \@ifundefined{\expandafter\@gobble\string#1}%
6     {\def#1{#2}}%
7     {\expandafter\def\expandafter#1\expandafter{#1#2}}
8 \def\bbl@csarg#1#2{\expandafter#1\csname bbl@#2\endcsname}%
9 \long\def\bbl@afterelse#1\else#2\fi{\fi#1}
```

```

10 \long\def\bb@afterfi#1\fi{\fi#1}
11 \def\bb@loop#1#2#3{\bb@loop#1{#3}#2,\@nnil,}
12 \def\bb@loopx#1#2{\expandafter\bb@loop\expandafter#1\expandafter{#2}}
13 \def\bb@loop#1#2#3,{%
14   \ifx\@nnil#3\relax\else
15     \def#1{#3}#2\bb@afterfi\bb@loop#1{#2}%
16   \fi}
17 \def\bb@for#1#2#3{\bb@loopx#1{#2}{\ifx#1\empty\else#3\fi}}
18 <</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  $\LaTeX$ .

```

19 <<*Make sure ProvidesFile is defined>> ≡
20 \ifx\ProvidesFile\@undefined
21   \def\ProvidesFile#1[#2 #3 #4]{%
22     \wlog{File: #1 #4 #3 <#2>}%
23     \let\ProvidesFile\@undefined}
24 \fi
25 <</Make sure ProvidesFile is defined>>

```

The following code is used in `babel.sty` and `babel.def`, and makes sure the current version of `switch.ldf` is used, if different from that in the format.

```

26 <<*Load switch if newer>> ≡
27 \def\bb@tempa{<<version>>}%
28 \ifx\bb@version\bb@tempa\else
29   \input switch.def\relax
30 \fi
31 <</Load switch if newer>>

```

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

```

32 <<*Load macros for plain if not LaTeX>> ≡
33 \ifx\AtBeginDocument\@undefined
34   \input plain.def\relax
35 \fi
36 <</Load macros for plain if not LaTeX>>

```

## 5.1 Multiple languages

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

```

37 <<*Define core switching macros>> ≡
38 \ifx\language\@undefined
39   \csname newcount\endcsname\language
40 \fi
41 <</Define core switching macros>>

```

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

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

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

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

```

42 <<*Define core switching macros>> ≡
43 \ifx\newlanguage\undefined
44   \csname newcount\endcsname\last@language
45   \def\addlanguage#1{%
46     \global\advance\last@language\@ne
47     \ifnum\last@language<\@ccclvi
48       \else
49         \errmessage{No room for a new \string\language!}%
50       \fi
51     \global\chardef#1\last@language
52     \wlog{\string#1 = \string\language\the\last@language}}
53 \else
54   \countdef\last@language=19
55   \def\addlanguage{\alloc@9\language\chardef\@ccclvi}
56 \fi
57 <</Define core switching macros>>

```

Identify each file that is produced from this source file.

```

58 <*driver&!user>
59 \ProvidesFile{babel.drv}[<<date>> <<version>>]
60 </driver&!user>
61 <*driver & user>
62 \ProvidesFile{user.drv}[<<date>> <<version>>]
63 </driver & user>

```

Now we make sure all required files are loaded. When the command `\AtBeginDocument` doesn't exist we assume that we are dealing with a plain-based format or  $\LaTeX 2.09$ . In that case the file `plain.def` is needed (which also defines `\AtBeginDocument`, and therefore it is not loaded twice). We need the first part when the format is created, and `\orig@dump` is used as a flag. Otherwise, we need to use the second part, so `\orig@dump` is not defined (`plain.def` undefines it). Check if the current version of `switch.def` has been previously loaded (mainly, `hyphen.cfg`). If not, load it now. We cannot load `babel.def` here because we first need to declare and process the package options.

## 6 The Package File ( $\LaTeX$ )

In order to make use of the features of  $\LaTeX 2_{\epsilon}$ , 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 behaviour 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.

## 6.1 base

The first option to be processed is base, which set the hyphenation patterns then resets `ver@babel.sty` so that L<sup>A</sup>T<sub>E</sub>X forgets about the first loading. After `switch.def` has been loaded (above) and `\AfterBabelLanguage` defined, exits.

```
64 (*package)
65 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
66 \ProvidesPackage{babel}[<<date>> <<version>> The Babel package]
67 \ifpackagewith{babel}{debug}
68 {\input switch.def\relax}
69 {<<Load switch if newer>>}
70 <<Basic macros>>
71 \def\AfterBabelLanguage#1{%
72   \global\expandafter\bbl@add\csname#1.ldf-h@k\endcsname}%
73 \ifpackagewith{babel}{base}{%
74   \DeclareOption*{\bbl@patterns{\CurrentOption}}%
75   \DeclareOption{base}{}%
76   \ProcessOptions
77   \global\expandafter\let\csname opt@babel.sty\endcsname\relax
78   \global\expandafter\let\csname ver@babel.sty\endcsname\relax
79   \global\let@ifl@ter@@@%
80   \def@ifl@ter#1#2#3#4#5{\global\let@ifl@ter@@@%
81   \endinput}{}}%
```

## 6.2 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`.

```
82 \bbl@csarg\let{tempa\expandafter}\csname opt@babel.sty\endcsname
83 \def\bbl@tempb#1.#2{%
84   #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%
85 \def\bbl@tempd#1.#2\@nnil{%
86   \ifx\@empty#2%
87     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
88   \else
89     \in@{=}{#1}\ifin@
90     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
91   \else
92     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
93     \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
94   \fi
95   \fi}
96 \let\bbl@tempc\@empty
97 \bbl@for\bbl@tempa\bbl@tempa{%
98   \expandafter\bbl@tempd\bbl@tempa.\@empty\@nnil}
99 \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.

```
100 \DeclareOption{KeepShorthandsActive}{}%
```

```

101 \DeclareOption{activeacute}{}
102 \DeclareOption{activegrave}{}
103 \DeclareOption{debug}{}
104 \DeclareOption{noconfigs}{}
105 \DeclareOption{showlanguages}{}
106 \DeclareOption{silent}{}
107 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
108 <<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.

```

109 \let\bbl@opt@shorthands\@nnil
110 \let\bbl@opt@config\@nnil
111 \let\bbl@opt@main\@nnil
112 \let\bbl@opt@headfoot\@nnil

```

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

```

113 \def\bbl@tempa#1=#2\bbl@tempa{%
114   \expandafter\ifx\csname bbl@opt@#1\endcsname\@nnil
115   \expandafter\edef\csname bbl@opt@#1\endcsname{#2}%
116   \else
117     \bbl@error{%
118       Bad option '#1=#2'. Either you have misspelled the\\%
119       key or there is a previous setting of '#1'}{%
120       Valid keys are 'shorthands', 'config', 'strings', 'main',\\%
121       'headfoot', 'safe', 'math'}
122   \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.

```

123 \let\bbl@language@opts\@empty
124 \DeclareOption*{%
125   \@expandtwoargs\in@{\string=}{\CurrentOption}%
126   \ifin@
127     \expandafter\bbl@tempa\CurrentOption\bbl@tempa
128   \else
129     \edef\bbl@language@opts{%
130       \ifx\bbl@language@opts\@empty\else\bbl@language@opts,\fi
131       \CurrentOption}%
132   \fi}

```

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

```

133 \ProcessOptions*

```

### 6.3 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@ifshorthands is always true, and it is always false if shorthands is empty. Also, some code makes sense only with shorthands=...

```

134 \def\bbl@sh@string#1{%
135   \ifx#1\@empty\else
136     \ifx#1t\string~%
137     \else\ifx#1c\string,%
138     \else\string#1%
139     \fi\fi
140   \expandafter\bbl@sh@string
141   \fi}
142 \ifx\bbl@opt@shorthands\@nnil
143   \def\bbl@ifshorthand#1#2#3{#2}%
144 \else\ifx\bbl@opt@shorthands\@empty
145   \def\bbl@ifshorthand#1#2#3{#3}%
146 \else

```

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

```

147 \def\bbl@ifshorthand#1{%
148   \@expandtwoargs\in@{\string#1}{\bbl@opt@shorthands}%
149   \ifin@
150     \expandafter\@firstoftwo
151   \else
152     \expandafter\@secondoftwo
153   \fi}

```

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

```

154 \edef\bbl@opt@shorthands{%
155   \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.

```

156 \bbl@ifshorthand{'}%
157   {\PassOptionsToPackage{activeacute}{babel}}{}
158 \bbl@ifshorthand{'}%
159   {\PassOptionsToPackage{activegrave}{babel}}{}
160 \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.

```

161 \ifx\bbl@opt@headfoot\@nnil\else
162   \g@addto@macro\@resetactivechars{%
163     \set@typeset@protect
164     \expandafter\select@language@x\expandafter{\bbl@opt@headfoot}%
165     \let\protect\noexpand}
166 \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.

```

167 \@ifundefined{bbl@opt@safe}{\def\bbl@opt@safe{BR}}{}
168 \ifx\bbl@opt@main\@nnil\else
169   \edef\bbl@language@opts{%
170     \ifx\bbl@language@opts\@empty\else\bbl@language@opts,\fi
171     \bbl@opt@main}
172 \fi

```

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.

```

173 \ifx\bbl@languages\@undefined\else

```



```

174 \begingroup
175   \catcode'\^^I=12
176   \@ifpackagewith{babel}{showlanguages}{%
177     \begingroup
178       \def\bbel@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}%
179       \wlog{<*languages>}%
180       \bbel@languages
181       \wlog{</languages>}%
182     \endgroup}{%}
183 \endgroup
184 \def\bbel@elt#1#2#3#4{%
185   \ifnum#2=\z@
186     \gdef\bbel@nulllanguage{#1}%
187     \def\bbel@elt##1##2##3##4{%
188       \fi}%
189   \bbel@languages
190 \fi

```

## 6.4 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).

```

191 \let\bbel@afterlang\relax
192 \let\BabelModifiers\relax
193 \let\bbel@loaded\@empty
194 \def\bbel@load@language#1{%
195   \InputIfFileExists{#1.ldf}%
196   {\edef\bbel@loaded{\CurrentOption
197     \ifx\bbel@loaded\@empty\else,\bbel@loaded\fi}%
198   \expandafter\let\expandafter\bbel@afterlang
199     \csname\CurrentOption.ldf-h@k\endcsname
200   \expandafter\let\expandafter\BabelModifiers
201     \csname bbl@mod@\CurrentOption\endcsname}%
202   {\bbel@error{%
203     Unknown option '\CurrentOption'. Either you misspelled it\\%
204     or the language definition file \CurrentOption.ldf was not found}}%
205   Valid options are: shorthands=, KeepShorthandsActive,\\%
206   activeacute, activegrave, noconfigs, safe=, main=, math=\\%
207   headfoot=, strings=, config=, hyphenmap=, or a language name.}}

```

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

```

208 \DeclareOption{acadian}{\bbel@load@language{frenchb}}
209 \DeclareOption{afrikaans}{\bbel@load@language{dutch}}
210 \DeclareOption{american}{\bbel@load@language{english}}
211 \DeclareOption{australian}{\bbel@load@language{english}}
212 \DeclareOption{bahasa}{\bbel@load@language{bahasai}}
213 \DeclareOption{bahasai}{\bbel@load@language{bahasai}}
214 \DeclareOption{bahasam}{\bbel@load@language{bahasam}}
215 \DeclareOption{brazil}{\bbel@load@language{portuges}}
216 \DeclareOption{brazilian}{\bbel@load@language{portuges}}
217 \DeclareOption{british}{\bbel@load@language{english}}
218 \DeclareOption{canadian}{\bbel@load@language{english}}
219 \DeclareOption{canadien}{\bbel@load@language{frenchb}}
220 \DeclareOption{français}{\bbel@load@language{frenchb}}

```

```

221 \DeclareOption{french}{\bbl@load@language{frenchb}}%
222 \DeclareOption{hebrew}{%
223   \input{rlbabel.def}%
224   \bbl@load@language{hebrew}}
225 \DeclareOption{hungarian}{\bbl@load@language{magyar}}
226 \DeclareOption{indon}{\bbl@load@language{bahasai}}
227 \DeclareOption{indonesian}{\bbl@load@language{bahasai}}
228 \DeclareOption{lowersorbian}{\bbl@load@language{lsorbian}}
229 \DeclareOption{malay}{\bbl@load@language{bahasam}}
230 \DeclareOption{meyalu}{\bbl@load@language{bahasam}}
231 \DeclareOption{melayu}{\bbl@load@language{bahasam}}
232 \DeclareOption{newzealand}{\bbl@load@language{english}}
233 \DeclareOption{nynorsk}{\bbl@load@language{norsk}}
234 \DeclareOption{polutonikogreek}{%
235   \bbl@load@language{greek}%
236   \languageattribute{greek}{polutoniko}}
237 \DeclareOption{portuguese}{\bbl@load@language{portuges}}
238 \DeclareOption{russian}{\bbl@load@language{russianb}}
239 \DeclareOption{UKenglish}{\bbl@load@language{english}}
240 \DeclareOption{ukrainian}{\bbl@load@language{ukraineb}}
241 \DeclareOption{uppersorbian}{\bbl@load@language{usorbian}}
242 \DeclareOption{USenglish}{\bbl@load@language{english}}

```

Another way to extend the list of ‘known’ options for babel is 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.

```

243 \ifx\bbl@opt@config\@nnil
244   \@ifpackagewith{babel}{noconfigs}}}%
245   {\InputIfFileExists{bblopts.cfg}%
246     {\typeout{*****^J%
247               * Local config file bblopts.cfg used^^J%
248               *}}}%
249   {}}}%
250 \else
251   \InputIfFileExists{\bbl@opt@config.cfg}%
252     {\typeout{*****^J%
253               * Local config file \bbl@opt@config.cfg used^^J%
254               *}}}%
255   {\bbl@error{%
256     Local config file ‘\bbl@opt@config.cfg’ not found}%
257     Perhaps you misspelled it.}}}%
258 \fi

```

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

```

259 \bbl@for\bbl@tempa\bbl@language@opts{%
260   \@ifundefined{ds@\bbl@tempa}%
261     {\edef\bbl@tempb{%
262       \noexpand\DeclareOption
263         {\bbl@tempa}%
264       {\noexpand\bbl@load@language{\bbl@tempa}}}}%

```

```

265 \bbl@tempb}%
266 \@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.

```

267 \bbl@for\bbl@tempa\@classoptionslist{%
268 \ifundefined{ds@\bbl@tempa}%
269 {\IfFileExists{\bbl@tempa.ldf}%
270 {\edef\bbl@tempb{%
271 \noexpand\DeclareOption
272 {\bbl@tempa}%
273 {\noexpand\bbl@load@language{\bbl@tempa}}}}%
274 \bbl@tempb}%
275 \@empty}%
276 \@empty}

```

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

```

277 \ifx\bbl@opt@main\@nnil\else
278 \expandafter
279 \let\expandafter\bbl@loadmain\csname ds@\bbl@opt@main\endcsname
280 \DeclareOption{\bbl@opt@main}{%}
281 \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 L<sup>A</sup>T<sub>E</sub>X processes before):

```

282 \def\AfterBabelLanguage#1{%
283 \bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang}{%}
284 \DeclareOption*{%}
285 \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.

```

286 \ifx\bbl@opt@main\@nnil
287 \edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
288 \let\bbl@tempc\@empty
289 \bbl@for\bbl@tempb\bbl@tempa{%
290 \expandtwoargs\in{,\bbl@tempb,}{,\bbl@loaded,}%
291 \ifin@edef\bbl@tempc{\bbl@tempb}\fi}
292 \def\bbl@tempa#1,#2\@nnil{\def\bbl@tempb{#1}}
293 \expandafter\bbl@tempa\bbl@loaded,\@nnil
294 \ifx\bbl@tempb\bbl@tempc\else
295 \bbl@warning{%
296 Last declared language option is '\bbl@tempc,\,%
297 but the last processed one was '\bbl@tempb'.\,%
298 The main language cannot be set as both a global\,%
299 and a package option. Use 'main=\bbl@tempc' as\,%
300 option. Reported}%
301 \fi
302 \else
303 \DeclareOption{\bbl@opt@main}{\bbl@loadmain}

```

```

304 \ExecuteOptions{\bbl@opt@main}
305 \DeclareOption*{}
306 \ProcessOptions*
307 \fi
308 \def\AfterBabelLanguage{%
309   \bbl@error
310   {Too late for \string\AfterBabelLanguage}%
311   {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.

```

312 \ifx\bbl@main@language\undefined
313   \bbl@error{%
314     You haven't specified a language option}{%
315     You need to specify a language, either as a global option\\%
316     or as an optional argument to the \string\usepackage\space
317     command;\\%
318     You shouldn't try to proceed from here, type x to quit.}
319 \fi
320 \end{package}

```

## 7 The kernel of Babel (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 L<sup>A</sup>T<sub>E</sub>X-specific stuff.

Because plain T<sub>E</sub>X users might want to use some of the features of the babel system too, care has to be taken that plain T<sub>E</sub>X can process the files. For this reason the current format will have to be checked in a number of places. Some of the code below is common to plain T<sub>E</sub>X and L<sup>A</sup>T<sub>E</sub>X, some of it is for the L<sup>A</sup>T<sub>E</sub>X case only.

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

### 7.1 Tools

`\bbl@engine` takes the following values: 0 is pdfT<sub>E</sub>X, 1 is `luatex`, and 2 is `xetex`. You may use it in your language style if necessary.

```

321 \*core)
322 \langle\langle Make sure ProvidesFile is defined \rangle\rangle
323 \ProvidesFile{babel.def}[\langle\langle date \rangle\rangle] [\langle\langle version \rangle\rangle] Babel common definitions]
324 \langle\langle Load macros for plain if not LaTeX \rangle\rangle
325 \langle\langle Load switch if newer \rangle\rangle
326 \ifx\bbl@languages\undefined
327   \openin1 = language.def

```

```

328 \ifeof1
329 \closein1
330 \message{I couldn't find the file language.def}
331 \else
332 \closein1
333 \begingroup
334 \def\addlanguage#1#2#3#4#5{%
335 \expandafter\ifx\csname lang@#1\endcsname\relax\else
336 \global\expandafter\let\csname l@#1\endcsname
337 \csname lang@#1\endcsname
338 \fi}%
339 \def\uselanguage#1{%
340 \input language.def
341 \endgroup
342 \fi
343 \chardef\l@english\z@
344 \fi
345 <<Basic macros>>
346 \def\bbl@csarg#1#2{\expandafter#1\csname bbl@#2\endcsname}%
347 \chardef\bbl@engine=%
348 \ifx\directlua\@undefined
349 \ifx\XeTeXinputencoding\@undefined
350 \z@
351 \else
352 \tw@
353 \fi
354 \else
355 \@ne
356 \fi

```

`\bbl@afterelse` Because the code that is used in the handling of active characters may need to  
`\bbl@afterfi` look ahead, we take extra care to ‘throw’ it over the `\else` and `\fi` parts of an  
`\if`-statement<sup>25</sup>. These macros will break if another `\if... \fi` statement appears  
in one of the arguments and it is not enclosed in braces.

```

357 \long\def\bbl@afterelse#1\else#2\fi{\fi#1}
358 \long\def\bbl@afterfi#1\fi{\fi#1}

```

`\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.

```

359 \def\addto#1#2{%
360 \ifx#1\@undefined
361 \def#1{#2}%
362 \else
363 \ifx#1\relax

```

<sup>25</sup>This code is based on code presented in TUGboat vol. 12, no2, June 1991 in “An expansion Power Lemma” by Sonja Maus.

```

364     \def#1{#2}%
365     \else
366     {\toks@\expandafter{#1#2}%
367     \xdef#1{\the\toks@}}%
368     \fi
369     \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.

```

370 \def\bb@withactive#1#2{%
371   \begingroup
372   \lccode'~=#2\relax
373   \lowercase{\endgroup#1~}}

```

`\bb@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 L<sup>A</sup>T<sub>E</sub>X 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 `\bb@redefine` which takes care of this. It creates a new control sequence, `\org@...`

```

374 \def\bb@redefine#1{%
375   \edef\bb@tempa{\expandafter\@gobble\string#1}%
376   \expandafter\let\csname org@\bb@tempa\endcsname#1%
377   \expandafter\def\csname\bb@tempa\endcsname}

```

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

```

378 \@onlypreamble\bb@redefine

```

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

```

379 \def\bb@redefine@long#1{%
380   \edef\bb@tempa{\expandafter\@gobble\string#1}%
381   \expandafter\let\csname org@\bb@tempa\endcsname#1%
382   \expandafter\long\expandafter\def\csname\bb@tempa\endcsname}
383 \@onlypreamble\bb@redefine@long

```

`\bb@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_`.

```

384 \def\bb@redefineroobust#1{%
385   \edef\bb@tempa{\expandafter\@gobble\string#1}%
386   \expandafter\ifx\csname\bb@tempa\space\endcsname\relax
387     \expandafter\let\csname org@\bb@tempa\endcsname#1%
388     \expandafter\edef\csname\bb@tempa\endcsname{\noexpand\protect
389     \expandafter\noexpand\csname\bb@tempa\space\endcsname}%
390   \else
391     \expandafter\let\csname org@\bb@tempa\expandafter\endcsname
392     \csname\bb@tempa\space\endcsname
393   \fi
394   \expandafter\def\csname\bb@tempa\space\endcsname}

```

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

```

395 \@onlypreamble\bb@redefineroobust

```

## 7.2 Hooks

Note they are loaded in `babel.def`. `switch.def` only provides a “hook” for hooks (with a default value which is a no-op, below). Admittedly, the current implementation is a somewhat simplistic and does very little to catch errors, but it is intended for developers, after all. `\bbl@usehooks` is the commands used by babel to execute hooks defined for an event.

```
396 \def\AddBabelHook#1#2{%
397   \ifundefined{bbl@hk@#1}{\EnableBabelHook{#1}}{%
398     \def\bbl@tempa##1,#2=##2,##3\@empty{\def\bbl@tempb{##2}}%
399     \expandafter\bbl@tempa\bbl@evargs,#2=,\@empty
400     \ifundefined{bbl@ev@#1@#2}%
401       {\bbl@csarg\bbl@add{ev@#2}{\bbl@elt{#1}}}%
402       \bbl@csarg\newcommand}%
403       {\bbl@csarg\let{ev@#1@#2}\relax
404       \bbl@csarg\newcommand}%
405       {ev@#1@#2}[\bbl@tempb]}
406 \def\EnableBabelHook#1{\bbl@csarg\let{hk@#1}\@firstofone}
407 \def\DisableBabelHook#1{\bbl@csarg\let{hk@#1}\@gobble}
408 \def\bbl@usehooks#1#2{%
409   \def\bbl@elt##1{%
410     \@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1}#2}}%
411   \@nameuse{bbl@ev@#1}}
```

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

```
412 \def\bbl@evargs{,% don't delete the comma
413   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
414   adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
415   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
416   hyphenation=2,initiateactive=3,afterreset=0}
```

`\babelensure` The user command just parses the optional argument and creates a new macro named `\bbl@ens@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@ens@language` contains

`\bbl@ensure{<include>}{<exclude>}{<fontenc>}`, which in turn loops over the macros names in `\bbl@ensured`, 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.

`\bbl@ensured` is the list of macros supposed to be “ensured”.

```
417 \newcommand\babelensure[2][]{%
418   \AddBabelHook{babel-ensure}{afterextras}{%
419     \ifcase\bbl@select@type
420       \@nameuse{bbl@e@languagename}%
421       \fi}%
422   \beginngroup
423     \let\bbl@ens@include\@empty
```

```

424 \let\bb@ens@exclude\@empty
425 \def\bb@ens@fontenc{\relax}%
426 \def\bb@tempb##1{%
427   \ifx\@empty##1\else\noexpand##1\expandafter\bb@tempb\fi}%
428 \edef\bb@tempa{\bb@tempb##1\@empty}%
429 \def\bb@tempb##1=##2\@{\@namedef{bb@ens@##1}{##2}}%
430 \bb@for\bb@tempa\bb@tempa{\expandafter\bb@tempb\bb@tempa\@}%
431 \def\bb@tempc{\bb@ensure}%
432 \expandafter\bb@add\expandafter\bb@tempc\expandafter{%
433   \expandafter{\bb@ens@include}}%
434 \expandafter\bb@add\expandafter\bb@tempc\expandafter{%
435   \expandafter{\bb@ens@exclude}}%
436 \toks@\expandafter{\bb@tempc}%
437 \edef\x{%
438   \endgroup
439   \noexpand\@namedef{bb@e@#2}{\the\toks@{\bb@ens@fontenc}}}%
440 \x}
441 \def\bb@ensure#1#2#3{%
442 \def\bb@tempb##1{% elt for \bb@ensured list
443   \ifx##1\@empty\else
444     \in{##1}{#2}%
445     \ifin\else
446       \toks@\expandafter{##1}%
447       \edef\bb@tempa{%
448         \noexpand\DeclareRobustCommand
449         \bb@csarg\noexpand{ensure@\language}\[1]{%
450           \noexpand\foreignlanguage{\language}%
451           {\ifx\relax#3\else
452             \noexpand\fontencoding{#3}\noexpand\selectfont
453             \fi
454             #####1}}}%
455         \bb@tempa
456         \edef##1{%
457           \bb@csarg\noexpand{ensure@\language}%
458           {\the\toks@}}
459         \fi
460         \expandafter\bb@tempb
461         \fi}%
462 \expandafter\bb@tempb\bb@ensured\@empty
463 \def\bb@tempa##1{% elt for include list
464   \ifx##1\@empty\else
465     \bb@csarg\in{ensure@\language\expandafter}\expandafter{##1}%
466     \ifin\else
467       \bb@tempb##1\@empty
468       \fi
469     \expandafter\bb@tempa
470     \fi}%
471 \bb@tempa#1\@empty}
472 \def\bb@ensured{%
473 \prefacename\refname\abstractname\bibname\chaptername\appendixname
474 \contentsname\listfigurename\listtablename\indexname\figurename
475 \tablename\partname\enclname\ccname\headtoname\pagename\seename
476 \alsoname\proofname\glossaryname\today}

```



### 7.3 Setting up language files

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

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

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

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

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

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

```
477 \def\LdfInit#1#2{%
478   \chardef\atcatcode=\catcode'\@
479   \catcode'\@=11\relax
480   \chardef\eqcatcode=\catcode'\=
481   \catcode'\==12\relax
482   \expandafter\if\expandafter\@backslashchar
483     \expandafter\@car\string#2\@nil
484     \ifx#2\@undefined\else
485       \ldf@quit{#1}%
486     \fi
487 \else
488   \expandafter\ifx\csname#2\endcsname\relax\else
489     \ldf@quit{#1}%
490   \fi
491 \fi
492 \let\bbl@screset\@empty
493 \let\BabelStrings\bbl@opt@strings
494 \let\BabelOptions\@empty
495 \let\BabelLanguages\relax
496 \ifx\originalTeX\@undefined
497   \let\originalTeX\@empty
498 \else
499   \originalTeX
500 \fi}
```

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

```
501 \def\ldf@quit#1{%
502   \expandafter\main@language\expandafter{#1}%
503   \catcode'\@=\atcatcode \let\atcatcode\relax
504   \catcode'\==\eqcatcode \let\eqcatcode\relax
505   \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.

```
506 \def\ldf@finish#1{%
507   \loadlocalcfg{#1}%
508   \bbl@afterlang
509   \let\bbl@afterlang\relax
510   \let\BabelModifiers\relax
511   \let\bbl@screset\relax
512   \expandafter\main@language\expandafter{#1}%
513   \catcode'\@=\atcatcode \let\atcatcode\relax
514   \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 L<sup>A</sup>T<sub>E</sub>X.

```
515 \@onlypreamble\LdfInit
516 \@onlypreamble\ldf@quit
517 \@onlypreamble\ldf@finish
```

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

```
518 \def\main@language#1{%
519   \def\bbl@main@language{#1}%
520   \let\languagename\bbl@main@language
521   \bbl@patterns{\languagename}}
```

We also have to make sure that some code gets executed at the beginning of the document.

```
522 \AtBeginDocument{%
523   \expandafter\selectlanguage\expandafter{\bbl@main@language}}
```

## 7.4 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 L<sup>A</sup>T<sub>E</sub>X is used). To keep all changes local, we begin a new group. Then we redefine the macros `\do` and `\@makeother` to add themselves and the given character without expansion. To add the character to the macros, we expand the original macros with the additional character inside the redefinition of the macros. Because `\@sanitize` can be undefined, we put the definition inside a conditional.

```
524 \def\bbl@add@special#1{%
525   \begingroup
526   \def\do{\noexpand\do\noexpand}%
527   \def\@makeother{\noexpand\@makeother\noexpand}%
528   \edef\x{\endgroup
529     \def\noexpand\dospecials{\dospecials\do#1}%
530     \expandafter\ifx\csname @sanitize\endcsname\relax \else
531       \def\noexpand\@sanitize{\@sanitize\@makeother#1}%
532     \fi}%
533   \x}
```

The macro `\x` contains at this moment the following:

```
\endgroup\def\dospecials{old contents \do<char>}
```

If `\@sanitize` is defined, it contains an additional definition of this macro. The last thing we have to do, is the expansion of `\x`. Then `\endgroup` is executed, which restores the old meaning of `\x`, `\do` and `\@makeoother`. After the group is closed, the new definition of `\dospecials` (and `\@sanitize`) is assigned.

`\bbl@remove@special` The companion of the former macro is `\bbl@remove@special`. It is used to remove a character from the set macros `\dospecials` and `\@sanitize`.

To keep all changes local, we begin a new group. Then we define a help macro `\x`, which expands to empty if the characters match, otherwise it expands to its nonexpandable input. Because  $\TeX$  inserts a `\relax`, if the corresponding `\else` or `\fi` is scanned before the comparison is evaluated, we provide a ‘stop sign’ which should expand to nothing.

With the help of this macro we define `\do` and `\make@other`.

The rest of the work is similar to `\bbl@add@special`.

```
534 \def\bbl@remove@special#1{%
535   \begingroup
536   \def\x##1##2{\ifnum'#1='##2\noexpand\@empty
537             \else\noexpand##1\noexpand##2\fi}%
538   \def\do{\x\do}%
539   \def\@makeoother{\x\@makeoother}%
540 \edef\x{\endgroup
541   \def\noexpand\dospecials{\dospecials}%
542   \expandafter\ifx\csname @sanitize\endcsname\relax \else
543     \def\noexpand\@sanitize{\@sanitize}%
544   \fi}%
545 \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).

```
546 \def\bbl@active@def#1#2#3#4{%
547   \namedef{#3#1}{%
548     \expandafter\ifx\csname#2@sh@#1\endcsname\relax
```

```

549     \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
550     \else
551     \bbl@afterfi\csname#2@sh@#1@endcsname
552     \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.

```

553 \long\@namedef{#3@arg#1}##1{%
554   \expandafter\ifx\csname#2@sh@#1@string##1@endcsname\relax
555     \bbl@afterelse\csname#4#1@endcsname##1%
556   \else
557     \bbl@afterfi\csname#2@sh@#1@string##1@endcsname
558   \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.

```

559 \def\initiate@active@char#1{%
560   \expandafter\ifx\csname active@char\string#1@endcsname\relax
561     \bbl@withactive
562     {\expandafter\@initiate@active@char\expandafter}#1\string#1#1%
563   \fi}

```

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`).

```

564 \def\@initiate@active@char#1#2#3{%
565   \expandafter\edef\csname bbl@oricat@#2@endcsname{%
566     \catcode'#2=\the\catcode'#2\relax}%
567   \ifx#1\undefined
568     \expandafter\edef\csname bbl@oridef@#2@endcsname{%
569       \let\noexpand#1\noexpand\@undefined}%
570   \else
571     \expandafter\let\csname bbl@oridef@#2@endcsname#1%
572     \expandafter\edef\csname bbl@oridef@#2@endcsname{%
573       \let\noexpand#1%
574       \expandafter\noexpand\csname bbl@oridef@#2@endcsname}%
575   \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).

```

576   \ifx#1#3\relax
577     \expandafter\let\csname normal@char#2@endcsname#3%
578   \else
579     \bbl@info{Making #2 an active character}%
580     \ifnum\mathcode'#2="8000
581       \@namedef{normal@char#2}{%
582         \textormath{#3}{\csname bbl@oridef@#2@endcsname}}%
583     \else
584       \@namedef{normal@char#2}{#3}%
585     \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).

```

586 \bbl@restoreactive{#2}%
587 \AtBeginDocument{%
588   \catcode'#2\active
589   \if@filesw
590     \immediate\write\@mainaux{\catcode'\string#2\active}%
591   \fi}%
592 \expandafter\bbl@add@special\csname#2\endcsname
593 \catcode'#2\active
594 \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⟩`).

```

595 \let\bbl@tempa\@firstoftwo
596 \if\string^#2%
597 \def\bbl@tempa{\noexpand\textormath}%
598 \else
599 \ifx\bbl@mathnormal\@undefined\else
600 \let\bbl@tempa\bbl@mathnormal
601 \fi
602 \fi
603 \expandafter\edef\csname active@char#2\endcsname{%
604 \bbl@tempa
605   {\noexpand\if@safe@actives
606     \noexpand\expandafter
607     \expandafter\noexpand\csname normal@char#2\endcsname
608     \noexpand\else
609     \noexpand\expandafter
610     \expandafter\noexpand\csname bbl@doactive#2\endcsname
611     \noexpand\fi}%
612   {\expandafter\noexpand\csname normal@char#2\endcsname}}%
613 \bbl@csarg\edef{doactive#2}{%
614 \expandafter\noexpand\csname user@active#2\endcsname}%

```

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

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

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

```

615 \bbl@csarg\edef{active@#2}{%
616 \noexpand\active@prefix\noexpand#1%
617 \expandafter\noexpand\csname active@char#2\endcsname}%
618 \bbl@csarg\edef{normal@#2}{%
619 \noexpand\active@prefix\noexpand#1%
620 \expandafter\noexpand\csname normal@char#2\endcsname}%

```

```
621 \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.

```
622 \bbl@active@def#2\user@group{user@active}{language@active}%
623 \bbl@active@def#2\language@group{language@active}{system@active}%
624 \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.

```
625 \expandafter\edef\csname\user@group @sh@#2@@\endcsname
626   {\expandafter\noexpand\csname normal@char#2\endcsname}%
627 \expandafter\edef\csname\user@group @sh@#2@\string\protect@\endcsname
628   {\expandafter\noexpand\csname user@active#2\endcsname}%
```

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

```
629 \if\string'#2%
630   \let\prim@s\bbl@prim@s
631   \let\active@math@prime#1%
632 \fi
633 \bbl@usehooks{initiateactive}{{#1}{#2}{#3}}
```

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

```
634 <<{*More package options}>> ≡
635 \DeclareOption{math=active}{}
636 \DeclareOption{math=normal}{\def\bbl@mathnormal{\noexpand\textormath}}
637 <</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.

```
638 \ifpackagewith{babel}{KeepShorthandsActive}%
639   {\let\bbl@restoreactive@gobble}%
640   {\def\bbl@restoreactive#1{%
641     \edef\bbl@tempa{%
642       \noexpand\AfterBabelLanguage\noexpand\CurrentOption
643       {\catcode'#1=\the\catcode'#1\relax}%
644       \noexpand\AtEndOfPackage{\catcode'#1=\the\catcode'#1\relax}}%
645     \bbl@tempa}%
646   \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.

```
647 \def\bbl@sh@select#1#2{%
648   \expandafter\ifx\csname#1@sh@#2@sel\endcsname\relax
649     \bbl@afterelse\bbl@scndcs
650   \else
651     \bbl@afterfi\csname#1@sh@#2@sel\endcsname
652   \fi}
```

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

```
653 \def\active@prefix#1{%
654   \ifx\protect\@typeset@protect
655   \else
```

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

```
656   \ifx\protect\@unexpandable@protect
657     \noexpand#1%
658   \else
659     \protect#1%
660   \fi
661   \expandafter\@gobble
662   \fi}
```

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

```
663 \newif\if@safe@actives
664 \@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.

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

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

```
666 \def\bbl@activate#1{%
667   \bbl@withactive{\expandafter\let\expandafter}#1%
668   \csname bbl@active@\string#1\endcsname}
669 \def\bbl@deactivate#1{%
670   \bbl@withactive{\expandafter\let\expandafter}#1%
671   \csname bbl@normal@\string#1\endcsname}
```

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

```
672 \def\bbl@firstcs#1#2{\csname#1\endcsname}
673 \def\bbl@scndcs#1#2{\csname#2\endcsname}
```

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

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

```

674 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil}
675 \def\@decl@short#1#2#3\@nil#4{%
676   \def\bbl@tempa{#3}%
677   \ifx\bbl@tempa\@empty
678     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs
679     \@ifundefined{#1@sh@\string#2@}{}%
680     {\def\bbl@tempa{#4}%
681      \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa
682      \else
683        \bbl@info
684          {Redefining #1 shorthand \string#2\%
685           in language \CurrentOption}%
686        \fi}%
687     \@namedef{#1@sh@\string#2@}{#4}%
688   \else
689     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs
690     \@ifundefined{#1@sh@\string#2@\string#3@}{}%
691     {\def\bbl@tempa{#4}%
692      \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
693      \else
694        \bbl@info
695          {Redefining #1 shorthand \string#2\string#3\%
696           in language \CurrentOption}%
697        \fi}%
698     \@namedef{#1@sh@\string#2@\string#3@}{#4}%
699   \fi}

```

`\textormath` Some of the shorthands that will be declared by the language definition files have to be usable in both text and mathmode. To achieve this the helper macro `\textormath` is provided.

```

700 \def\textormath{%
701   \ifmmode
702     \expandafter\@secondoftwo
703   \else
704     \expandafter\@firstoftwo
705   \fi}

```

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

```

706 \def\user@group{user}
707 \def\language@group{english}
708 \def\system@group{system}

```

`\usesshorthands` This is the user level command to tell L<sup>A</sup>T<sub>E</sub>X 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.

```

709 \def\useshorthands{%
710   \ifstar\bb@usesh@s{\bb@usesh@x{}}
711 \def\bb@usesh@s#1{%
712   \bb@usesh@x
713   {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bb@activate{#1}}}%
714   {#1}}
715 \def\bb@usesh@x#1#2{%
716   \bb@ifshorthand{#2}%
717   {\def\user@group{user}%
718     \initiate@active@char{#2}%
719     #1%
720     \bb@activate{#2}}%
721   {\bb@error
722     {Cannot declare a shorthand turned off (\string#2)}
723     {Sorry, but you cannot use shorthands which have been\\%
724       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 `\bb@set@user@generic`); we make also sure `{}` and `\protect` are taken into account in this new top level.

```

725 \def\user@language@group{user@\language@group}
726 \def\bb@set@user@generic#1#2{%
727   \ifundefined{user@generic@active#1}%
728     {\bb@active@def#1\user@language@group{user@active}{user@generic@active}%
729     \bb@active@def#1\user@group{user@generic@active}{language@active}%
730     \expandafter\edef\csname#2@sh@#1@\endcsname{%
731       \expandafter\noexpand\csname normal@char#1\endcsname}%
732     \expandafter\edef\csname#2@sh@#1@\string\protect\endcsname{%
733       \expandafter\noexpand\csname user@active#1\endcsname}}%
734   \@empty}
735 \newcommand\defineshorthand[3][user]{%
736   \edef\bb@tempa{\zap@space#1 \@empty}%
737   \bb@for\bb@tempb\bb@tempa{%
738     \if*\expandafter\@car\bb@tempb\@nil
739       \edef\bb@tempb{user\expandafter\@gobble\bb@tempb}%
740       \@expandtwoargs
741       \bb@set@user@generic{\expandafter\string\@car#2\@nil}\bb@tempb
742     \fi
743     \declare@shorthand{\bb@tempb}{#2}{#3}}}
```

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

```

744 \def\languageshorthands#1{\def\language@group{#1}}
```

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

```

745 \def\aliasshorthand#1#2{%
746   \bb@ifshorthand{#2}%
```

```

747     {\expandafter\ifx\csname active@char\string#2\endcsname\relax
748       \ifx\document\@notprerr
749         \@notshorthand{#2}%
750       \else
751         \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`.

```

752         \expandafter\let\csname active@char\string#2\expandafter\endcsname
753         \csname active@char\string#1\endcsname
754     \expandafter\let\csname normal@char\string#2\expandafter\endcsname
755     \csname normal@char\string#1\endcsname
756     \bbl@activate{#2}%
757     \fi
758 \fi}%
759 {\bbl@error
760   {Cannot declare a shorthand turned off (\string#2)}
761   {Sorry, but you cannot use shorthands which have been\\%
762     turned off in the package options}}}

```

`\@notshorthand`

```

763 \def\@notshorthand#1{%
764   \bbl@error{%
765     The character '\string #1' should be made a shorthand character;\\%
766     add the command \string\usesshorthands\string{#1\string} to
767     the preamble.\\%
768     I will ignore your instruction}%
769   {You may proceed, but expect unexpected results}}

```

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

```

770 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
771 \DeclareRobustCommand*\shorthandoff{%
772   \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
773 \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.

```

774 \def\bbl@switch@sh#1#2{%
775   \ifx#2\@nnil\else
776     \@ifundefined{bbl@active@\string#2}%
777     {\bbl@error
778       {I cannot switch '\string#2' on or off--not a shorthand}%
779       {This character is not a shorthand. Maybe you made\\%
780         a typing mistake? I will ignore your instruction}}%
781     {\ifcase#1%

```

```

782     \catcode'#212\relax
783     \or
784     \catcode'#2\active
785     \or
786     \csname bbl@oricat@string#2\endcsname
787     \csname bbl@oridef@string#2\endcsname
788     \fi}%
789     \bbl@afterfi\bbl@switch@sh#1%
790 \fi}

```

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

```

791 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
792 \def\bbl@putsh#1{%
793   \ifundefined{bbl@active@string#1}%
794     {\bbl@putsh@i#1@empty\@nnil}%
795     {\csname bbl@active@string#1\endcsname}}
796 \def\bbl@putsh@i#1#2\@nnil{%
797   \csname\languagename @sh@string#1@%
798     \ifx\@empty#2\else\string#2@\fi\endcsname}
799 \ifx\bbl@opt@shorthands\@nnil\else
800   \let\bbl@s@initiate@active@char\initiate@active@char
801   \def\initiate@active@char#1{%
802     \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
803   \let\bbl@s@switch@sh\bbl@switch@sh
804   \def\bbl@switch@sh#1#2{%
805     \ifx#2\@nnil\else
806       \bbl@afterfi
807       \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
808     \fi}
809   \let\bbl@s@activate\bbl@activate
810   \def\bbl@activate#1{%
811     \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
812   \let\bbl@s@deactivate\bbl@deactivate
813   \def\bbl@deactivate#1{%
814     \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
815 \fi

```

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

```

816 \def\bbl@prim@s{%
817   \prime\futurelet\@let@token\bbl@pr@m@s}
818 \def\bbl@if@primes#1#2{%
819   \ifx#1\@let@token
820     \expandafter\@firstoftwo
821   \else\ifx#2\@let@token
822     \bbl@afterelse\expandafter\@firstoftwo
823   \else
824     \bbl@afterfi\expandafter\@secondoftwo
825   \fi\fi}
826 \begingroup
827 \catcode'\^=7 \catcode'\*=\active \lccode'\*='\^
828 \catcode'\'=12 \catcode'\"=\active \lccode'\"='\^
829 \lowercase{%

```

```

830 \gdef\bbI@pr@ms{%
831 \bbI@if@primes" '%
832 \pr@@@s
833 {\bbI@if@primes*^\pr@@@t\egroup}}
834 \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).

```

835 \initiate@active@char{~}
836 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
837 \bbI@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.

```

838 \expandafter\def\csname OT1dqpos\endcsname{127}
839 \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

```

840 \ifx\f@encoding\@undefined
841 \def\f@encoding{OT1}
842 \fi

```

## 7.5 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.

```

843 \newcommand\languageattribute[2]{%
844 \def\bbI@tempc{#1}%
845 \bbI@fixname\bbI@tempc
846 \bbI@iflanguage\bbI@tempc{%
847 \bbI@loopx\bbI@attr{#2}{%

```

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

```

848 \ifx\bbI@known@attrs\@undefined
849 \in@false
850 \else

```

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

```

851 \@expandtwoargs\in{,\bbI@tempc-\bbI@attr,}{,\bbI@known@attrs,}%
852 \fi

```

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

```

853     \ifin@
854     \bbl@warning{%
855         You have more than once selected the attribute '\bbl@attr'\%
856         for language #1}%
857     \else

```

When we end up here the attribute is not selected before. So, we add it to the list of selected attributes and execute the associated T<sub>E</sub>X-code.

```

858     \edef\bbl@tempa{%
859         \noexpand\bbl@add@list
860         \noexpand\bbl@known@attribs{\bbl@tempc-\bbl@attr}}%
861     \bbl@tempa
862     \edef\bbl@tempa{\bbl@tempc-\bbl@attr}%
863     \expandafter\bbl@ifknown@attrib\expandafter{\bbl@tempa}\bbl@attributes%
864     {\csname\bbl@tempc @attr@\bbl@attr\endcsname}%
865     {\@attrerr{\bbl@tempc}{\bbl@attr}}%
866     \fi}}

```

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

```

867 \@onlypreamble\languageattribute

```

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

```

868 \newcommand*{\@attrerr}[2]{%
869     \bbl@error
870     {The attribute #2 is unknown for language #1.}%
871     {Your command will be ignored, type <return> to proceed}}

```

`\bbl@declare@attribute` 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}`.

```

872 \def\bbl@declare@attribute#1#2#3{%
873     \@expandtwoargs\in@{,#2,}{,\BabelModifiers,}%
874     \ifin@
875     \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
876     \fi
877     \bbl@add@list\bbl@attributes{#1-#2}%
878     \expandafter\def\csname#1@attr@#2\endcsname{#3}}

```

`\bbl@ifattributeset` This internal macro has 4 arguments. It can be used to interpret T<sub>E</sub>X 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.

```

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

```

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

```

880     \ifx\bbl@known@attribs\undefined
881     \in@false
882     \else

```

The we need to check the list of known attributes.

```
883   \expandtwoargs\in@{,#1-#2,}{,\bbl@known@attribs,}%  
884   \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`'.

```
885   \ifin@  
886     \bbl@afterelse#3%  
887   \else  
888     \bbl@afterfi#4%  
889   \fi  
890 }
```

`\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

```
891 \def\bbl@add@list#1#2{%  
892   \ifx#1\@undefined  
893     \def#1{#2}%  
894   \else  
895     \ifx#1\@empty  
896       \def#1{#2}%  
897     \else  
898       \edef#1{#1,#2}%  
899   \fi  
900 \fi  
901 }
```

`\bbl@ifknown@attrib` 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.

```
902 \def\bbl@ifknown@attrib#1#2{%  
    We first assume the attribute is unknown.  
903   \let\bbl@tempa\@secondoftwo  
    Then we loop over the list of known attributes, trying to find a match.  
904   \bbl@loopx\bbl@tempb{#2}{%  
905     \expandafter\in@\expandafter{\expandafter,\bbl@tempb,}{,#1,}%  
906     \ifin@  
    When a match is found the definition of \bbl@tempa is changed.  
907     \let\bbl@tempa\@firstoftwo  
908   \else  
909     \fi}%  
    Finally we execute \bbl@tempa.  
910   \bbl@tempa  
911 }
```

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

```
912 \def\bbl@clear@attribs{%  
913   \ifx\bbl@attributes\@undefined\else  
914     \bbl@loopx\bbl@tempa{\bbl@attributes}{%  
915       \expandafter\bbl@clear@attrib\bbl@tempa.  
916     }%
```

```

917 \let\bbl@attributes\@undefined
918 \fi
919 }
920 \def\bbl@clear@ttrib#1-#2.{%
921 \expandafter\let\csname#1@attr@#2\endcsname\@undefined}
922 \AtBeginDocument{\bbl@clear@ttribs}

```

## 7.6 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` 923 `\def\babel@beginsave{\babel@savecnt\z@}`

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

```

924 \newcount\babel@savecnt
925 \babel@beginsave

```

`\babel@save` The macro `\babel@save<csname>` saves the current meaning of the control sequence `<csname>` to `\originalTeX`<sup>26</sup>. To do this, we let the current meaning to a temporary control sequence, the restore commands are appended to `\originalTeX` and the counter is incremented.

```

926 \def\babel@save#1{%
927 \expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax
928 \begingroup
929 \toks@\expandafter{\originalTeX\let#1=}%
930 \edef\x{\endgroup
931 \def\noexpand\originalTeX{\the\toks@ \expandafter\noexpand
932 \csname babel@number\babel@savecnt\endcsname\relax}}%
933 \x
934 \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.

```

935 \def\babel@savevariable#1{\begingroup
936 \toks@\expandafter{\originalTeX #1=}%
937 \edef\x{\endgroup
938 \def\noexpand\originalTeX{\the\toks@ \the#1\relax}}%
939 \x}

```

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

```

940 \def\bbl@frenchspacing{%
941 \ifnum\the\sfcode'\.=\@m
942 \let\bbl@nonfrenchspacing\relax

```

<sup>26</sup>`\originalTeX` has to be expandable, i. e. you shouldn't let it to `\relax`.

```

943 \else
944   \frenchspacing
945   \let\bbbl@nonfrenchspacing\nonfrenchspacing
946 \fi}
947 \let\bbbl@nonfrenchspacing\nonfrenchspacing

```

## 7.7 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.

```

948 \def\babeltags#1{%
949   \edef\bbbl@tempa{\zap@space#1 \@empty}%
950   \def\bbbl@tempb##1=##2\@{%
951     \edef\bbbl@tempc{%
952       \noexpand\newcommand
953       \expandafter\noexpand\csname ##1\endcsname{%
954         \noexpand\protect
955         \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
956       \noexpand\newcommand
957       \expandafter\noexpand\csname text##1\endcsname{%
958         \noexpand\foreignlanguage{##2}}
959     \bbbl@tempc}%
960   \bbbl@for\bbbl@tempa\bbbl@tempa{%
961     \expandafter\bbbl@tempb\bbbl@tempa\@{}}

```

## 7.8 Hyphens

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

```

962 \@onlypreamble\babelhyphenation
963 \AtEndOfPackage{%
964   \newcommand\babelhyphenation[2][\@empty]{%
965     \ifx\bbbl@hyphenation@relax
966       \let\bbbl@hyphenation@\@empty
967     \fi
968     \ifx\bbbl@hyphlist\@empty\else
969       \bbbl@warning{%
970         You must not intermingle \string\selectlanguage\space and\%
971         \string\babelhyphenation\space or some exceptions will not\%
972         be taken into account. Reported}%
973     \fi
974     \ifx\@empty#1%
975       \protected@edef\bbbl@hyphenation@{\bbbl@hyphenation@\space#2}%
976     \else
977       \edef\bbbl@tempb{\zap@space#1 \@empty}%
978       \bbbl@for\bbbl@tempa\bbbl@tempb{%
979         \bbbl@fixname\bbbl@tempa
980         \bbbl@iflanguage\bbbl@tempa{%
981           \bbbl@csarg\protected@edef{hyphenation@\bbbl@tempa}{%
982             \@ifundefined{bbbl@hyphenation@\bbbl@tempa}%
983             \@empty

```



```

984         {\csname bbl@hyphenation@\bbl@tempa\endcsname\space}%
985         #2}}}%
986     \fi}}

```

`\bbl@allowhyphens` This macro makes hyphenation possible. Basically its definition is nothing more than `\nobreak \hskip 0pt plus 0pt`<sup>27</sup>.

```

987 \def\bbl@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}
988 \def\bbl@t@one{T1}
989 \def\allowhyphens{\ifx\cf@encoding\bbl@t@one\else\bbl@allowhyphens\fi}

```

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

```

990 \newcommand\babelnullhyphen{\char\hyphenchar\font}
991 \def\babelhyphen{\active@prefix\babelhyphen\bbl@hyphen}
992 \def\bbl@hyphen{%
993   \@ifstar{\bbl@hyphen@i @}{\bbl@hyphen@i \@empty}}
994 \def\bbl@hyphen@i#1#2{%
995   \@ifundefined{bbl@hy@#1#2\@empty}%
996   {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%
997   {\csname bbl@hy@#1#2\@empty\endcsname}}

```

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

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

```

998 \def\bbl@usehyphen#1{%
999   \leavevmode
1000   \ifdim\lastskip>\z@\mbox{#1}\nobreak\else\nobreak#1\fi
1001   \hskip\z@skip}
1002 \def\bbl@@usehyphen#1{%
1003   \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

The following macro inserts the hyphen char.

```

1004 \def\bbl@hyphenchar{%
1005   \ifnum\hyphenchar\font=\m@ne
1006     \babelnullhyphen
1007   \else
1008     \char\hyphenchar\font
1009   \fi}

```

Finally, we define the hyphen “types”. Their names will not change, so you may use them in `ldf`’s.

```

1010 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1011 \def\bbl@hy@soft{\bbl@@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1012 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
1013 \def\bbl@hy@hard{\bbl@@usehyphen\bbl@hyphenchar}
1014 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}\nobreak}}
1015 \def\bbl@hy@@nobreak{\mbox{\bbl@hyphenchar}}

```

<sup>27</sup>`TEX` begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.

```

1016 \def\bb@hy@repeat{%
1017   \bb@usehyphen{%
1018     \discretionary{\bb@hyphenchar}{\bb@hyphenchar}{\bb@hyphenchar}%
1019     \nobreak}}
1020 \def\bb@hy@@repeat{%
1021   \bb@usehyphen{%
1022     \discretionary{\bb@hyphenchar}{\bb@hyphenchar}{\bb@hyphenchar}}}
1023 \def\bb@hy@empty{\hskip\z@skip}
1024 \def\bb@hy@@empty{\discretionary{}{}{}}

```

`\bb@disc` For some languages the macro `\bb@disc` is used to ease the insertion of discretionaries for letters that behave ‘abnormally’ at a breakpoint.

```

1025 \def\bb@disc#1#2{\nobreak\discretionary{#2-}{#1}\bb@allowhyphens}

```

## 7.9 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.

```

1026 \def\bb@tglobal#1{\global\let#1#1}
1027 \def\bb@recatcode#1{%
1028   \@tempcnta="7F
1029   \def\bb@tempa{%
1030     \ifnum\@tempcnta>"FF\else
1031       \catcode\@tempcnta=#1\relax
1032       \advance\@tempcnta\@ne
1033       \expandafter\bb@tempa
1034     \fi}%
1035   \bb@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 `\bb@uclc`. The parser is restarted inside `\(lang)\bb@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\bb@tolower\@empty\bb@toupper\@empty
%

```

and starts over (and similarly when lowercasing).

```

1036 \@ifpackagewith{babel}{nocase}%
1037   {\let\bb@patchuclc\relax}%
1038   {\def\bb@patchuclc{%
1039     \global\let\bb@patchuclc\relax
1040     \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bb@uclc}}%
1041     \gdef\bb@uclc##1{%
1042       \let\bb@encoded\bb@encoded@uclc
1043       \@ifundefined{\language @\bb@uclc}% and resumes it

```

```

1044     {##1}%
1045     {\let\bbbl@tempa##1\relax % Used by LANG@bbbl@uclc
1046     \csname\languagename @bbbl@uclc\endcsname}%
1047     {\bbbl@tolower\@empty}{\bbbl@toupper\@empty}}%
1048     \gdef\bbbl@tolower{\csname\languagename @bbbl@lc\endcsname}%
1049     \gdef\bbbl@toupper{\csname\languagename @bbbl@uc\endcsname}}}}
1050 <<(*More package options)>> ≡
1051 \DeclareOption{nocase}{}
1052 <</More package options>>

```

The following package options control the behaviour of \SetString.

```

1053 <<(*More package options)>> ≡
1054 \let\bbbl@opt@strings\@nnil % accept strings=value
1055 \DeclareOption{strings}{\def\bbbl@opt@strings\BabelStringsDefault}}
1056 \DeclareOption{strings=encoded}{\let\bbbl@opt@strings\relax}
1057 \def\BabelStringsDefault{generic}
1058 <</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.

```

1059 \@onlypreamble\StartBabelCommands
1060 \def\StartBabelCommands{%
1061   \begingroup
1062   \bbbl@recatcode{11}%
1063   <<Macros local to BabelCommands>>
1064   \def\bbbl@provstring##1##2{%
1065     \providecommand##1{##2}%
1066     \bbbl@toggle##1}%
1067   \global\let\bbbl@scafter\@empty
1068   \let\StartBabelCommands\bbbl@startcmds
1069   \ifx\BabelLanguages\relax
1070     \let\BabelLanguages\CurrentOption
1071   \fi
1072   \begingroup
1073   \let\bbbl@screset\@nnil % local flag - disable 1st stopcommands
1074   \StartBabelCommands}
1075 \def\bbbl@startcmds{%
1076   \ifx\bbbl@screset\@nnil\else
1077     \bbbl@usehooks{stopcommands}{}%
1078   \fi
1079   \endgroup
1080   \begingroup
1081   \@ifstar
1082     {\ifx\bbbl@opt@strings\@nnil
1083       \let\bbbl@opt@strings\BabelStringsDefault
1084     \fi
1085     \bbbl@startcmds@i}%
1086   \bbbl@startcmds@i}
1087 \def\bbbl@startcmds@i#1#2{%
1088   \edef\bbbl@L{\zap@space#1 \@empty}%
1089   \edef\bbbl@G{\zap@space#2 \@empty}%
1090   \bbbl@startcmds@ii}

```

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

Select the behaviour 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.

```

1091 \newcommand\bbstartcmds@ii[1][\@empty]{%
1092   \let\SetString@gobbletwo
1093   \let\bbstringdef@gobbletwo
1094   \let\AfterBabelCommands@gobble
1095   \ifx\@empty#1%
1096     \def\bb@sc@label{generic}%
1097     \def\bb@encstring##1##2{%
1098       \ProvideTextCommandDefault##1{##2}%
1099       \bb@tglobal##1%
1100       \expandafter\bb@tglobal\csname\string?\string##1\endcsname}%
1101     \let\bb@sctest\in@true
1102   \else
1103     \let\bb@sc@charset\space % <- zapped below
1104     \let\bb@sc@fontenc\space % <- " "
1105     \def\bb@tempa##1=##2\@nil{%
1106       \bb@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
1107     \bb@for\bb@tempb{label=#1}{\expandafter\bb@tempa\bb@tempb\@nil}%
1108     \def\bb@tempa##1 ##2{% space -> comma
1109       ##1%
1110       \ifx\@empty##2\else\ifx,##1,\else,\fi\bb@afterfi\bb@tempa##2\fi}%
1111     \edef\bb@sc@fontenc{\expandafter\bb@tempa\bb@sc@fontenc\@empty}%
1112     \edef\bb@sc@label{\expandafter\zap@space\bb@sc@label\@empty}%
1113     \edef\bb@sc@charset{\expandafter\zap@space\bb@sc@charset\@empty}%
1114     \def\bb@encstring##1##2{%
1115       \bb@for\bb@tempc\bb@sc@fontenc{%
1116         \ifundefined{T\bb@tempc}%
1117           {}%
1118           {\ProvideTextCommand##1\bb@tempc{##2}%
1119             \bb@tglobal##1%
1120             \expandafter
1121             \bb@tglobal\csname\bb@tempc\string##1\endcsname}}}%
1122     \def\bb@sctest{%
1123       \@expandtwoargs
1124       \in{\, \bb@opt@strings,}{, \bb@sc@label, \bb@sc@fontenc,}%
1125     \fi
1126     \ifx\bb@opt@strings\@nnil % ie, no strings key -> defaults
1127     \else\ifx\bb@opt@strings\relax % ie, strings=encoded
1128       \let\AfterBabelCommands\bb@aftercmds
1129       \let\SetString\bb@setstring
1130       \let\bbstringdef\bb@encstring
1131     \else % ie, strings=value
1132     \bb@sctest
1133     \ifin@
1134       \let\AfterBabelCommands\bb@aftercmds
1135       \let\SetString\bb@setstring

```

```

1136 \let\bbL@stringdef\bbL@provstring
1137 \fi\fi\fi
1138 \bbL@scswitch
1139 \ifx\bbL@G\@empty
1140 \def\SetString##1##2{%
1141 \bbL@error{Missing group for string \string##1}%
1142 {You must assign strings to some category, typically\%
1143 captions or extras, but you set none}}%
1144 \fi
1145 \ifx\@empty#1%
1146 \@expandtwoargs
1147 \bbL@usehooks{defaultcommands}{}%
1148 \else
1149 \@expandtwoargs
1150 \bbL@usehooks{encodedcommands}{\bbL@sc@charset}{\bbL@sc@fontenc}}%
1151 \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) .

```

1152 \def\bbL@forlang#1#2{%
1153 \bbL@for#1\bbL@L{%
1154 \@expandtwoargs\in@{,#1,}{,\BabelLanguages,}%
1155 \ifin#2\relax\fi}}
1156 \def\bbL@scswitch{%
1157 \bbL@forlang\bbL@tempa{%
1158 \ifx\bbL@G\@empty\else
1159 \ifx\SetString@gobbletwo\else
1160 \edef\bbL@GL{\bbL@G\bbL@tempa}%
1161 \@expandtwoargs\in@{,\bbL@GL,}{,\bbL@screset,}%
1162 \ifin@\else
1163 \global\expandafter\let\csname\bbL@GL\endcsname\@undefined
1164 \xdef\bbL@screset{\bbL@screset,\bbL@GL}%
1165 \fi
1166 \fi
1167 \fi}}
1168 \AtEndOfPackage{%
1169 \def\bbL@forlang#1#2{\bbL@for#1\bbL@L{\@ifundefined{date#1}{#2}}}%
1170 \let\bbL@scswitch\relax}
1171 \@onlypreamble\EndBabelCommands
1172 \def\EndBabelCommands{%
1173 \bbL@usehooks{stopcommands}{}%
1174 \endgroup
1175 \endgroup
1176 \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.

```

1177 \def\bb@setstring#1#2{%
1178   \bb@forlang\bb@tempa{%
1179     \edef\bb@LC{\bb@tempa\expandafter@gobble\string#1}%
1180     \@ifundefined{\bb@LC}% eg, \germanchaptername
1181     {\global\expandafter
1182      \bb@add\csname\bb@G\bb@tempa\expandafter\endcsname\expandafter
1183      {\expandafter\bb@scset\expandafter#1\csname\bb@LC\endcsname}}%
1184     }%
1185   \def\BabelString{#2}%
1186   \bb@usehooks{stringprocess}{}%
1187   \expandafter\bb@stringdef
1188   \csname\bb@LC\expandafter\endcsname\expandafter{\BabelString}}

```

Now, some additional stuff to be used when encoded strings are used. Captions then include `\bb@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`.

```

1189 \ifx\bb@opt@strings\relax
1190   \def\bb@scset#1#2{\def#1{\bb@encoded#2}}
1191   \bb@patchuclc
1192   \let\bb@encoded\relax
1193   \def\bb@encoded@uclc#1{%
1194     \@inmathwarn#1%
1195     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
1196       \expandafter\ifx\csname ?\string#1\endcsname\relax
1197         \TextSymbolUnavailable#1%
1198       \else
1199         \csname ?\string#1\endcsname
1200       \fi
1201     \else
1202       \csname\cf@encoding\string#1\endcsname
1203     \fi}
1204 \else
1205   \def\bb@scset#1#2{\def#1{#2}}
1206 \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).

```

1207 <<(*Macros local to BabelCommands)>> ≡
1208 \def\SetStringLoop###1##2{%
1209   \def\bb@templ####1{\expandafter\noexpand\csname##1\endcsname}%
1210   \count@z@
1211   \bb@loop\bb@tempa{##2}{%
1212     \advance\count@\@ne
1213     \toks@\expandafter{\bb@tempa}%
1214     \edef\bb@tempb{%
1215       \bb@templ{\romannumeral\count@}{\the\toks@}%
1216       \count@=\the\count@\relax}%
1217     \expandafter\SetString\bb@tempb}}%
1218 <</Macros local to BabelCommands>>

```

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

```
1219 \def\bbl@aftercmds#1{%
1220   \toks@\expandafter{\bbl@scafter#1}%
1221   \xdef\bbl@scafter{\the\toks@}}
```

**Case mapping** The command `\SetCase` provides a way to change the behaviour of `\MakeUppercase` and `\MakeLowercase`. `\bbl@tempa` is set by the patched `\@uclclist` to the parsing command.

```
1222 <<(*Macros local to BabelCommands)>> ≡
1223   \newcommand\SetCase[3][]{%
1224     \bbl@patchuclc
1225     \bbl@forlang\bbl@tempa{%
1226       \expandafter\bbl@encstring
1227       \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
1228       \expandafter\bbl@encstring
1229       \csname\bbl@tempa @bbl@uc\endcsname{##2}%
1230       \expandafter\bbl@encstring
1231       \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
1232 <</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.

```
1233 <<(*Macros local to BabelCommands)>> ≡
1234   \newcommand\SetHyphenMap[1]{%
1235     \bbl@forlang\bbl@tempa{%
1236       \expandafter\bbl@stringdef
1237       \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}%
1238 <</Macros local to BabelCommands>>
```

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

```
1239 \newcommand\BabelLower[2]{% one to one.
1240   \ifnum\lccode#1=#2\else
1241     \babel@savevariable{\lccode#1}%
1242     \lccode#1=#2\relax
1243   \fi}
1244 \newcommand\BabelLowerMM[4]{% many-to-many
1245   \@tempcnta=#1\relax
1246   \@tempcntb=#4\relax
1247   \def\bbl@tempa{%
1248     \ifnum\@tempcnta>#2\else
1249       \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
1250       \advance\@tempcnta#3\relax
1251       \advance\@tempcntb#3\relax
1252       \expandafter\bbl@tempa
1253     \fi}%
1254   \bbl@tempa}
1255 \newcommand\BabelLowerM0[4]{% many-to-one
1256   \@tempcnta=#1\relax
1257   \def\bbl@tempa{%
1258     \ifnum\@tempcnta>#2\else
1259       \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
1260       \advance\@tempcnta#3
1261       \expandafter\bbl@tempa
1262     \fi}%
```

```
1263 \bbl@tempa}
```

The following package options control the behaviour of hyphenation mapping.

```
1264 <<{*More package options}>> ≡
1265 \DeclareOption{hyphenmap=off}{\chardef\bbl@hymapopt\z@}
1266 \DeclareOption{hyphenmap=first}{\chardef\bbl@hymapopt\@ne}
1267 \DeclareOption{hyphenmap=select}{\chardef\bbl@hymapopt\tw@}
1268 \DeclareOption{hyphenmap=other}{\chardef\bbl@hymapopt\thr@@}
1269 \DeclareOption{hyphenmap=other*}{\chardef\bbl@hymapopt4\relax}
1270 <</More package options>>
```

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

```
1271 \AtEndOfPackage{%
1272 \ifx\bbl@hymapopt\undefined
1273 \expandtwoargs\in@{,}{\bbl@language@opts}%
1274 \chardef\bbl@hymapopt\ifin@4\else\@ne\fi
1275 \fi}
```

## 7.10 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.

```
1276 \def\set@low@box#1{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}%
1277 \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@%
1278 \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.

```
1279 \def\save@sf@q#1{\leavevmode
1280 \begingroup
1281 \edef\@SF{\spacefactor\the\spacefactor}#1\@SF
1282 \endgroup}
```

## 7.11 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.

### 7.11.1 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.

```
1283 \ProvideTextCommand{\quotedblbase}{OT1}{%
1284 \save@sf@q{\set@low@box{\textquotedblright\}}%
1285 \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.

```
1286 \ProvideTextCommandDefault{\quotedblbase}{%
1287 \UseTextSymbol{OT1}{\quotedblbase}}
```

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

```
1288 \ProvideTextCommand{\quotesinglbase}{OT1}{%
1289 \save@sf@q{\set@low@box{\textquoteright\}}%
1290 \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.

```
1291 \ProvideTextCommandDefault{\quotesinglbase}{%
1292 \UseTextSymbol{OT1}{\quotesinglbase}}
```

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

```
\guillemotright 1293 \ProvideTextCommand{\guillemotleft}{OT1}{%
1294 \ifmmode
1295 \ll
1296 \else
1297 \save@sf@q{\nobreak
1298 \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bbl@allowhyphens}%
1299 \fi}
1300 \ProvideTextCommand{\guillemotright}{OT1}{%
1301 \ifmmode
1302 \gg
1303 \else
1304 \save@sf@q{\nobreak
1305 \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bbl@allowhyphens}%
1306 \fi}
```

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

```
1307 \ProvideTextCommandDefault{\guillemotleft}{%
1308 \UseTextSymbol{OT1}{\guillemotleft}}
1309 \ProvideTextCommandDefault{\guillemotright}{%
1310 \UseTextSymbol{OT1}{\guillemotright}}
```

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

```
\guilsinglright 1311 \ProvideTextCommand{\guilsinglleft}{OT1}{%
1312 \ifmmode
1313 <%
1314 \else
1315 \save@sf@q{\nobreak
1316 \raise.2ex\hbox{$\scriptscriptstyle<$}\bbl@allowhyphens}%
1317 \fi}
1318 \ProvideTextCommand{\guilsinglright}{OT1}{%
1319 \ifmmode
1320 >%
1321 \else
1322 \save@sf@q{\nobreak
1323 \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
1324 \fi}
```

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

```
1325 \ProvideTextCommandDefault{\guilsinglleft}{%
1326 \UseTextSymbol{OT1}{\guilsinglleft}}
1327 \ProvideTextCommandDefault{\guilsinglright}{%
1328 \UseTextSymbol{OT1}{\guilsinglright}}
```

### 7.11.2 Letters

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

```
1329 \DeclareTextCommand{\ij}{OT1}{%}
```

```

1330 i\kern-0.02em\bbl@allowhyphens j}
1331 \DeclareTextCommand{\IJ}{OT1}{%
1332 I\kern-0.02em\bbl@allowhyphens J}
1333 \DeclareTextCommand{\ij}{T1}{\char188}
1334 \DeclareTextCommand{\IJ}{T1}{\char156}

```

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

```

1335 \ProvideTextCommandDefault{\ij}{%
1336 \UseTextSymbol{OT1}{\ij}}
1337 \ProvideTextCommandDefault{\IJ}{%
1338 \UseTextSymbol{OT1}{\IJ}}

```

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

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

```

1339 \def\crrtic@{\hrule height0.1ex width0.3em}
1340 \def\crttic@{\hrule height0.1ex width0.33em}
1341 \def\ddj@{%
1342 \setbox0\hbox{d}\dimen@=\ht0
1343 \advance\dimen@1ex
1344 \dimen@.45\dimen@
1345 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1346 \advance\dimen@ii.5ex
1347 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
1348 \def\DDJ@{%
1349 \setbox0\hbox{D}\dimen@=.55\ht0
1350 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1351 \advance\dimen@ii.15ex % correction for the dash position
1352 \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
1353 \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
1354 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
1355 %
1356 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
1357 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}

```

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

```

1358 \ProvideTextCommandDefault{\dj}{%
1359 \UseTextSymbol{OT1}{\dj}}
1360 \ProvideTextCommandDefault{\DJ}{%
1361 \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.

```

1362 \DeclareTextCommand{\SS}{OT1}{SS}
1363 \ProvideTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}

```

### 7.11.3 Shorthands for quotation marks

Shorthands are provided for a number of different quotation marks, which make them usable both outside and inside mathmode.

`\glq` The ‘german’ single quotes.

```

\grq 1364 \ProvideTextCommand{\glq}{OT1}{%

```

```

1365 \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}
1366 \ProvideTextCommand{\glq}{T1}{%
1367 \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}
1368 \ProvideTextCommandDefault{\glq}{\UseTextSymbol{OT1}\glq}

```

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

```

1369 \ProvideTextCommand{\grq}{T1}{%
1370 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
1371 \ProvideTextCommand{\grq}{OT1}{%
1372 \save@sf@q{\kern-.0125em%
1373 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
1374 \kern.07em\relax}}
1375 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}

```

`\glqq` The ‘german’ double quotes.

```

\grqq 1376 \ProvideTextCommand{\glqq}{OT1}{%
1377 \textormath{\quotedblbase}{\mbox{\quotedblbase}}}
1378 \ProvideTextCommand{\glqq}{T1}{%
1379 \textormath{\quotedblbase}{\mbox{\quotedblbase}}}
1380 \ProvideTextCommandDefault{\glqq}{\UseTextSymbol{OT1}\glqq}

```

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

```

1381 \ProvideTextCommand{\grqq}{T1}{%
1382 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1383 \ProvideTextCommand{\grqq}{OT1}{%
1384 \save@sf@q{\kern-.07em%
1385 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
1386 \kern.07em\relax}}
1387 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}

```

`\flq` The ‘french’ single guillemets.

```

\frq 1388 \ProvideTextCommand{\flq}{OT1}{%
1389 \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
1390 \ProvideTextCommand{\flq}{T1}{%
1391 \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
1392 \ProvideTextCommandDefault{\flq}{\UseTextSymbol{OT1}\flq}

1393 \ProvideTextCommand{\frq}{OT1}{%
1394 \textormath{\guilsinglright}{\mbox{\guilsinglright}}}
1395 \ProvideTextCommand{\frq}{T1}{%
1396 \textormath{\guilsinglright}{\mbox{\guilsinglright}}}
1397 \ProvideTextCommandDefault{\frq}{\UseTextSymbol{OT1}\frq}

```

`\flqq` The ‘french’ double guillemets.

```

\frqq 1398 \ProvideTextCommand{\flqq}{OT1}{%
1399 \textormath{\guillemotleft}{\mbox{\guillemotleft}}}
1400 \ProvideTextCommand{\flqq}{T1}{%
1401 \textormath{\guillemotleft}{\mbox{\guillemotleft}}}
1402 \ProvideTextCommandDefault{\flqq}{\UseTextSymbol{OT1}\flqq}

1403 \ProvideTextCommand{\frqq}{OT1}{%
1404 \textormath{\guillemotright}{\mbox{\guillemotright}}}
1405 \ProvideTextCommand{\frqq}{T1}{%
1406 \textormath{\guillemotright}{\mbox{\guillemotright}}}
1407 \ProvideTextCommandDefault{\frqq}{\UseTextSymbol{OT1}\frqq}

```

### 7.11.4 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).

```
1408 \def\umlauthigh{%
1409   \def\bb@umlauta##1{\leavevmode\bgroup%
1410     \expandafter\accent\csname\fontencoding dqpos\endcsname
1411       ##1\bb@allowhyphens\egroup}%
1412   \let\bb@umlaute\bb@umlauta}
1413 \def\umlautlow{%
1414   \def\bb@umlauta{\protect\lower@umlaut}}
1415 \def\umlaute\low{%
1416   \def\bb@umlaute{\protect\lower@umlaut}}
1417 \umlauthigh
```

`\lower@umlaut` The command `\lower@umlaut` is used to position the `\` closer the the letter.  
We want the umlaut character lowered, nearer to the letter. To do this we need an extra `\dimen` register.

```
1418 \expandafter\ifx\csname U@D\endcsname\relax
1419   \csname newdimen\endcsname\U@D
1420 \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.

```
1421 \def\lower@umlaut#1{%
1422   \leavevmode\bgroup
1423   \U@D lex%
1424   {\setbox\z@\hbox{%
1425     \expandafter\char\csname\fontencoding dqpos\endcsname}%
1426     \dimen@ -.45ex\advance\dimen@\ht\z@
1427     \ifdim lex<\dimen@ \fontdimen5\font\dimen@ \fi}%
1428   \expandafter\accent\csname\fontencoding dqpos\endcsname
1429   \fontdimen5\font\U@D #1%
1430 \egroup}
```

For all vowels we declare `\` to be a composite command which uses `\bb@umlauta` or `\bb@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).

```
1431 \AtBeginDocument{%
1432   \DeclareTextCompositeCommand{"}{OT1}{a}{\bbl@umlauta{a}}%
1433   \DeclareTextCompositeCommand{"}{OT1}{e}{\bbl@umlaute{e}}%
1434   \DeclareTextCompositeCommand{"}{OT1}{i}{\bbl@umlaute{i}}%
1435   \DeclareTextCompositeCommand{"}{OT1}{\i}{\bbl@umlaute{\i}}%
1436   \DeclareTextCompositeCommand{"}{OT1}{o}{\bbl@umlauta{o}}%
1437   \DeclareTextCompositeCommand{"}{OT1}{u}{\bbl@umlauta{u}}%
1438   \DeclareTextCompositeCommand{"}{OT1}{A}{\bbl@umlauta{A}}%
1439   \DeclareTextCompositeCommand{"}{OT1}{E}{\bbl@umlaute{E}}%
1440   \DeclareTextCompositeCommand{"}{OT1}{I}{\bbl@umlaute{I}}%
1441   \DeclareTextCompositeCommand{"}{OT1}{O}{\bbl@umlauta{O}}%
1442   \DeclareTextCompositeCommand{"}{OT1}{U}{\bbl@umlauta{U}}%
1443 }
```

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

```
1444 \ifx\l@english\undefined
1445   \chardef\l@english\z@
1446 \fi
1447 \main@language{english}
```

Now we load definition files for engines.

```
1448 \ifcase\bbl@engine\or
1449   \input luababel.def
1450 \or
1451   \input xebabel.def
1452 \fi
```

## 8 The kernel of Babel (only L<sup>A</sup>T<sub>E</sub>X)

### 8.1 The redefinition of the style commands

The rest of the code in this file can only be processed by L<sup>A</sup>T<sub>E</sub>X, so we check the current format. If it is plain T<sub>E</sub>X, 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 T<sub>E</sub>X 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`.

```
1453 {\def\format{lplain}
1454 \ifx\fmtname\format
1455 \else
1456   \def\format{LaTeX2e}
1457   \ifx\fmtname\format
1458   \else
1459     \aftergroup\endinput
1460   \fi
1461 \fi}
```

### 8.2 Cross referencing macros

The L<sup>A</sup>T<sub>E</sub>X 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 T<sub>E</sub>Xbook [1] (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.

```
1462 %\bbl@redefine\newlabel#1#2{%
1463 % \@safe@activestrue\org@newlabel{#1}{#2}\@safe@activesfalse}
```

`\@newl@bel` We need to change the definition of the L<sup>A</sup>T<sub>E</sub>X-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.

```
1464 <<(*More package options)>> ≡
1465 \DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
1466 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
1467 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
1468 <</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.

```
1469 \ifx\bbl@opt@safe\@empty\else
1470 \def\@newl@bel#1#2#3{%
1471   {\@safe@activestrue
1472     \ifundefined{#1@#2}%
1473       \relax
1474       {\gdef\@multiplelabels{%
1475         \@latex@warning@no@line{There were multiply-defined labels}}%
1476         \@latex@warning@no@line{Label ‘#2’ multiply defined}}%
1477       \global\@namedef{#1@#2}{#3}}
```

`\@testdef` An internal L<sup>A</sup>T<sub>E</sub>X 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 L<sup>A</sup>T<sub>E</sub>X keeps reporting that the labels may have changed.

```
1478 \CheckCommand*\@testdef[3]{%
1479   \def\reserved@a{#3}%
1480   \expandafter\ifx\csname#1@#2\endcsname\reserved@a
1481   \else
1482     \@tempwattrue
1483   \fi}
```

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

```
1484 \def\@testdef#1#2#3{%
1485   \@safe@activestrue
```

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

```
1486   \expandafter\let\expandafter\bbl@tempa\csname #1#2\endcsname
```

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

```
1487   \def\bbl@tempb{#3}%
1488   \@safe@activesfalse
```

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

```
1489   \ifx\bbl@tempa\relax
1490   \else
1491     \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
1492   \fi
```

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

```
1493   \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
```

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

```
1494   \ifx\bbl@tempa\bbl@tempb
1495   \else
1496     \@tempswatrue
1497   \fi}
1498 \fi
```

`\ref` `\pageref` The same holds for the macro `\ref` that references a label and `\pageref` to reference a 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.

```
1499 \@expandtwoargs\in@{R}\bbl@opt@safe
1500 \ifin@
1501   \bbl@redefineroobust\ref#1{%
1502     \@safe@activestrue\org@ref{#1}\@safe@activesfalse}
1503   \bbl@redefineroobust\pageref#1{%
1504     \@safe@activestrue\org@pageref{#1}\@safe@activesfalse}
1505 \else
1506   \let\org@ref\ref
1507   \let\org@pageref\pageref
1508 \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.

```
1509 \@expandtwoargs\in@{B}\bbl@opt@safe
1510 \ifin@
1511   \bbl@redefine\@citex[#1]#2{%
1512     \@safe@activestrue\edef\@tempa{#2}\@safe@activesfalse
1513     \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.

```
1514 \AtBeginDocument{%
1515   \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.)

```
1516   \def\@citex[#1][#2]#3{%
1517     \@safe@activestruedef\@tempa{#3}\@safe@activesfalse
1518     \org@citex[#1][#2]{\@tempa}}%
1519   }{}}
```

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

```
1520 \AtBeginDocument{%
1521   \ifpackageloaded{cite}{%
1522     \def\@citex[#1]#2{%
1523       \@safe@activestruedef\org@citex[#1][#2]\@safe@activesfalse}%
1524     }{}}
```

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

```
1525 \bbl@redefine\nocite#1{%
1526   \@safe@activestruedef\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@activestruedef` 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.

```
1527 \bbl@redefine\bibcite{%
```

We call `\bbl@cite@choice` to select the proper definition for `\bibcite`. This new definition is then activated.

```
1528   \bbl@cite@choice
1529   \bibcite}
```

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

```
1530 \def\bbl@bibcite#1#2{%
1531   \org@bibcite{#1}{\@safe@activesfalse#2}}
```

`\bbl@cite@choice` The macro `\bbl@cite@choice` determines which definition of `\bibcite` is needed.

```
1532 \def\bbl@cite@choice{%
```

First we give `\bibcite` its default definition.

```
1533   \global\let\bibcite\bbl@bibcite
```

Then, when `natbib` is loaded we restore the original definition of `\bibcite`.

```
1534   \ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{}}
```



For cite we do the same.

```
1535 \ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{}%
```

Make sure this only happens once.

```
1536 \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.

```
1537 \AtBeginDocument{\bbl@cite@choice}
```

`\@bibitem` One of the two internal L<sup>A</sup>T<sub>E</sub>X macros called by \bibitem that write the citation label on the .aux file.

```
1538 \bbl@redefine\@bibitem#1{%
```

```
1539 \@safe@activestruer\org@@bibitem{#1}\@safe@activesfalse}
```

```
1540 \else
```

```
1541 \let\org@nocite\nocite
```

```
1542 \let\org@@citex\@citex
```

```
1543 \let\org@bibcite\bibcite
```

```
1544 \let\org@@bibitem\@bibitem
```

```
1545 \fi
```

### 8.3 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.

```
1546 \bbl@redefine\markright#1{%
```

First of all we temporarily store the language switching command, using an expanded definition in order to get the current value of \language.

```
1547 \edef\bbl@tempb{\noexpand\protect
```

```
1548 \noexpand\foreignlanguage{\language}}%
```

Then, we check whether the argument is empty; if it is, we just make sure the scratch token register is empty.

```
1549 \def\bbl@arg{#1}%
```

```
1550 \ifx\bbl@arg\@empty
```

```
1551 \toks@{}%
```

```
1552 \else
```

Next, we store the argument to \markright in the scratch token register, together with the expansion of \bbl@tempb (containing the language switching command) as defined before. This way these commands will not be expanded by using \edef later on, 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@activestruer is in effect.

```
1553 \expandafter\toks@\expandafter{%
```

```
1554 \bbl@tempb{\protect\bbl@restore@actives#1}}%
```

```
1555 \fi
```

Then we define a temporary control sequence using \edef.

```
1556 \edef\bbl@tempa{%
```

When `\bbl@tempa` is executed, only `\language` will be expanded, because of the way the token register was filled.

```
1557 \noexpand\org@markright{\the\toks@}%
1558 \bbl@tempa
1559 }
```

`\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 `\makrboth`.

```
1560 \ifx\@mkboth\markboth
1561 \def\bbl@tempc{\let\@mkboth\markboth}
1562 \else
1563 \def\bbl@tempc{}
1564 \fi
```

Now we can start the new definition of `\markboth`

```
1565 \bbl@redefine\markboth#1#2{%
1566 \edef\bbl@tempb{\noexpand\protect
1567 \noexpand\foreignlanguage{\language}}%
1568 \def\bbl@arg{#1}%
1569 \ifx\bbl@arg@empty
1570 \toks@{}%
1571 \else
1572 \expandafter\toks@\expandafter{%
1573 \bbl@tempb{\protect\bbl@restore@actives#1}}%
1574 \fi
1575 \def\bbl@arg{#2}%
1576 \ifx\bbl@arg@empty
1577 \toks8{}%
1578 \else
1579 \expandafter\toks8\expandafter{%
1580 \bbl@tempb{\protect\bbl@restore@actives#2}}%
1581 \fi
1582 \edef\bbl@tempa{%
1583 \noexpand\org@markboth{\the\toks@}{\the\toks8}}%
1584 \bbl@tempa
1585 }
```

and copy it to `\@mkboth` if necessary.

```
1586 \bbl@tempc
```

## 8.4 Preventing clashes with other packages

### 8.4.1 `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.

```
1587 \@expandtwoargs\in@{R}\bbl@opt@safe
1588 \ifin@
1589   \AtBeginDocument{%
1590     \@ifpackageloaded{ifthen}{%
```

Then we can redefine `\ifthenelse`:

```
1591     \bbl@redefine@long\ifthenelse#1#2#3{%
```

We want to revert the definition of `\pageref` and `\ref` to their original definition for the duration of `\ifthenelse`, so we first need to store their current meanings.

```
1592       \let\bbl@tempa\pageref
1593       \let\pageref\org@pageref
1594       \let\bbl@tempb\ref
1595       \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.

```
1596         \@safe@activestru
1597         \org@ifthenelse{#1}{%
1598           \let\pageref\bbl@tempa
1599           \let\ref\bbl@tempb
1600           \@safe@activesfalse
1601           #2}{%
1602           \let\pageref\bbl@tempa
1603           \let\ref\bbl@tempb
1604           \@safe@activesfalse
1605           #3}%
1606       }%
1607   }{}%
1608 }
```

#### 8.4.2 varioref

`\@vpageref` When the package `varioref` is in use we need to modify its internal command  
`\vrefpagemum` `\@vpageref` in order to prevent problems when an active character ends up in the  
`\Ref` argument of `\vref`.

```
1609 \AtBeginDocument{%
1610   \@ifpackageloaded{varioref}{%
1611     \bbl@redefine\@vpageref#1[#2]#3{%
1612       \@safe@activestru
1613       \org\@vpageref{#1}[#2]{#3}%
1614       \@safe@activesfalse}%
```

The same needs to happen for `\vrefpagemum`.

```
1615     \bbl@redefine\vrefpagemum#1#2{%
1616       \@safe@activestru
1617       \org\vrefpagemum{#1}[#2]%
1618       \@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.

```

1619     \expandafter\def\csname Ref \endcsname#1{%
1620         \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
1621     }{}%
1622 }
1623 \fi

```

### 8.4.3 `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.

```

1624 \AtEndOfPackage{%
1625     \AtBeginDocument{%
1626         \@ifpackageloaded{hhline}%

```

Then we check whether the expansion of `\normal@char:` is not equal to `\relax`.

```

1627         {\expandafter\ifx\csname normal@char\string:\endcsname\relax
1628         \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.

```

1629             \makeatletter
1630             \def\@currname{hhline}\input{hhline.sty}\makeatother
1631         \fi}%
1632     {}}}

```

### 8.4.4 `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.

```

1633 \AtBeginDocument{%
1634     \ifundefined{pdfstringdefDisableCommands}%
1635     {}%
1636     {\pdfstringdefDisableCommands%
1637         \languageshorthands{system}}%
1638     }%
1639 }

```

### 8.4.5 `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.

```
1640 \DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
1641   \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.

```
1642 \def\substitutefontfamily#1#2#3{%
1643   \lowercase{\immediate\openout15=#1#2.fd\relax}%
1644   \immediate\write15{%
1645     \string\ProvidesFile{#1#2.fd}%
1646     [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
1647     \space generated font description file]^J
1648     \string\DeclareFontFamily{#1}{#2}{}^J
1649     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{}^J
1650     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{}^J
1651     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{}^J
1652     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{}^J
1653     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{}^J
1654     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{}^J
1655     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{}^J
1656     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{}^J
1657   }%
1658   \closeout15
1659 }
```

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

```
1660 \@onlypreamble\substitutefontfamily
```

## 8.5 Encoding issues

Because documents may use non-ASCII font encodings, we make sure that the logos of  $\TeX$  and  $\LaTeX$  always come out in the right encoding. There is a list of non-ASCII encodings. Unfortunately, `fontenc` deletes its package options, so we must guess which encodings 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
```

```
1661 \newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU,}
1662 \let\org@TeX\TeX
1663 \let\org@LaTeX\LaTeX
1664 \let\ensureascii\@firstofone
1665 \AtBeginDocument{%
1666   \in@false
1667   \bbl@loopx\bbl@tempa\BabelNonASCII{% is there a non-ascii enc?
1668     \ifin@else
1669       \edef\bbl@tempb{\,\bbl@tempa enc.def,}{,\@filelist,}}%
1670     \lowercase\expandafter{\expandafter\in@\bbl@tempb}%
1671     \fi}
1672 \ifin@ % if a non-ascii has been loaded
1673   \def\ensureascii#1{\fontencoding{OT1}\selectfont#1}}%
```

```

1674 \DeclareTextCommandDefault{\TeX}{\org@TeX}%
1675 \DeclareTextCommandDefault{\LaTeX}{\org@LaTeX}%
1676 \def\bbl@tempb#1\@@{\uppercase{\bbl@tempc#1}ENC.DEF\@empty\@@}%
1677 \def\bbl@tempc#1ENC.DEF#2\@@{%
1678   \ifx\@empty#2\else
1679     \ifundefined{T@#1}%
1680       {%
1681         \@expandtwoargs\in@{,#1,}{,\BabelNonASCII,}%
1682         \ifin@
1683           \DeclareTextCommand{\TeX}{#1}{\ensureascii{\org@TeX}}%
1684           \DeclareTextCommand{\LaTeX}{#1}{\ensureascii{\org@LaTeX}}%
1685         \else
1686           \def\ensureascii##1{\fontencoding{#1}\selectfont##1}%
1687           \fi}%
1688       \fi}%
1689 \bbl@loopx\bbl@tempa\@filelist{\expandafter\bbl@tempb\bbl@tempa\@@}%
1690 \@expandtwoargs\in@{,\cf@encoding,}{,\BabelNonASCII,}%
1691 \ifin@else
1692   \edef\ensureascii#1{%
1693     \noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}}%
1694 \fi
1695 \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.

```

1696 \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.

```

1697 \AtBeginDocument{%
1698   \@ifpackageloaded{fontspec}%
1699     {\xdef\latinencoding{EU\ifcase\bbl@engine\or2\or1\fi}}%
1700     {\gdef\latinencoding{OT1}%
1701       \ifx\cf@encoding\bbl@t@one
1702         \xdef\latinencoding{\bbl@t@one}%
1703       \else
1704         \ifl@aded{def}{tlenc}{\xdef\latinencoding{\bbl@t@one}}}%
1705     \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.

```

1706 \DeclareRobustCommand{\latintext}{%
1707   \fontencoding{\latinencoding}\selectfont
1708   \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.

```

1709 \ifx\@undefined\DeclareTextFontCommand
1710 \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
1711 \else
1712 \DeclareTextFontCommand{\textlatin}{\latintext}
1713 \fi

```

## 8.6 Local Language Configuration

`\loadlocalcfg` At some sites it may be necessary to add site-specific actions to a language definition file. This can be done by creating a file with the same name as the language definition file, but with the extension `.cfg`. For instance the file `norsk.cfg` will be loaded when the language definition file `norsk.ldf` is loaded. For plain-based formats we don't want to override the definition of `\loadlocalcfg` from `plain.def`.

```

1714 \ifx\loadlocalcfg\@undefined
1715 \ifpackagewith{babel}{noconfigs}%
1716 {\let\loadlocalcfg@gobble}%
1717 {\def\loadlocalcfg#1{%
1718 \InputIfFileExists{#1.cfg}%
1719 {\typeout{*****^J%
1720 * Local config file #1.cfg used^^J%
1721 *}}%
1722 \@empty}}
1723 \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:

```

1724 \ifx\@unexpandable@protect\@undefined
1725 \def\@unexpandable@protect{\noexpand\protect\noexpand}
1726 \long\def\protected@write#1#2#3{%
1727 \begingroup
1728 \let\thepage\relax
1729 #2%
1730 \let\protect\@unexpandable@protect
1731 \edef\reserved@a{\write#1{#3}}%
1732 \reserved@a
1733 \endgroup
1734 \if@nobreak\ifvmode\nobreak\fi\fi}
1735 \fi
1736 </core>

```

## 9 Internationalizing L<sup>A</sup>T<sub>E</sub>X 2.09

Now that we're sure that the code is seen by L<sup>A</sup>T<sub>E</sub>X only, we have to find out what the main (primary) document style is because we want to redefine some macros. This is only necessary for releases of L<sup>A</sup>T<sub>E</sub>X dated before December 1991. Therefore this part of the code can optionally be included in `babel.def` by specifying the `docstrip` option names.

The standard styles can be distinguished by checking whether some macros are defined. In table 1 an overview is given of the macros that can be used for this purpose.

The macros that have to be redefined for the `report` and `book` document styles happen to be the same, so there is no need to distinguish between those two styles.

|                 |   |   |
|-----------------|---|---|
| article         | : | both the <code>\chapter</code> and <code>\opening</code> macros are undefined         |
| report and book | : | the <code>\chapter</code> macro is defined and the <code>\opening</code> is undefined |
| letter          | : | the <code>\chapter</code> macro is undefined and the <code>\opening</code> is defined |

Table 1: How to determine the main document style

`\doc@style` First a parameter `\doc@style` is defined to identify the current document style. This parameter might have been defined by a document style that already uses macros instead of hard-wired texts, such as `artikel1.sty` [6], so the existence of `\doc@style` is checked. If this macro is undefined, i. e., if the document style is unknown and could therefore contain hard-wired texts, `\doc@style` is defined to the default value '0'.

```
1737 (*names)
1738 \ifx\@undefined\doc@style
1739 \def\doc@style{0}%
```

This parameter is defined in the following if construction (see table 1):

```
1740 \ifx\@undefined\opening
1741 \ifx\@undefined\chapter
1742 \def\doc@style{1}%
1743 \else
1744 \def\doc@style{2}%
1745 \fi
1746 \else
1747 \def\doc@style{3}%
1748 \fi%
1749 \fi%
```

Now here comes the real work: we start to redefine things and replace hard-wired texts by macros. These redefinitions should be carried out conditionally, in case it has already been done.

For the figure and table environments we have in all styles:

```
1750 \@ifundefined{figurename}{\def\fnm@figure{\figurename} \thefigure}}{}
1751 \@ifundefined{tablename}{\def\fnm@table{\tablename} \thetable}}{}
```

The rest of the macros have to be treated differently for each style. When `\doc@style` still has its default value nothing needs to be done.

```
1752 \ifcase \doc@style\relax
1753 \or
```

This means that `babel.def` is read after the article style, where no `\chapter` and `\opening` commands are defined<sup>28</sup>.

First we have the `\tableofcontents`, `\listoffigures` and `\listoftables`:

```
1754 \@ifundefined{contentsname}%
1755 {\def\tableofcontents{\section*{\contentsname\mkboth
1756 \uppercase{\contentsname}}{\uppercase{\contentsname}}}%
1757 \@starttoc{toc}}{}
1758 \@ifundefined{listfigurename}%
```

<sup>28</sup>A fact that was pointed out to me by Nico Poppelier and was already used in Piet van Oostrum's document style option `n1`.



```

1759     {\def\listoffigures{\section*{\listfigurename\mkboth
1760         {\uppercase{\listfigurename}}{\uppercase{\listfigurename}}}%
1761         \@starttoc{lof}}}{\fi}
1762 \ifundefined{listtablename}%
1763     {\def\listoftables{\section*{\listtablename\mkboth
1764         {\uppercase{\listtablename}}{\uppercase{\listtablename}}}%
1765         \@starttoc{lot}}}{\fi}

```

Then the \thebibliography and \theindex environments.

```

1766 \ifundefined{refname}%
1767     {\def\thebibliography#1{\section*{\refname
1768         \@mkboth{\uppercase{\refname}}{\uppercase{\refname}}}%
1769         \list{[\arabic{enumi}]}{\settowidth\labelwidth{[#1]}%
1770         \leftmargin\labelwidth
1771         \advance\leftmargin\labelsep
1772         \usecounter{enumi}}}%
1773         \def\newblock{\hskip.11em plus.33em minus.07em}%
1774         \sloppy\clubpenalty4000\widowpenalty\clubpenalty
1775         \sfcode'\.=1000\relax}}{\fi}
1776 \ifundefined{indexname}%
1777     {\def\theindex{\@restonecoltrue\if@twocolumn\@restonecolfalse\fi
1778         \columnseprule \z@
1779         \columnsep 35pt\twocolumn[\section*{\indexname}]%
1780         \@mkboth{\uppercase{\indexname}}{\uppercase{\indexname}}}%
1781         \thispagestyle{plain}%
1782         \parskip\z@ plus.3pt\parindent\z@\let\item\@idxitem}}{\fi}

```

The abstract environment:

```

1783 \ifundefined{abstractname}%
1784     {\def\abstract{\if@twocolumn
1785         \section*{\abstractname}%
1786         \else \small
1787         \begin{center}%
1788         {\bf \abstractname\vspace{-.5em}\vspace{\z@}}%
1789         \end{center}%
1790         \quotation
1791         \fi}}{\fi}

```

And last but not least, the macro \part:

```

1792 \ifundefined{partname}%
1793     {\def\part[#1]#2{\ifnum \c@secnumdepth >\m@ne
1794         \refstepcounter{part}%
1795         \addcontentsline{toc}{part}{\thepart
1796         \hspace{1em}#1}\else
1797         \addcontentsline{toc}{part}{#1}\fi
1798     {\parindent\z@ \raggedright
1799     \ifnum \c@secnumdepth >\m@ne
1800         \Large \bf \partname{} \thepart
1801         \par \nobreak
1802     \fi
1803     \huge \bf
1804     #2\markboth{}{\par}%
1805     \nobreak
1806     \vskip 3ex\@afterheading}%
1807 }{\fi}

```

This is all that needs to be done for the article style.

1808 \or

The next case is formed by the two styles book and report. Basically we have to do the same as for the article style, except now we must also change the \chapter command.

The tables of contents, figures and tables:

```
1809 \@ifundefined{contentsname}%
1810   {\def\tableofcontents{\@restonecolfalse
1811     \if@twocolumn\@restonecoltrue\onecolumn
1812     \fi\chapter*{\contentsname\@mkboth
1813       {\uppercase{\contentsname}}{\uppercase{\contentsname}}}%
1814     \@starttoc{toc}%
1815     \csname if@restonecol\endcsname\twocolumn
1816     \csname fi\endcsname}}{}
1817 \@ifundefined{listfigurename}%
1818   {\def\listoffigures{\@restonecolfalse
1819     \if@twocolumn\@restonecoltrue\onecolumn
1820     \fi\chapter*{\listfigurename\@mkboth
1821       {\uppercase{\listfigurename}}{\uppercase{\listfigurename}}}%
1822     \@starttoc{lof}%
1823     \csname if@restonecol\endcsname\twocolumn
1824     \csname fi\endcsname}}{}
1825 \@ifundefined{listtablename}%
1826   {\def\listoftables{\@restonecolfalse
1827     \if@twocolumn\@restonecoltrue\onecolumn
1828     \fi\chapter*{\listtablename\@mkboth
1829       {\uppercase{\listtablename}}{\uppercase{\listtablename}}}%
1830     \@starttoc{lot}%
1831     \csname if@restonecol\endcsname\twocolumn
1832     \csname fi\endcsname}}{}
```

Again, the bibliography and index environments; notice that in this case we use \bibname instead of \refname as in the definitions for the article style. The reason for this is that in the article document style the term 'References' is used in the definition of \thebibliography. In the report and book document styles the term 'Bibliography' is used.

```
1833 \@ifundefined{bibname}%
1834   {\def\thebibliography#1{\chapter*{\bibname
1835     \@mkboth{\uppercase{\bibname}}{\uppercase{\bibname}}}%
1836     \list{[\arabic{enumi}]}{\settowidth\labelwidth{[#1]}%
1837     \leftmargin\labelwidth \advance\leftmargin\labelsep
1838     \usecounter{enumi}}%
1839     \def\newblock{\hskip.11em plus.33em minus.07em}%
1840     \sloppy\clubpenalty4000\widowpenalty\clubpenalty
1841     \sfcode'\.=1000\relax}}{}
1842 \@ifundefined{indexname}%
1843   {\def\theindex{\@restonecoltrue\if@twocolumn\@restonecolfalse\fi
1844     \columnseprule \z@
1845     \columnsep 35pt\twocolumn[\@makeschapterhead{\indexname}]%
1846     \@mkboth{\uppercase{\indexname}}{\uppercase{\indexname}}}%
1847     \thispagestyle{plain}%
1848     \parskip\z@ plus.3pt\parindent\z@ \let\item\@idxitem}}{}
```

Here is the abstract environment:

```
1849 \@ifundefined{abstractname}%
1850   {\def\abstract{\titlepage
1851     \null\vfil
```

```

1852 \begin{center}%
1853 {\bf \abstractname}%
1854 \end{center}}{}

```

And last but not least the `\chapter`, `\appendix` and `\part` macros.

```

1855 \@ifundefined{chaptername}{\def\@chapapp{\chaptername}}{}
1856 %
1857 \@ifundefined{appendixname}%
1858   {\def\appendix{\par
1859     \setcounter{chapter}{0}%
1860     \setcounter{section}{0}%
1861     \def\@chapapp{\appendixname}%
1862     \def\thechapter{\Alph{chapter}}}}{}
1863 %
1864 \@ifundefined{partname}%
1865   {\def\@part[#1]#2{\ifnum \c@secnumdepth >-2\relax
1866     \refstepcounter{part}%
1867     \addcontentsline{toc}{part}{\thepart
1868       \hspace{1em}#1}\else
1869     \addcontentsline{toc}{part}{#1}\fi
1870     \markboth{}{}}%
1871   {\centering
1872     \ifnum \c@secnumdepth >-2\relax
1873       \huge\bf \partname{} \thepart
1874     \par
1875     \vskip 20pt \fi
1876     \Huge \bf
1877     #1\par}\@endpart}}{}%
1878 \or

```

Now we address the case where `babel.def` is read after the letter style. The letter document style defines the macro `\opening` and some other macros that are specific to letter. This means that we have to redefine other macros, compared to the previous two cases.

First two macros for the material at the end of a letter, the `\cc` and `\encl` macros.

```

1879 \@ifundefined{ccname}%
1880   {\def\cc#1{\par\noindent
1881     \parbox[t]{\textwidth}%
1882     {\@hangfrom{\rm \ccname : }\ignorespaces #1\strut}\par}}{}
1883 \@ifundefined{enclname}%
1884   {\def\encl#1{\par\noindent
1885     \parbox[t]{\textwidth}%
1886     {\@hangfrom{\rm \enclname : }\ignorespaces #1\strut}\par}}{}

```

The last thing we have to do here is to redefine the headings `pagestyle`:

```

1887 \@ifundefined{headtoname}%
1888   {\def\ps@headings{%
1889     \def\@oddhead{\sl \headtoname} \ignorespaces\toname \hfil
1890     \@date \hfil \pagename{} \thepage}%
1891   \def\@oddfoot{}}}{}

```

This was the last of the four standard document styles, so if `\doc@style` has another value we do nothing and just close the `if` construction.

```

1892 \fi
1893 \endnames

```

Here ends the code that can be optionally included when a version of L<sup>A</sup>T<sub>E</sub>X is in use that is dated *before* December 1991.

We also need to redefine a number of commands to ensure that the right font encoding is used, but this can't be done before `babel.def` is loaded.

## 10 Multiple languages

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.

```
1894 (*kernel)
1895 <<Make sure ProvidesFile is defined>>
1896 \ProvidesFile{switch.def}[\langle date \rangle \langle version \rangle Babel switching mechanism]
1897 <<Load macros for plain if not LaTeX>>
1898 <<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.

```
1899 \def\bbl@version{\langle version \rangle}
1900 \def\bbl@date{\langle date \rangle}
1901 \def\adddialect#1#2{%
1902   \global\chardef#1#2\relax
1903   \bbl@usehooks{adddialect}{\#1}{\#2}}%
1904   \wlog{\string#1 = a dialect from \string\language#2}}
```

`\bbl@iflanguage` executes code only if the language 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).

```
1905 \def\bbl@fixname#1{%
1906   \begingroup
1907     \def\bbl@tempe{l}%
1908     \edef\bbl@tempd{\noexpand\ifundefined{\noexpand\bbl@tempe#1}}%
1909     \bbl@tempd
1910     {\lowercase\expandafter{\bbl@tempd}}%
1911     {\uppercase\expandafter{\bbl@tempd}}%
1912     \@empty
1913     {\edef\bbl@tempd{\def\noexpand#1{\#1}}%
1914      \uppercase\expandafter{\bbl@tempd}}}%
1915     {\edef\bbl@tempd{\def\noexpand#1{\#1}}%
1916      \lowercase\expandafter{\bbl@tempd}}}%
1917   \@empty
1918   \edef\bbl@tempd{\endgroup\def\noexpand#1{\#1}}%
1919   \bbl@tempd}
1920 \def\bbl@iflanguage#1{%
1921   \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.

```

1922 \def\iflanguage#1{%
1923   \bbl@iflanguage{#1}{%
1924     \ifnum\csname l@#1\endcsname=\language
1925       \expandafter\@firstoftwo
1926     \else
1927       \expandafter\@secondoftwo
1928     \fi}}

```

## 10.1 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.

```

1929 \let\bbl@select@type\z@
1930 \edef\selectlanguage{%
1931   \noexpand\protect
1932   \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`.

```

1933 \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.

```

1934 \ifx\documentclass\@undefined
1935   \def\xstring{\string\string\string}
1936 \else
1937   \let\xstring\string
1938 \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.

```
1939 \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`

```
1940 \def\bbl@push@language{%
```

```
1941 \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.

```
1942 \def\bbl@pop@lang#1+#2-#3{%
```

```
1943 \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.

```
1944 \def\bbl@pop@language{%
```

```
1945 \expandafter\bbl@pop@lang\bbl@language@stack-\bbl@language@stack
```

```
1946 \expandafter\bbl@set@language\expandafter{\language}}
```

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.

```
1947 \expandafter\def\csname selectlanguage \endcsname#1{%
```

```
1948 \ifnum\bbl@hymapsel=\cclv\let\bbl@hymapsel\tw\fi
```

```
1949 \bbl@push@language
```

```
1950 \aftergroup\bbl@pop@language
```

```
1951 \bbl@set@language{#1}}
```

`\bbl@set@language` The macro `\bbl@set@language` takes care of switching the language environment *and* of writing entries on the auxiliary files. For historical reasons, language names can be either `language` or `\language`. To catch either form a trick is used, but unfortunately as a side effect the catcodes of letters in `\language` are not well

defined. 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.

```

1952 \def\BabelContentsFiles{toc,lof,lot}
1953 \def\bbl@set@language#1{%
1954   \edef\language#1{%
1955     \ifnum\escapechar=\expandafter'\string#1\@empty
1956     \else\string#1\@empty\fi}%
1957   \select@language{\language}%
1958   \expandafter\ifx\csname date\language\endcsname\relax\else
1959     \if@filesw
1960       \protected@write\@auxout{{\string\select@language{\language}}}%
1961       \bbl@for\bbl@tempa\BabelContentsFiles{%
1962         \addtocontents{\bbl@tempa}{\xstring\select@language{\language}}}%
1963       \bbl@usehooks{write}{}%
1964     \fi
1965   \fi}
1966 \def\select@language#1{%
1967   \ifnum\bbl@hymapsel=@cclv\chardef\bbl@hymapsel4\relax\fi
1968   \edef\language#1{%
1969     \bbl@fixname\language
1970     \bbl@iflanguage\language{%
1971       \expandafter\ifx\csname date\language\endcsname\relax
1972         \bbl@error
1973           {Unknown language '#1'. Either you have\\%
1974            misspelled its name, it has not been installed,\\%
1975            or you requested it in a previous run. Fix its name,\\%
1976            install it or just rerun the file, respectively}%
1977           {You may proceed, but expect unexpected results}%
1978       \else
1979         \let\bbl@select@type\z@
1980         \expandafter\bbl@switch\expandafter{\language}%
1981       \fi}}
1982 % A bit of optimization:
1983 \def\select@language@x#1{%
1984   \ifcase\bbl@select@type
1985     \bbl@ifsamestring\language#1}{\select@language#1}%
1986   \else
1987     \select@language#1}%
1988 \fi}

```

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 `\leftthyphenmin` and `\rightthyphenmin` is somewhat

different. First we save their current values, then we check if `\langle lang \rangle hyphenmins` is defined. If it is not, we set default values (2 and 3), otherwise the values in `\langle lang \rangle hyphenmins` will be used.

```

1989 \def\bbl@switch#1{%
1990   \originalTeX
1991   \expandafter\def\expandafter\originalTeX\expandafter{%
1992     \csname noextras#1\endcsname
1993     \let\originalTeX\@empty
1994     \babel@beginsave}%
1995   \bbl@usehooks{afterreset}{}%
1996   \languageshorthands{none}%
1997   \ifcase\bbl@select@type
1998     \csname captions#1\endcsname
1999     \csname date#1\endcsname
2000   \fi
2001   \bbl@usehooks{beforeextras}{}%
2002   \csname extras#1\endcsname\relax
2003   \bbl@usehooks{afterextras}{}%
2004   \ifcase\bbl@hymapopt\or
2005     \def\BabelLower##1##2{\lccode##1=##2\relax}%
2006     \ifnum\bbl@hymapsel>4\else
2007       \csname\languagename @bbl@hyphenmap\endcsname
2008       \fi
2009     \chardef\bbl@hymapopt\z@
2010   \else
2011     \ifnum\bbl@hymapsel>\bbl@hymapopt\else
2012       \csname\languagename @bbl@hyphenmap\endcsname
2013     \fi
2014   \fi
2015   \global\let\bbl@hymapsel\@cclv
2016   \bbl@patterns{#1}%
2017   \babel@savevariable\lefthyphenmin
2018   \babel@savevariable\rightthyphenmin
2019   \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2020     \set@hyphenmins\tw@\thr@\relax
2021   \else
2022     \expandafter\expandafter\expandafter\set@hyphenmins
2023     \csname #1hyphenmins\endcsname\relax
2024   \fi}
2025 \def\bbl@ifsamestring#1#2{%
2026   \protected@edef\bbl@tempb{#1}%
2027   \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
2028   \protected@edef\bbl@tempc{#2}%
2029   \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
2030   \ifx\bbl@tempb\bbl@tempc
2031     \expandafter\@firstoftwo
2032   \else
2033     \expandafter\@secondoftwo
2034   \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 first thing this environment does is store the name of the language in



`\language`; it then calls `\selectlanguage_` to switch on everything that is needed for this language. The `\ignorespaces` command is necessary to hide the environment when it is entered in horizontal mode.

```
2035 \long\def\otherlanguage#1{%
2036   \ifnum\bbbl@hymapsel=\@ccclv\let\bbbl@hymapsel\thr@@\fi
2037   \csname selectlanguage \endcsname{#1}%
2038   \ignorespaces}
```

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

```
2039 \long\def\endotherlanguage{%
2040   \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`.

```
2041 \expandafter\def\csname otherlanguage*\endcsname#1{%
2042   \ifnum\bbbl@hymapsel=\@ccclv\chardef\bbbl@hymapsel4\relax\fi
2043   \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”.

```
2044 \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`.

```
2045 \edef\foreignlanguage{%
2046   \noexpand\protect
2047   \expandafter\noexpand\csname foreignlanguage \endcsname}
2048 \expandafter\def\csname foreignlanguage \endcsname#1#2{%
2049   \begingroup
2050     \foreign@language{#1}%
2051     #2%
2052   \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 `bbbl@switch`.

```
2053 \def\foreign@language#1{%
2054   \edef\languagename{#1}%
2055   \bbbl@fixname\languagename
2056   \bbbl@iflanguage\languagename{%
2057     \expandafter\ifx\csname date\languagename\endcsname\relax
2058     \bbbl@warning
2059     {You haven’t loaded the language \languagename\space yet\\%
2060     I’ll proceed, but expect unexpected results.\\%
2061     Reported}%
2062     \fi
2063     \let\bbbl@select@type\@ne
2064     \expandafter\bbbl@switch\expandafter{\languagename}}
```

`\bbl@patterns` This macro selects the hyphenation patterns by changing the `\language` register. If special hyphenation patterns are available specifically for the current font encoding, use them instead of the default. It also sets hyphenation exceptions, but only once, because they are global (here language `\lccode's` has been set, too). `\bbl@hyphenation@` is set to relax until the very first `\babelhyphenation`, so do nothing with this value. If the exceptions for a language (by its number, not its name, so that `:ENC` is taken into account) has been set, then use `\hyphenation` with both global and language exceptions and empty the latter to mark they must not be set again.

```

2065 \let\bbl@hyphlist@empty
2066 \let\bbl@hyphenation@\relax
2067 \let\bbl@pttnlist@empty
2068 \let\bbl@patterns@\relax
2069 \let\bbl@hymapsel=\ccclv
2070 \def\bbl@patterns#1{%
2071   \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
2072     \csname l@#1\endcsname
2073     \edef\bbl@tempa{#1}%
2074   \else
2075     \csname l@#1:\f@encoding\endcsname
2076     \edef\bbl@tempa{#1:\f@encoding}%
2077   \fi\relax
2078   \@expandtwoargs\bbl@usehooks{patterns}{{#1}{\bbl@tempa}}%
2079   \@ifundefined{bbl@hyphenation@}{}{%
2080     \begingroup
2081       \@expandtwoargs\in@{,\number\language,}{,\bbl@hyphlist}%
2082       \ifin@else
2083         \@expandtwoargs\bbl@usehooks{hyphenation}{{#1}{\bbl@tempa}}%
2084         \hyphenation{%
2085           \bbl@hyphenation@
2086           \@ifundefined{bbl@hyphenation@#1}%
2087             \@empty
2088             {\space\csname bbl@hyphenation@#1\endcsname}}%
2089         \xdef\bbl@hyphlist{\bbl@hyphlist\bbl@hyphlist\number\language,}%
2090       \fi
2091     \endgroup}}

```

`hyphenrules` The environment `hyphenrules` can be used to select *just* the hyphenation rules. This environment does *not* change `\languagename` and when the hyphenation rules specified were not loaded it has no effect. Note however, `\lccode's` and font encodings are not set at all, so in most cases you should use `otherlanguage*`.

```

2092 \def\hyphenrules#1{%
2093   \edef\languagename{#1}%
2094   \bbl@fixname\languagename
2095   \bbl@iflanguage\languagename{%
2096     \expandafter\bbl@patterns\expandafter{\languagename}%
2097     \languageshorthands{none}%
2098     \expandafter\ifx\csname\languagename hyphenmins\endcsname\relax
2099       \set@hyphenmins\tw@\thr@\relax
2100   \else
2101     \expandafter\expandafter\expandafter\set@hyphenmins
2102     \csname\languagename hyphenmins\endcsname\relax
2103   \fi}}
2104 \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 `\langle lang \rangle hyphenmins` is already defined this command has no effect.

```
2105 \def\providehyphenmins#1#2{%
2106   \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2107     \namedef{#1hyphenmins}{#2}%
2108   \fi}
```

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

```
2109 \def\set@hyphenmins#1#2{\lefthyphenmin#1\relax\righthyphenmin#2\relax}
```

`\ProvidesLanguage` The identification code for each file is something that was introduced in L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>. When the command `\ProvidesFile` does not exist, a dummy definition is provided temporarily. For use in the language definition file the command `\ProvidesLanguage` is defined by `babel`. Depending on the format, ie, on if the former is defined, we use a similar definition or not.

```
2110 \ifx\ProvidesFile\undefined
2111   \def\ProvidesLanguage#1[#2 #3 #4]{%
2112     \wlog{Language: #1 #4 #3 <#2>}%
2113   }
2114 \else
2115   \def\ProvidesLanguage#1{%
2116     \begingroup
2117     \catcode'\ 10 %
2118     \@makeother\/%
2119     \@ifnextchar[%]
2120       {@provideslanguage{#1}}{@provideslanguage{#1}[]}
2121   \def{@provideslanguage#1[#2]}{%
2122     \wlog{Language: #1 #2}%
2123     \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
2124     \endgroup}
2125 \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`

```
2126 \def\LdfInit{%
2127   \chardef\atcatcode=\catcode'\@
2128   \catcode'\@=11\relax
2129   \input babel.def\relax
2130   \catcode'\@=\atcatcode \let\atcatcode\relax
2131   \LdfInit}
```

`\originalTeX` The macro `\originalTeX` should be known to T<sub>E</sub>X at this moment. As it has to be expandable we `\let` it to `\@empty` instead of `\relax`.

```
2132 \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.

```
2133 \ifx\babel@beginsave\@undefined\let\babel@beginsave\relax\fi
```

## 10.2 Errors

`\@nolanerr` `\@nopatterns` 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 L<sup>A</sup>T<sub>E</sub>X<sub>2 $\epsilon$</sub> , so we can safely use its error handling interface. Otherwise we'll have to 'keep it simple'.

```
2134 \edef\bbl@nulllanguage{\string\language=0}
```

```
2135 \ifx\PackageError\@undefined
```

```
2136   \def\bbl@error#1#2{%
```

```
2137     \begingroup
```

```
2138       \newlinechar='^^J
```

```
2139       \def\{^^J(babel) }%
```

```
2140       \errhelp{#2}\errmessage{\#\#1}%
```

```
2141     \endgroup}
```

```
2142 \def\bbl@warning#1{%
```

```
2143   \begingroup
```

```
2144     \newlinechar='^^J
```

```
2145     \def\{^^J(babel) }%
```

```
2146     \message{\#\#1}%
```

```
2147   \endgroup}
```

```
2148 \def\bbl@info#1{%
```

```
2149   \begingroup
```

```
2150     \newlinechar='^^J
```

```
2151     \def\{^^J}%
```

```
2152     \wlog{#1}%
```

```
2153   \endgroup}
```

```
2154 \else
```

```
2155   \def\bbl@error#1#2{%
```

```
2156     \begingroup
```

```
2157       \def\{\MessageBreak}%
```

```
2158       \PackageError{babel}{#1}{#2}%
```

```
2159     \endgroup}
```

```
2160 \def\bbl@warning#1{%
```

```
2161   \begingroup
```

```
2162     \def\{\MessageBreak}%
```

```
2163     \PackageWarning{babel}{#1}%
```

```
2164   \endgroup}
```

```
2165 \def\bbl@info#1{%
```

```
2166   \begingroup
```

```
2167     \def\{\MessageBreak}%
```

```
2168     \PackageInfo{babel}{#1}%
```

```
2169   \endgroup}
```

```
2170 \fi
```

```
2171 \@ifpackagewith{babel}{silent}
```

```

2172 {\let\bbl@info@gobble
2173  \let\bbl@warning@gobble}
2174 {}
2175 \def\nolanerr#1{%
2176  \bbl@error
2177   {You haven't defined the language #1\space yet}%
2178   {Your command will be ignored, type <return> to proceed}}
2179 \def\nopatpatterns#1{%
2180  \bbl@warning
2181   {No hyphenation patterns were preloaded for\\%
2182    the language '#1' into the format.\\%
2183    Please, configure your TeX system to add them and\\%
2184    rebuild the format. Now I will use the patterns\\%
2185    preloaded for \bbl@nulllanguage\space instead}}
2186 \let\bbl@usehooks@gobbletwo
2187 </kernel>

```

## 11 Loading hyphenation patterns

The following code is meant to be read by `iniTeX` because it should instruct `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.

`toks8` stores info to be shown when the program is run.

We want to add a message to the message `LATEX 2.09` puts in the `\everyjob` register. This could be done by the following code:

```

% \let\orgeveryjob\everyjob
% \def\everyjob#1{%
%   \orgeveryjob{#1}%
%   \orgeveryjob\expandafter{\the\orgeveryjob\immediate\write16{%
%     hyphenation patterns for \the\loaded@patterns loaded.}}%
%   \let\everyjob\orgeveryjob\let\orgeveryjob@undefined}
%

```

The code above redefines the control sequence `\everyjob` in order to be able to add something to the current contents of the register. This is necessary because the processing of hyphenation patterns happens long before `LATEX` fills the register. There are some problems with this approach though.

- When someone wants to use several hyphenation patterns with `LATEX` the above scheme won't work. The reason is that `LATEX` overwrites the contents of the `\everyjob` register with its own message.
- Plain `TeX` does not use the `\everyjob` register so the message would not be displayed.

To circumvent this a 'dirty trick' can be used. As this code is only processed when creating a new format file there is one command that is sure to be used, `\dump`. Therefore the original `\dump` is saved in `\org@dump` and a new definition is supplied.

To make sure that `LATEX 2.09` executes the `\@begindocumenthook` we would want to alter `\begin{document}`, but as this done too often already, we add the new code at the front of `\@preamblecmds`. But we can only do that after it has been defined, so we add this piece of code to `\dump`.

This new definition starts by adding an instruction to write a message on the terminal and in the transcript file to inform the user of the preloaded hyphenation patterns.

Then everything is restored to the old situation and the format is dumped.

```

2188 (*patterns)
2189 <<Make sure ProvidesFile is defined>>
2190 \ProvidesFile{hyphen.cfg}[<<date>> <<version>>] Babel hyphens]
2191 \xdef\bbl@format{\jobname}
2192 \ifx\AtBeginDocument\@undefined
2193   \def\@empty{}
2194   \let\orig@dump\dump
2195   \def\dump{%
2196     \ifx\@ztryfc\@undefined
2197       \else
2198         \toks0=\expandafter{\@preamblecmds}%
2199         \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
2200         \def\@begindocumenthook{}%
2201       \fi
2202       \let\dump\orig@dump\let\orig@dump\@undefined\dump}
2203 \fi
2204 <<Define core switching macros>>
2205 \toks8{Babel <<@version@>> and hyphenation patterns for }%

```

`\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.

```

2206 \def\process@line#1#2 #3 #4 {%
2207   \ifx=#1%
2208     \process@synonym{#2}%
2209   \else
2210     \process@language{#1#2}{#3}{#4}%
2211   \fi
2212   \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.

```

2213 \toks@{}
2214 \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.

```

2215 \def\process@synonym#1{%
2216   \ifnum\last@language=\m@ne
2217     \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
2218   \else
2219     \expandafter\chardef\csname l@#1\endcsname\last@language
2220     \wlog{\string\l@#1=\string\language\the\last@language}%
2221     \expandafter\let\csname #1hyphenmins\expandafter\endcsname
2222       \csname\languagename hyphenmins\endcsname
2223     \let\bbl@elt\relax

```

```

2224 \edef\bbbl@languages{\bbbl@languages\bbbl@elt{#1}{\the\last@language}{}}%
2225 \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 ‘name’ of the language that will be loaded now is added to the token register `\toks8`. and finally 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 `\bbbl@get@enc` extracts the font encoding from the language name and stores it in `\bbbl@hyph@enc`. The latter can be used in hyphenation files if you need to set a behaviour 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 `\langle lang \rangle hyphenmins` macro. When no assignments were made we provide a default setting.

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

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

`\bbbl@languages` saves a snapshot of the loaded languages in the form `\bbbl@elt{\langle language-name \rangle}{\langle number \rangle}{\langle patterns-file \rangle}{\langle exceptions-file \rangle}`. 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.

```

2226 \def\process@language#1#2#3{%
2227 \expandafter\addlanguage\csname l@#1\endcsname
2228 \expandafter\language\csname l@#1\endcsname
2229 \edef\languagename{#1}%
2230 \bbbl@hook@everylanguage{#1}%
2231 \bbbl@get@enc#1::\@@@
2232 \begingroup
2233 \lefthyphenmin\m@ne
2234 \bbbl@hook@loadpatterns{#2}%
2235 \ifnum\lefthyphenmin=\m@ne
2236 \else
2237 \expandafter\xdef\csname #1hyphenmins\endcsname{%
2238 \the\lefthyphenmin\the\righthyphenmin}%
2239 \fi
2240 \endgroup
2241 \def\bbbl@tempa{#3}%
2242 \ifx\bbbl@tempa\@empty\else
2243 \bbbl@hook@loadexceptions{#3}%

```

```

2244 \fi
2245 \let\bbl@elt\relax
2246 \edef\bbl@languages{%
2247   \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
2248 \ifnum\the\language=\z@
2249   \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2250     \set@hyphenmins\tw@\thr@\relax
2251   \else
2252     \expandafter\expandafter\expandafter\set@hyphenmins
2253     \csname #1hyphenmins\endcsname
2254   \fi
2255   \the\toks@
2256   \toks@{}%
2257 \fi}

```

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

```

2258 \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.

```

2259 \def\bbl@hook@everylanguage#1{}
2260 \def\bbl@hook@loadpatterns#1{\input #1\relax}
2261 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
2262 \let\bbl@hook@loadkernel\bbl@hook@loadpatterns
2263 \begingroup
2264   \def\AddBabelHook#1#2{%
2265     \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
2266       \def\next{\toks1}%
2267     \else
2268       \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname###1}%
2269     \fi
2270     \next}
2271 \ifx\directlua\@undefined
2272   \ifx\XeTeXinputencoding\@undefined\else
2273     \input xebabel.def
2274   \fi
2275 \else
2276   \input luababel.def
2277 \fi
2278 \openin1 = babel-\bbl@format.cfg
2279 \ifeof1
2280 \else
2281   \input babel-\bbl@format.cfg\relax
2282 \fi
2283 \closein1
2284 \endgroup
2285 \bbl@hook@loadkernel{switch.def}

```

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

```

2286 \openin1 = language.dat

```

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

```

2287 \def\languagename{english}%
2288 \ifeof1

```



```

2289 \message{I couldn't find the file language.dat,\space
2290         I will try the file hyphen.tex}
2291 \input hyphen.tex\relax
2292 \chardef\l@english\z@
2293 \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`.

```
2294 \last@language\m@ne
```

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

```
2295 \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.

```

2296 \endlinechar\m@ne
2297 \read1 to \bbl@line
2298 \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.

```

2299 \if T\ifeof1F\fi T\relax
2300 \ifx\bbl@line\empty\else
2301 \edef\bbl@line{\bbl@line\space\space\space}%
2302 \expandafter\process@line\bbl@line\relax
2303 \fi
2304 \repeat

```

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

```

2305 \begingroup
2306 \def\bbl@elt#1#2#3#4{%
2307 \global\language=#2\relax
2308 \gdef\languagename{#1}%
2309 \def\bbl@elt##1##2##3##4{}}%
2310 \bbl@languages
2311 \endgroup
2312 \fi

```

and close the configuration file.

```
2313 \closein1
```

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

```

2314 \if/\the\toks@/\else
2315 \errhelp{language.dat loads no language, only synonyms}
2316 \errmessage{Orphan language synonym}
2317 \fi
2318 \edef\bbl@tempa{%
2319 \everyjob{%

```

```

2320 \the\everyjob
2321 \ifx\typeout\@undefined
2322 \immediate\writel6%
2323 \else
2324 \noexpand\typeout
2325 \fi
2326 {\the\toks8 \the\last@language\space languages loaded.}}
2327 \bbl@tempa

```

Also remove some macros from memory and raise an error if `\toks@` is not empty. Finally load `switch.def`, but the letter is not required and the line inputting it may be commented out.

```

2328 \let\bbl@line\@undefined
2329 \let\process@line\@undefined
2330 \let\process@synonym\@undefined
2331 \let\process@language\@undefined
2332 \let\bbl@get@enc\@undefined
2333 \let\bbl@hyph@enc\@undefined
2334 \let\bbl@tempa\@undefined
2335 \let\bbl@hook@loadkernel\@undefined
2336 \let\bbl@hook@everylanguage\@undefined
2337 \let\bbl@hook@loadpatterns\@undefined
2338 \let\bbl@hook@loadexceptions\@undefined
2339 </patterns>

```

Here the code for `iniTEX` ends.

## 12 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.

```

2340 <*nil>
2341 \ProvidesLanguage{nil}[<<date>> <<version>> Nil language]
2342 \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.

```

2343 \ifx\l@nohyphenation\@undefined
2344 \@nopatterns{nil}
2345 \adddialect\l@nil0
2346 \else
2347 \let\l@nil\l@nohyphenation
2348 \fi

```

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

```

2349 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}

```

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

```

\captionnil
\datenil 2350 \let\captionnil\@empty
2351 \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.

```
2352 \ldf@finish{nil}
2353 \end{document}
```

## 13 Support for Plain T<sub>E</sub>X

### 13.1 Not renaming `hyphen.tex`

As Don Knuth has declared that the filename `hyphen.tex` may only be used to designate *his* version of the american English hyphenation patterns, a new solution has to be found in order to be able to load hyphenation patterns for other languages in a plain-based T<sub>E</sub>X-format. When asked he responded:

That file name is “sacred”, and if anybody changes it they will cause severe upward/downward compatibility headaches.

People can have a file `locallyhyphen.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 `lplain.tex` can be used as replacement wrappers around `plain.tex` and `lplain.tex` to achieve the desired effect, based on the `babel` package. If you load each of them with `iniTEX`, you will get a file called either `bplain.fmt` or `lplain.fmt`, which you can use as replacements for `plain.fmt` and `lplain.fmt`.

As these files are going to be read as the first thing `iniTEX` sees, we need to set some category codes just to be able to change the definition of `\input`

```
2354 (*bplain | lplain)
2355 \catcode'\{=1 % left brace is begin-group character
2356 \catcode'\}=2 % right brace is end-group character
2357 \catcode'\#=6 % hash mark is macro parameter character
```

Now let's see if a file called `hyphen.cfg` can be found somewhere on T<sub>E</sub>X's input path by trying to open it for reading...

```
2358 \openin 0 hyphen.cfg
```

If the file wasn't found the following test turns out true.

```
2359 \ifeof0
2360 \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).

```
2361 \let\input
```

Then `\input` is defined to forget about its argument and load `hyphen.cfg` instead.

```
2362 \def\input #1 {%
2363   \let\input a
2364   a hyphen.cfg
```

Once that's done the original meaning of `\input` can be restored and the definition of `\a` can be forgotten.

```
2365 \let\a\undefined
2366 }
2367 \fi
2368 </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`.

```
2369 <bplain>\a plain.tex
2370 <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.

```
2371 <bplain>\def\fmtname{babel-plain}
2372 <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.

## 13.2 Emulating some L<sup>A</sup>T<sub>E</sub>X features

The following code duplicates or emulates parts of L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> that are needed for `babel`.

We need to define `\loadlocalcfg` for plain users as the L<sup>A</sup>T<sub>E</sub>X definition uses `\InputIfFileExists`. We have to execute `\@endofldf` in this case.

```
2373 (*plain)
2374 \def\@empty{}
2375 \def\loadlocalcfg#1{%
2376   \openin0#1.cfg
2377   \ifeof0
2378     \closein0
2379   \else
2380     \closein0
2381     {\immediate\write16{*****}%
2382      \immediate\write16{* Local config file #1.cfg used}%
2383      \immediate\write16{*}%
2384     }
2385     \input #1.cfg\relax
2386   \fi
2387   \@endofldf}
```

## 13.3 General tools

A number of L<sup>A</sup>T<sub>E</sub>X macro's that are needed later on.

```
2388 \long\def\@firstofone#1{#1}
2389 \long\def\@firstoftwo#1#2{#1}
2390 \long\def\@secondoftwo#1#2{#2}
2391 \def\@nnil{\@nil}
2392 \def\@gobbletwo#1#2{}
2393 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
2394 \def\@star@or@long#1{%
2395   \@ifstar
2396   {\let\@ngrel@x\relax#1}%
2397   {\let\@ngrel@x\long#1}}
2398 \let\@ngrel@x\relax
```

```

2399 \def\@car#1#2\@nil{#1}
2400 \def\@cdr#1#2\@nil{#2}
2401 \let\@typeset@protect\relax
2402 \let\protected@edef\edef
2403 \long\def\@gobble#1{}
2404 \edef\@backslashchar{\expandafter\@gobble\string\}
2405 \def\strip@prefix#1>{}
2406 \def\g@addto@macro#1#2{%
2407     \toks@\expandafter{#1#2}%
2408     \xdef#1{\the\toks@}}
2409 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
2410 \def\@nameuse#1{\csname #1\endcsname}
2411 \def\@ifundefined#1{%
2412     \expandafter\ifx\csname#1\endcsname\relax
2413     \expandafter\@firstoftwo
2414     \else
2415     \expandafter\@secondoftwo
2416     \fi}
2417 \def\@expandtwoargs#1#2#3{%
2418     \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
2419 \def\zap@space#1 #2{%
2420     #1%
2421     \ifx#2\empty\else\expandafter\zap@space\fi
2422     #2}

```

$\LaTeX 2_{\epsilon}$  has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

2423 \ifx\@preamblecmds\@undefined
2424     \def\@preamblecmds{}
2425 \fi
2426 \def\@onlypreamble#1{%
2427     \expandafter\gdef\expandafter\@preamblecmds\expandafter{
2428         \@preamblecmds\do#1}}
2429 \@onlypreamble\@onlypreamble

```

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

```

2430 \def\begin{document}{%
2431     \@begin{document}hook
2432     \global\let\@begin{document}hook\@undefined
2433     \def\do##1{\global\let##1\@undefined}%
2434     \@preamblecmds
2435     \global\let\do\noexpand}
2436 \ifx\@begin{document}hook\@undefined
2437     \def\@begin{document}hook{}
2438 \fi
2439 \@onlypreamble\@begin{document}hook
2440 \def\AtBeginDocument{\g@addto@macro\@begin{document}hook}

```

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

```

2441 \def\AtEndOfPackage#1{\g@addto@macro\@endofldf{#1}}
2442 \@onlypreamble\AtEndOfPackage
2443 \def\@endofldf{}
2444 \@onlypreamble\@endofldf
2445 \let\bb@afterlang\empty
2446 \chardef\bb@hymapopt\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.

```

2447 \ifx\if@filesw\undefined
2448 \expandafter\let\csname if@filesw\expandafter\endcsname
2449 \csname iffalse\endcsname
2450 \fi

```

Mimick L<sup>A</sup>T<sub>E</sub>X's commands to define control sequences.

```

2451 \def\newcommand{\@star@or@long\new@command}
2452 \def\new@command#1{%
2453 \@testopt{\@newcommand#1}0}
2454 \def\@newcommand#1[#2]{%
2455 \@ifnextchar [{\@xargdef#1[#2]}%
2456 \argdef#1[#2]}
2457 \long\def\@argdef#1[#2]#3{%
2458 \@yargdef#1\@ne{#2}{#3}}
2459 \long\def\@xargdef#1[#2][#3]#4{%
2460 \expandafter\def\expandafter#1\expandafter{%
2461 \expandafter\@protected@testopt\expandafter #1%
2462 \csname\string#1\expandafter\endcsname{#3}}%
2463 \expandafter\@yargdef \csname\string#1\endcsname
2464 \tw@{#2}{#4}}
2465 \long\def\@yargdef#1#2#3{%
2466 \@tempcnta#3\relax
2467 \advance \@tempcnta \@ne
2468 \let\@hash@\relax
2469 \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
2470 \@tempcntb #2%
2471 \@whilenum \@tempcntb < \@tempcnta
2472 \do{%
2473 \edef\reserved@a{\reserved@a\@hash@\the\@tempcntb}%
2474 \advance\@tempcntb \@ne}%
2475 \let\@hash@##%
2476 \l@ngrel@x\expandafter\def\expandafter#1\reserved@a}
2477 \def\providecommand{\@star@or@long\provide@command}
2478 \def\provide@command#1{%
2479 \begingroup
2480 \escapechar\m@ne\xdef\@gtempa{\string#1}%
2481 \endgroup
2482 \expandafter\@ifundefined\@gtempa
2483 {\def\reserved@a{\new@command#1}}%
2484 {\let\reserved@a\relax
2485 \def\reserved@a{\new@command\reserved@a}}%
2486 \reserved@a}%
2487 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
2488 \def\declare@robustcommand#1{%
2489 \edef\reserved@a{\string#1}%
2490 \def\reserved@b{#1}%
2491 \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
2492 \edef#1{%
2493 \ifx\reserved@a\reserved@b
2494 \noexpand\x@protect
2495 \noexpand#1%
2496 \fi
2497 \noexpand\protect
2498 \expandafter\noexpand\csname

```

```

2499     \expandafter\@gobble\string#1 \endcsname
2500   }%
2501   \expandafter\new@command\csname
2502     \expandafter\@gobble\string#1 \endcsname
2503 }
2504 \def\x@protect#1{%
2505   \ifx\protect\@typeset@protect\else
2506     \@x@protect#1%
2507   \fi
2508 }
2509 \def\@x@protect#1\fi#2#3{%
2510   \fi\protect#1%
2511 }

```

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`.

```

2512 \def\bbl@tempa{\csname newif\endcsname\ifin@}
2513 \ifx\in@\@undefined
2514   \def\in@#1#2{%
2515     \def\in@##1#1##2##3\in@{%
2516       \ifx\in@##2\in@false\else\in@true\fi}%
2517     \in@##2#1\in@\in@}
2518 \else
2519   \let\bbl@tempa\@empty
2520 \fi
2521 \bbl@tempa

```

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

```

2522 \def\@ifpackagewith#1#2#3#4{#3}

```

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

```

2523 \def\@ifl@aded#1#2#3#4{}

```

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

```

2524 \ifx\@tempcnta\@undefined
2525   \csname newcount\endcsname\@tempcnta\relax
2526 \fi
2527 \ifx\@tempcntb\@undefined
2528   \csname newcount\endcsname\@tempcntb\relax
2529 \fi

```

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

```

2530 \ifx\bye\@undefined

```

```

2531 \advance\count10 by -2\relax
2532 \fi
2533 \ifx\@ifnextchar\@undefined
2534 \def\@ifnextchar#1#2#3{%
2535   \let\reserved@#1%
2536   \def\reserved@a{#2}\def\reserved@b{#3}%
2537   \futurelet\@let@token\@ifnch}
2538 \def\@ifnch{%
2539   \ifx\@let@token\@sptoken
2540     \let\reserved@c\@xifnch
2541   \else
2542     \ifx\@let@token\reserved@d
2543       \let\reserved@c\reserved@a
2544     \else
2545       \let\reserved@c\reserved@b
2546     \fi
2547   \fi
2548   \reserved@c}
2549 \def\:\let\@sptoken= } \: % this makes \@sptoken a space token
2550 \def\:\@xifnch \expandafter\def\:\futurelet\@let@token\@ifnch}
2551 \fi
2552 \def\@testopt#1#2{%
2553   \@ifnextchar[#{1}{#1[#2]}}
2554 \def\@protected@testopt#1{%
2555   \ifx\protect\@typeset@protect
2556     \expandafter\@testopt
2557   \else
2558     \@x@protect#1%
2559   \fi}
2560 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
2561   #2\relax}\fi}
2562 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
2563   \else\expandafter\@gobble\fi{#1}}

```

### 13.4 Encoding related macros

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

```

2564 \def\DeclareTextCommand{%
2565   \@dec@text@cmd\providecommand
2566 }
2567 \def\ProvideTextCommand{%
2568   \@dec@text@cmd\providecommand
2569 }
2570 \def\DeclareTextSymbol#1#2#3{%
2571   \@dec@text@cmd\chardef#1{#2}#3\relax
2572 }
2573 \def\@dec@text@cmd#1#2#3{%
2574   \expandafter\def\expandafter#2%
2575     \expandafter{%
2576       \csname#3-cmd\expandafter\endcsname
2577       \expandafter#2%
2578       \csname#3\string#2\endcsname
2579     }%
2580 % \let\@ifdefinable\@rc@ifdefinable
2581 \expandafter#1\csname#3\string#2\endcsname

```



```

2582 }
2583 \def\@current@cmd#1{%
2584   \ifx\protect\@typeset@protect\else
2585     \noexpand#1\expandafter\@gobble
2586   \fi
2587 }
2588 \def\@changed@cmd#1#2{%
2589   \ifx\protect\@typeset@protect
2590     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
2591       \expandafter\ifx\csname ?\string#1\endcsname\relax
2592         \expandafter\def\csname ?\string#1\endcsname{%
2593           \@changed@x@err{#1}%
2594         }%
2595     \fi
2596     \global\expandafter\let
2597       \csname\cf@encoding \string#1\expandafter\endcsname
2598       \csname ?\string#1\endcsname
2599     \fi
2600     \csname\cf@encoding\string#1%
2601     \expandafter\endcsname
2602   \else
2603     \noexpand#1%
2604   \fi
2605 }
2606 \def\@changed@x@err#1{%
2607   \errhelp{Your command will be ignored, type <return> to proceed}%
2608   \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
2609 \def\DeclareTextCommandDefault#1{%
2610   \DeclareTextCommand#1?%
2611 }
2612 \def\ProvideTextCommandDefault#1{%
2613   \ProvideTextCommand#1?%
2614 }
2615 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
2616 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
2617 \def\DeclareTextAccent#1#2#3{%
2618   \DeclareTextCommand#1{#2}[1]{\accent#3 ##1}
2619 }
2620 \def\DeclareTextCompositeCommand#1#2#3#4{%
2621   \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
2622   \edef\reserved@b{\string##1}%
2623   \edef\reserved@c{%
2624     \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
2625   \ifx\reserved@b\reserved@c
2626     \expandafter\expandafter\expandafter\ifx
2627       \expandafter\@car\reserved@a\relax\relax\@nil
2628     \@text@composite
2629   \else
2630     \edef\reserved@b##1{%
2631       \def\expandafter\noexpand
2632         \csname#2\string#1\endcsname###1{%
2633           \noexpand\@text@composite
2634             \expandafter\noexpand\csname#2\string#1\endcsname
2635             ###1\noexpand\@empty\noexpand\@text@composite
2636             {##1}%
2637         }%

```

```

2638         }%
2639         \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
2640     \fi
2641     \expandafter\def\csname\expandafter\string\csname
2642     #2\endcsname\string#1-\string#3\endcsname{#4}
2643 \else
2644     \errhelp{Your command will be ignored, type <return> to proceed}%
2645     \errmessage{\string\DeclareTextCompositeCommand\space used on
2646     inappropriate command \protect#1}
2647 \fi
2648 }
2649 \def\@text@composite#1#2#3\@text@composite{%
2650     \expandafter\@text@composite@x
2651     \csname\string#1-\string#2\endcsname
2652 }
2653 \def\@text@composite@x#1#2{%
2654     \ifx#1\relax
2655         #2%
2656     \else
2657         #1%
2658     \fi
2659 }
2660 %
2661 \def\@strip@args#1:#2-#3\@strip@args{#2}
2662 \def\DeclareTextComposite#1#2#3#4{%
2663     \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
2664     \bgroup
2665         \lccode'\@=#4%
2666         \lowercase{%
2667     \egroup
2668         \reserved@a @%
2669     }%
2670 }
2671 %
2672 \def\UseTextSymbol#1#2{%
2673 %     \let\@curr@enc\cf@encoding
2674 %     \@use@text@encoding{#1}%
2675     #2%
2676 %     \@use@text@encoding\@curr@enc
2677 }
2678 \def\UseTextAccent#1#2#3{%
2679 %     \let\@curr@enc\cf@encoding
2680 %     \@use@text@encoding{#1}%
2681 %     #2{\@use@text@encoding\@curr@enc\selectfont#3}%
2682 %     \@use@text@encoding\@curr@enc
2683 }
2684 \def\@use@text@encoding#1{%
2685 %     \edef\@encoding{#1}%
2686 %     \xdef\font@name{%
2687 %         \csname\curr@fontshape/\@size\endcsname
2688 %     }%
2689 %     \pickup@font
2690 %     \font@name
2691 %     \@@enc@update
2692 }
2693 \def\DeclareTextSymbolDefault#1#2{%

```

```

2694 \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
2695 }
2696 \def\DeclareTextAccentDefault#1#2{%
2697 \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
2698 }
2699 \def\cf@encoding{OT1}

```

Currently we only use the  $\LaTeX 2_{\epsilon}$  method for accents for those that are known to be made active in *some* language definition file.

```

2700 \DeclareTextAccent{"}{OT1}{127}
2701 \DeclareTextAccent{'}{OT1}{19}
2702 \DeclareTextAccent{^}{OT1}{94}
2703 \DeclareTextAccent{\'}{OT1}{18}
2704 \DeclareTextAccent{\~}{OT1}{126}

```

The following control sequences are used in `babel.def` but are not defined for plain  $\TeX$ .

```

2705 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
2706 \DeclareTextSymbol{\textquotedblright}{OT1}{'\'}
2707 \DeclareTextSymbol{\textquoteleft}{OT1}{'\'}
2708 \DeclareTextSymbol{\textquoteright}{OT1}{'\'}
2709 \DeclareTextSymbol{\i}{OT1}{16}
2710 \DeclareTextSymbol{\ss}{OT1}{25}

```

For a couple of languages we need the  $\LaTeX$ -control sequence `\scriptsize` to be available. Because plain  $\TeX$  doesn't have such a sophisticated font mechanism as  $\LaTeX$  has, we just `\let` it to `\sevenrm`.

```

2711 \ifx\scriptsize\undefined
2712 \let\scriptsize\sevenrm
2713 \fi

```

### 13.5 Babel options

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

```

2714 \let\bbbl@opt@shorthands@nnil
2715 \def\bbbl@ifshorthand#1#2#3{#2}%
2716 \ifx\babeloptionstrings\undefined
2717 \let\bbbl@opt@strings@nnil
2718 \else
2719 \let\bbbl@opt@strings\babeloptionstrings
2720 \fi
2721 \def\bbbl@tempa{normal}
2722 \ifx\babeloptionmath\bbbl@tempa
2723 \def\bbbl@mathnormal{\noexpand\textormath}
2724 \fi
2725 \def\BabelStringsDefault{generic}
2726 \ifx\BabelModifiers\undefined\let\BabelModifiers\relax\fi
2727 \let\bbbl@afterlang\relax
2728 \let\bbbl@language@opts@empty
2729 \ifx@uclclist\undefined\let@uclclist@empty\fi
2730 \def\AfterBabelLanguage#1#2{}

```

2731 </plain>

## 14 Tentative font handling

A general solution is far from trivial:

- `\addfontfeature` only sets it for the current family and it's not very efficient, and
- `\defaultfontfeatures` requires to redefine the font (and the options aren't "orthogonal").

```
2732 <<*Font selection>> ≡
2733 \def\babelFSstore#1{%
2734   \bbl@for\bbl@tempa{#1}{%
2735     \edef\bbl@tempb{\noexpand\bbl@FSstore{\bbl@tempa}}
2736     \bbl@tempb{rm}\rmdefault\bbl@save@rmdefault
2737     \bbl@tempb{sf}\sfdefault\bbl@save@sfdefault
2738     \bbl@tempb{tt}\ttdefault\bbl@save@ttdefault}}
2739 \def\bbl@FSstore#1#2#3#4{%
2740   \bbl@csarg\edef{#2default#1}{#3}%
2741   \expandafter\addto\csname extras#1\endcsname{%
2742     \let#4#3%
2743     \ifx#3\f@family
2744       \edef#3{\csname bbl@#2default#1\endcsname}%
2745       \fontfamily{#3}\selectfont
2746     \else
2747       \edef#3{\csname bbl@#2default#1\endcsname}%
2748       \fi}%
2749   \expandafter\addto\csname noextras#1\endcsname{%
2750     \ifx#3\f@family
2751       \fontfamily{#4}\selectfont
2752     \fi
2753     \let#3#4}}
2754 \let\bbl@langfeatures\@empty
2755 \def\babelFSfeatures{%
2756   \let\bbl@ori@fontspec\fontspec
2757   \renewcommand\fontspec[1][]{%
2758     \bbl@ori@fontspec[\bbl@langfeatures##1]}
2759   \let\babelFSfeatures\bbl@FSfeatures
2760   \babelFSfeatures}
2761 \def\bbl@FSfeatures#1#2{%
2762   \expandafter\addto\csname extras#1\endcsname{%
2763     \babel@save\bbl@langfeatures
2764     \edef\bbl@langfeatures{#2,}}}}
2765 <</Font selection>>
```

## 15 Hooks for XeTeX and LuaTeX

### 15.1 XeTeX

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

$\LaTeX$  sets many "codes" just before loading `hyphen.cfg`. That is not a problem in `luatex`, but in `xetex` they must be reset to the proper value. Most of the work is

done in `xe(la)tex.ini`, so here we just “undo” some of the changes done by `LATEX`.  
 Anyway, for consistency `LuaTEX` also resets the catcodes.

```

2766 <<{*Restore Unicode catcodes before loading patterns}>> ≡
2767 \AddBabelHook{xetex}{loadkernel}{%
2768   \begingroup
2769     % Reset chars "80-"C0 to category "other", no case mapping:
2770     \catcode'\@=11 \count@=128
2771     \loop\ifnum\count@<192
2772       \global\uccode\count@=0 \global\lccode\count@=0
2773       \global\catcode\count@=12 \global\sfcode\count@=1000
2774       \advance\count@ by 1 \repeat
2775     % Other:
2776     \def\0 ##1 {%
2777       \global\uccode"##1=0 \global\lccode"##1=0
2778       \global\catcode"##1=12 \global\sfcode"##1=1000 }%
2779     % Letter:
2780     \def\L ##1 ##2 ##3 {\global\catcode"##1=11
2781       \global\uccode"##1="##2
2782       \global\lccode"##1="##3
2783       % Uppercase letters have sfcode=999:
2784       \ifnum"##1="##3 \else \global\sfcode"##1=999 \fi }%
2785     % Letter without case mappings:
2786     \def\l ##1 {\L ##1 ##1 ##1 }%
2787     \l 00AA
2788     \L 00B5 039C 00B5
2789     \l 00BA
2790     \0 00D7
2791     \l 00DF
2792     \0 00F7
2793     \L 00FF 0178 00FF
2794   \endgroup
2795   \input #1\relax}
2796 <</Restore Unicode catcodes before loading patterns}>>

```

Now, the code.

```

2797 <{*xetex}
2798 \def\BabelStringsDefault{unicode}
2799 \let\xebbl@stop\relax
2800 \AddBabelHook{xetex}{encodedcommands}{%
2801   \def\bbl@tempa{#1}%
2802   \ifx\bbl@tempa\@empty
2803     \XeTeXinputencoding"bytes"%
2804   \else
2805     \XeTeXinputencoding"#1"%
2806   \fi
2807   \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
2808 \AddBabelHook{xetex}{stopcommands}{%
2809   \xebbl@stop
2810   \let\xebbl@stop\relax}
2811 <<Restore Unicode catcodes before loading patterns}>>
2812 <<Font selection}>>
2813 </xetex>

```

## 15.2 LuaTeX

This part relies on the lua strips in `luatex-hyphen` by Khaled Hosny, Élie Roux, and Manuel Pégourié-Gonnard. Élie also improved the code below.

```
2814 (*luatex)
2815 \begingroup
2816 \catcode'\%=12
2817 \catcode'\'=12
2818 \catcode'\ "=12
2819 \catcode'\:=12
2820 \directlua{
2821   require("luatex-hyphen")
2822   Babel = {}
2823   function Babel.bytes(line)
2824     return line:gsub(".",
2825       function (chr) return unicode.utf8.char(string.byte(chr)) end)
2826   end
2827   function Babel.begin_process_input()
2828     if luatexbase and luatexbase.add_to_callback then
2829       luatexbase.add_to_callback('process_input_buffer',
2830         Babel.bytes, 'Babel.bytes')
2831     else
2832       Babel.callback = callback.find('process_input_buffer')
2833       callback.register('process_input_buffer', Babel.bytes)
2834     end
2835   end
2836   function Babel.end_process_input ()
2837     if luatexbase and luatexbase.remove_from_callback then
2838       luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
2839     else
2840       callback.register('process_input_buffer', Babel.callback)
2841     end
2842   end
2843   function Babel.addpatterns(pp, lg)
2844     local lg = lang.new(lg)
2845     local pats = lang.patterns(lg) or ''
2846     lang.clear_patterns(lg)
2847     for p in pp:gmatch('[^s]+') do
2848       ss = ''
2849       for i in string.utfcharacters(p:gsub('%d', '')) do
2850         ss = ss .. '%d?' .. i
2851       end
2852       ss = ss:gsub('^%d%?%.', '%%.') .. '%d?'
2853       ss = ss:gsub('%.%d%?$', '%%.')
2854       pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
2855       if n == 0 then
2856         tex.sprint(
2857           [[\string\csname\space bbl@info\endcsname{New pattern: }]]
2858           .. p .. [[{}]])
2859         pats = pats .. ' ' .. p
2860       else
2861         tex.sprint(
2862           [[\string\csname\space bbl@info\endcsname{Renew pattern: }]]
2863           .. p .. [[{}]])
2864       end
2865     end
2866   end
end
```

```

2866   lang.patterns(lg, pats)
2867 end
2868 }
2869 \endgroup
2870 \def\BabelStringsDefault{unicode}
2871 \let\luabbl@stop\relax
2872 \AddBabelHook{luatex}{encodedcommands}{%
2873   \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
2874   \ifx\bbl@tempa\bbl@tempb\else
2875     \directlua{Babel.begin_process_input()}%
2876     \def\luabbl@stop{%
2877       \directlua{Babel.end_process_input()}}%
2878   \fi}%
2879 \AddBabelHook{luatex}{stopcommands}{%
2880   \luabbl@stop
2881   \let\luabbl@stop\relax}
2882 \AddBabelHook{luatex}{patterns}{%
2883   \ifcsname lu@texhyphen@loaded@the\language\endcsname \else
2884     \global\@namedef{lu@texhyphen@loaded@the\language}{}%
2885     \directlua{
2886       luatexhyphen.loadlanguage([[string#1]],
2887         '\the\language')}%
2888   \fi
2889   \@ifundefined{bbl@patterns@}{}%
2890   \begingroup
2891     \@expandtwoargs\in@{,\number\language,}{,\bbl@pttnlist}%
2892     \ifin@else
2893       \ifx\bbl@patterns@\@empty\else
2894         \directlua{ Babel.addpatterns(
2895           [[\bbl@patterns@]], \number\language) }%
2896       \fi
2897       \@ifundefined{bbl@patterns@#1}%
2898       \@empty
2899       {\directlua{ Babel.addpatterns(
2900         [[\space\csname bbl@patterns@#1\endcsname]],
2901         \number\language) }}%
2902       \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
2903     \fi
2904   \endgroup}}
2905 \AddBabelHook{luatex}{adddialect}{%
2906   \ifx\directlua\relax\else
2907     \directlua{
2908       luatexhyphen.adddialect([[string#1]],
2909         [[string#2]])
2910     }%
2911   \fi}
2912 \AddBabelHook{luatex}{everylanguage}{%
2913   \directlua{
2914     processnow = (tex.language == 0) or
2915       (luatexhyphen.lookupname([[string#1]]) == nil)}%
2916   \ifnum0=\directlua{tex.sprint(processnow and "0" or "1")}\relax
2917     \global\@namedef{lu@texhyphen@loaded@the\language}{}%
2918   \fi}
2919 \AddBabelHook{luatex}{loadpatterns}{%
2920   \ifnum0=\directlua{tex.sprint(processnow and "0" or "1")}\relax
2921     \input #1\relax

```

```

2922 \fi}
2923 \AddBabelHook{luatex}{loadexceptions}{%
2924 \ifnum0=\directlua{tex.sprint(processnow and "0" or "1")}\relax
2925 \input #1\relax
2926 \fi
2927 \directlua{processnow = nil}}

```

`\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.

```

2928 \onlypreamble\babelpatterns
2929 \AtEndOfPackage{%
2930 \newcommand\babelpatterns[2][\@empty]{%
2931 \ifx\bbl@patterns@\relax
2932 \let\bbl@patterns@\@empty
2933 \fi
2934 \ifx\bbl@pttnlist\@empty\else
2935 \bbl@warning{%
2936 You must not intermingle \string\selectlanguage\space and\\%
2937 \string\babelpatterns\space or some patterns will not\\%
2938 be taken into account. Reported}%
2939 \fi
2940 \ifx\@empty#1%
2941 \protected@edef\bbl@patterns@{\bbl@patterns@\space#2}%
2942 \else
2943 \edef\bbl@tempb{\zap@space#1 \@empty}%
2944 \bbl@for\bbl@tempa\bbl@tempb{%
2945 \bbl@fixname\bbl@tempa
2946 \bbl@iflanguage\bbl@tempa{%
2947 \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
2948 \@ifundefined{bbl@patterns@\bbl@tempa}%
2949 \@empty
2950 {\csname bbl@patterns@\bbl@tempa\endcsname\space}%
2951 #2}}}%
2952 \fi}}

```

Common stuff.

```

2953 <<Restore Unicode catcodes before loading patterns>>
2954 <<Font selection>>
2955 </luatex>

```

## 16 Conclusion

A system of document options has been presented that enable the user of L<sup>A</sup>T<sub>E</sub>X to adapt the standard document classes of L<sup>A</sup>T<sub>E</sub>X to the language he or she prefers to use. These options offer the possibility of switching between languages in one document. The basic interface consists of using one option, which is the same for *all* standard document classes.

In some cases the language definition files provide macros that can be useful to plain T<sub>E</sub>X users as well as to L<sup>A</sup>T<sub>E</sub>X users. The babel system has been implemented so that it can be used by both groups of users.



## 17 Acknowledgements

I would like to thank all who volunteered as  $\beta$ -testers for their time. I would like to mention Julio Sanchez who supplied the option file for the Spanish language and Maurizio Codogno who supplied the option file for the Italian language. 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] Donald E. Knuth, *The T<sub>E</sub>Xbook*, Addison-Wesley, 1986.
- [2] Leslie Lamport, *L<sup>A</sup>T<sub>E</sub>X, A document preparation System*, Addison-Wesley, 1986.
- [3] K.F. Treebus. *Tekstwijzer, een gids voor het grafisch verwerken van tekst*. SDU Uitgeverij ('s-Gravenhage, 1988). A Dutch book on layout design and typography.
- [4] Hubert Partl, *German T<sub>E</sub>X, TUGboat* 9 (1988) #1, p. 70–72.
- [5] Leslie Lamport, in: *T<sub>E</sub>Xhax Digest, Volume 89, #13*, 17 February 1989.
- [6] Johannes Braams, Victor Eijkhout and Nico Poppelier, *The development of national L<sup>A</sup>T<sub>E</sub>X styles, TUGboat* 10 (1989) #3, p. 401–406.
- [7] Joachim Schrod, *International L<sup>A</sup>T<sub>E</sub>X is ready to use, TUGboat* 11 (1990) #1, p. 87–90.
- [8] Apostolos Syropoulos, Antonis Tsolomitis and Nick Sofroniu, *Digital typography using L<sup>A</sup>T<sub>E</sub>X*, Springer, 2002, p. 301–373.
- [9] Yannis Haralambous, *Fonts & Encodings*, O'Reilly, 2007.