Babel

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, xetex and luatex to provide an environment in which documents can be typeset in a language other than US English, or in more than one language or script.

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

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

I  User guide 4

1  The user interface 4

1.1  Monolingual documents 4
1.2  Multilingual documents 5
1.3  Modifiers 6
1.4  xelatex and lualatex 7
1.5  Troubleshooting 7
1.6  Plain 8
1.7  Basic language selectors 8
1.8  Auxiliary language selectors 9
1.9  More on selection 10
1.10  Shorthands 11
1.11  Package options 14
1.12  The base option 16
1.13  ini files 17
1.14  Selecting fonts 24
1.15  Modifying a language 25
1.16  Creating a language 26
1.17  Digits 29
1.18  Getting the current language name 29
1.19  Hyphenation and line breaking 29
1.20  Selecting scripts 31
1.21  Selecting directions 32
1.22  Language attributes 36
1.23  Hooks 36
1.24  Languages supported by babel with ldf files 37
1.25  Tips, workarounds, know issues and notes 38
1.26  Current and future work 40
1.27  Tentative and experimental code 40

2  Loading languages with language.dat 40

2.1  Format 41

3  The interface between the core of babel and the language definition files 41

3.1  Guidelines for contributed languages 43
3.2  Basic macros 43
3.3  Skeleton 44
3.4  Support for active characters 45
3.5  Support for saving macro definitions 46
3.6  Support for extending macros 46
3.7  Macros common to a number of languages 46
3.8  Encoding-dependent strings 47

4  Changes 50

4.1  Changes in babel version 3.9 50

II  Source code 51

5  Identification and loading of required files 51

6  locale directory 51
14 Hooks for XeTeX and LuaTeX

14.1 XeTeX ................................................. 139
14.2 Layout ............................................... 141
14.3 LuaTeX ............................................... 144
14.4 Southeast Asian scripts ............................... 149
14.5 CJK line breaking .................................... 152
14.6 Layout ............................................... 152
14.7 Auto bidi with basic and basic-r ................. 154

15 Data for CJK ........................................... 165

16 The 'nil' language .................................... 165

17 Support for Plain TeX (plain.def) .................. 165

17.1 Not renaming hyphen.tex ........................... 165
17.2 Emulating some LmTeX features ..................... 167
17.3 General tools ........................................ 167
17.4 Encoding related macros ............................ 171

18 Acknowledgements ..................................... 173

Troubleshooting

<table>
<thead>
<tr>
<th>Error Message</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paragraph ended before \UTFviii@three@octets was complete</td>
<td>4</td>
</tr>
<tr>
<td>No hyphenation patterns were preloaded for (babel) the language 'LANG' into</td>
<td>5</td>
</tr>
<tr>
<td>the format</td>
<td></td>
</tr>
<tr>
<td>You are loading directly a language style</td>
<td>7</td>
</tr>
<tr>
<td>Unknown language 'LANG'</td>
<td>8</td>
</tr>
<tr>
<td>Argument of \language@active@arg&quot; has an extra }</td>
<td>11</td>
</tr>
<tr>
<td>Package fontspec Warning: 'Language 'LANG' not available for font 'FONT' with</td>
<td>25</td>
</tr>
<tr>
<td>script 'SCRIPT' 'Default' language used instead</td>
<td></td>
</tr>
</tbody>
</table>
Part I

User guide

• This user guide focuses on \LaTeX. There are also some notes on its use with Plain \LaTeX.

• Changes and new features with relation to version 3.8 are highlighted with $\text{New } XXX$. The most recent features could be still unstable. Please, report any issues you find on https://github.com/latex3/babel/issues, which is better than just complaining on an e-mail list or a web forum.

• If you are interested in the \TeX multilingual support, please join the kadingira list on http://tug.org/mailman/listinfo/kadingira. You can follow the development of babel on https://github.com/latex3/babel (which provides some sample files, too).

• See section 3.1 for contributing a language.

• The first sections describe the traditional way of loading a language (with .lf files). The alternative way based on ini files, which complements the previous one (it will not replace it), is described below.

1 The user interface

1.1 Monolingual documents

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

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

\documentclass{article}
\usepackage[T1]{fontenc}
\usepackage[utf8]{inputenc}
\usepackage[french]{babel}
\begin{document}
Plus ça change, plus c'est la même chose!
\end{document}

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

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

Make sure you set the encoding actually used by your editor.
Another approach is making the language (french in the example) a global option in order to let other packages detect and use it:

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

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

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

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

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

The document will be typeset, but very likely the text will not be correctly hyphenated. Some languages may be raising this warning wrongly (because they are not hyphenated); it is a bug to be fixed — just ignore it. See the manual of your distribution (Mac\TeX, Mi\TeX, \TeX\Live, etc.) for further info about how to configure it.

### 1.2 Multilingual documents

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

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

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

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

You can also set the main language explicitly:

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

**NOTE** Some classes load babel with a hardcoded language option. Sometimes, the main language could be overridden with something like that before \documentclass:
WARNING Languages may be set as global and as package option at the same time, but in such a case you should set explicitly the main language with the package option `main`:

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

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

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

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

\documentclass{article}
\usepackage[T1]{fontenc}
\usepackage[utf8]{inputenc}
\usepackage[english,french]{babel}
\begin{document}

Plus ça change, plus c'est la même chose!
\selectlanguage{english}

And an English paragraph, with a short text in \foreignlanguage{french}{français}.
\end{document}

### 1.3 Modifiers

**New 3.9c** The basic behavior of some languages can be modified when loading babel by means of *modifiers*. They are set after the language name, and are prefixed with a dot (only when the language is set as package option – neither global options nor the `main` key accept them). An example is (spaces are not significant and they can be added or removed):\footnote{No predefined “axis” for modifiers are provided because languages and their scripts have quite different needs.}

\usepackage[latin.medieval, spanish.notilde.roman, 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.
1.4 *xelatex and lualatex*

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

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

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

\begin{verbatim}
\documentclass{article}
\usepackage[vietnamese,danish]{babel}
\begin{document}
\prefacename{} -- \alsoname{} -- \today
\selectlanguage{vietnamese}
\prefacename{} -- \alsoname{} -- \today
\end{document}
\end{verbatim}

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

\begin{verbatim}
\documentclass{article}
\usepackage[russian]{babel}
\babelfont{rm}{DejaVu Serif}
\begin{document}
Россия, находящаяся на пересечении множества культур, а также с учётом многонационального характера её населения, — отличается высокой степенью этнокультурного многообразия и способностью к межкультурному диалогу.
\end{document}
\end{verbatim}

1.5 **Troubleshooting**

- Loading directly sty files in \texttt{\LaTeX} (i.e, \texttt{\usepackage{\langle language\rangle}}) is deprecated and you will get the error\footnote{In old versions the error read “You have used an old interface to call babel”, not very helpful.}:

  \begin{verbatim}
  ! Package babel Error: You are loading directly a language style.
  (babel) This syntax is deprecated and you must use
  (babel) \usepackage[language]{babel}.
  \end{verbatim}
Another typical error when using babel is the following:\footnote{In old versions the error read “You haven’t loaded the language LANG yet”.
Even in the babel kernel there were some macros not compatible with plain. Hopefully these issues have been fixed.}

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

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

### 1.6 Plain

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

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

**WARNING** Not all languages provide a sty file and some of them are not compatible with Plain.\footnote{In old versions the error read “You haven’t loaded the language LANG yet”.
Even in the babel kernel there were some macros not compatible with plain. Hopefully these issues have been fixed.}

### 1.7 Basic language selectors

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

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

```
\selectlanguage{⟨language⟩}
```

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

```
\selectlanguage{german}
```

This command can be used as environment, too.

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

**WARNING** If used inside braces there might be some non-local changes, as this would be roughly equivalent to:
If you want a change which is really local, you must enclose this code with an additional grouping level.

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

1.8 Auxiliary language selectors

\begin{otherlanguage} {⟨language⟩} ... \end{otherlanguage}
The environment otherlanguage does basically the same as \selectlanguage, except the language change is (mostly) local to the environment. Actually, there might be some non-local changes, as this environment is roughly equivalent to:

\begin{group}
\selectlanguage{⟨inner-language⟩}
...
\end{group}
\selectlanguage{⟨outer-language⟩}

If you want a change which is really local, you must enclose this environment with an additional grouping, like braces \{\}. Spaces after the environment are ignored.

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

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

1.9 More on selection

\babetags \{tag1\} = \{language1\}, \{tag2\} = \{language2\},...

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

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

**EXAMPLE** With

\babetags{de = german}

you can write

\text \textde{German text} \text

and

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

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

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

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

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

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

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

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

\bcelensure[include=\Today]{spanish}
\bcelensure[fontenc=TS]{vietnamese}

They are activated when the language is selected (at the after extras 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 of \dag).

With ini files (see below), captions are ensured by default.

1.10 Shorthands

A shorthand is a sequence of one or two characters that expands to arbitrary \TeX code.
Shorthands can be used for different kinds of things, as for example: (1) in some languages
shorthands such as "a are defined to be able to hyphenate the word if the encoding is OT1;
(2) in some languages shorthands such as ! are used to insert the right amount of white
space; (3) several kinds of 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 on the keyboards (and sometimes not even as separated or precomposed
Unicode characters). As to the point 2, now pdf\TeX provides \knbcode, and luatex can
manipulate the glyph list. Tools for point 3 can be still very useful in general.

There are three levels of shorthands: user, language, and system (by order of precedence).
Version 3.9 introduces the language user level on top of the user level, as described below.
In most cases, you will use only shorthands provided by languages.

**NOTE** Note the following:

1. Activated chars used for two-char shorthands cannot be followed by a closing brace
   } and the spaces following are gobbled. With one-char shorthands (eg, '.'), they are
   preserved.

2. If on a certain level (system, language, user) there is a one-char shorthand, two-char
   ones starting with that char and on the same level are ignored.

3. Since they are active, a shorthand cannot contain the same character in its definition
   (except if it is deactivated with, eg, string).

A typical error when using shorthands is the following:

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

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

\shorthandon \{\shorthandlist\}
\shorthandoff  *{(shorthand-list)}

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

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

\shorthandoff*{-^}

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

\useshorthands  *{(char)}

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

New 3.9a User shorthands are not always alive, as they may be deactivated by languages
(for example, if you use * for your user shorthands and switch from german to french, they
stop working). Therefore, a starred version \useshorthands*{(char)} is provided, which
makes sure shorthands are always activated.
Currently, if the package option shorthands is used, you must include any character to be
activated with \useshorthands. This restriction will be lifted in a future release.

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

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

New 3.9a An optional argument allows to (re)define language and system shorthands
(some languages do not activate shorthands, so you may want to add
\languageshorthands{(lang)} to the corresponding \extras{lang}, as explained below).
By default, user shorthands are (re)defined.
User shorthands override language ones, which in turn override system shorthands.
Language-dependent user shorthands (new in 3.9) take precedence over “normal” user
shorthands.

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

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

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

5 With it encoded string may not work as expected.
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.

\texttt{\textbackslash aliasshorthand \{\langle original\rangle\}\{\langle alias\rangle\}}

The command \texttt{\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 \texttt{\aliasshorthand{"}{/}}.

\textbf{NOTE} The substitute character must not have been declared before as shorthand (in such a case, \texttt{\aliasshorthand} is ignored).

\textbf{EXAMPLE} The following example shows how to replace a shorthand by another

\begin{verbatim}
\aliasshorthand{-}{^}
\AtBeginDocument{\shorthandoff{-}}
\end{verbatim}

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

\texttt{\textbackslash languageshorthands \{\langle language\rangle\}}

The command \texttt{\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)\footnote{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

\begin{verbatim}
\addto\extrasenglish{\languageshorthands{ngerman}}
\end{verbatim}

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

\begin{verbatim}
\newcommand{\myipa}[1]{\{\languageshorthands{none}\texttt{\tipaencoding\#1}}
\end{verbatim}
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, \belflashand{"u} or \belflashand{:}. (You can conveniently define your own macros, or even your own user shorthands provided they do not overlap.)

For your records, here is a list of shorthands, but you must double check them, as they may change.

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

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

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

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

**ifbelflashand** \{(character)\}\{\{true\}\}\{\{false\}\}

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

### 1.11 Package options

**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.

---

*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.*

*Thanks to Enrico Gregorio*  

*This declaration serves to nothing, but it is preserved for backward compatibility.*
activegrave  Same for `.

shorthands=  ⟨char⟩⟨char⟩... | off

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

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

If ` is included, activeacute is set; if ` is included, activegrave is set. Active characters (like ~) should be preceded by \string (otherwise they will be expanded by \LaTeX 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 \LaTeX macros are redefined so that using shorthands is safe. With safe=bib only \nocite, \bibcite and \bibitem are redefined. With safe=ref only \newlabel, \ref and \pageref are redefined (as well as a few macros from varioref and ifthen). With safe=none no macro is redefined. This option is strongly recommended, because a good deal of incompatibilities and errors are related to these redefinitions – of course, in such a case you cannot use shorthands in these macros, but this is not a real problem (just use “allowed” characters).

math=  active | normal

Shorthands are mainly intended for text, not for math. By setting this option with the value normal they are deactivated in math mode (default is active) and things like $\{a\}$$ (a closing brace after a shorthand) are not a source of trouble any more.

config=  ⟨file⟩

Load ⟨file⟩.cfg instead of the default config file bblopts.cfg (the file is loaded even with noconfigs).

main=  ⟨language⟩

Sets the main language, as explained above, ie, this language is always loaded last. If it is not given as package or global option, it is added to the list of requested languages.

headfoot=  ⟨language⟩

By default, headlines and footlines are not touched (only marks), and if they contain language dependent macros (which is not usual) there may be unexpected results. With this option you may set the language in heads and foots.

noconfigs  Global and language default config files are not loaded, so you can make sure your document is not spoil 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 3.9i  Language settings for uppercase and lowercase mapping (as set by \SetCase) are ignored. Use only if there are incompatibilities with other packages.

**silent**  New 3.9i  No warnings and no *infos* are written to the log file. You can use alternatively the package silence.

**strings=**  New 3.9i  Selects the encoding of strings in languages supporting this feature. Predefined labels are *generic* (for traditional \TeX\, LIC and ASCII strings), *unicode* (for engines like xetex and luatex) and *encoded* (for special cases requiring mixed encodings). Other allowed values are font encoding codes (T1, T2A, LGR, L7X...), but only in languages supporting them. Be aware with encoded captions are protected, but they work in \MakeUpperCase and the like (this feature misuses some internal \TeX\ tools, so use it only as a last resort).

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

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

- **off**  deactivates this feature and no case mapping is applied;
- **first**  sets it at the first switching commands in the current or parent scope (typically, when the aux file is first read and at \begin{document}, but also the first \selectlanguage in the preamble), and it's the default if a single language option has been stated.
- **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.

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

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

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

### 1.12 The base option

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

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

---

*You can use alternatively the package silence.
10* Turned off in plain.
11* Duplicated options count as several ones.
12* Providing foreign is pointless, because the case mapping applied is that at the end of paragraph, but if either xetex or luatex change this behavior it might be added. On the other hand, other is provided even if I [JBL] think it isn't really useful, but who knows.
This command is currently the only provided by base. Executes \texttt{⟨code⟩} when the file loaded by the corresponding package option is finished (at \texttt{\ldf@finish}). The setting is global. So does ... at the end of french.1df. It can be used in 1df files, too, but in such a case the code is executed only if \texttt{(option-name)} is the same as \texttt{\CurrentOption} (which could not be the same as the option name as set in \texttt{\usepackage!}).

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

\begin{verbatim}
\usepackage{base}{babel}
\AfterBabelLanguage{foo}{\%
  \let\macroFoo\macro
  \let\macro\relax}
\end{verbatim}

1.13 ini files

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

**EXAMPLE** Although Georgian has its own 1df file, here is how to declare this language with an ini file in Unicode engines.

\begin{verbatim}
\documentclass{book}
\usepackage{babel}
\ babelprovide[import, main]{georgian}
\babelfont{rm}{DejaVu Sans}
\begin{document}
\tableofcontents
\chapter{სამზარეულო და სუფრის ტრადიციები}
ქართული ტრადიციული სამზარეულო ერთ-ერთი უმდიდრესია მთელ მსოფლიოში.
\end{document}
\end{verbatim}

**NOTE** The ini files just define and set some parameters, but the corresponding behavior is not always implemented. Also, there are some limitations in the engines. A few remarks follows:
**Arabic**  Monolingual documents mostly work in luatex, but it must be fine tuned, and a recent version of fontspec/loaotfload is required. In xetex babel resorts to the bidi package, which seems to work.

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

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

\newfontscript{Devanagari}{deva}

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

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

\babelprovide[import,hyphenrules=+]{lao}
\babelpatterns[lao]{1ດ 1ມ 1 1ອ 1ງ 1ກ 1າ} % Random

Khemer clusters are rendered wrongly.

**East Asia scripts**  Internal inconsistencies in script and language names must be sorted out, so you may need to set them explicitly in \babelfont, as well as CJKShape. luatex does basic line breaking, but currently xetex does not (you may load zhspacing). Anyway, CJK texts are are best set with a dedicated framework (CJK, luatexja, kotex, CTeX...), although for a few words and shorts texts babel should be fine.

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

<table>
<thead>
<tr>
<th>af</th>
<th>Afrikaans-ul</th>
<th>bo</th>
<th>Tibetan-u</th>
</tr>
</thead>
<tbody>
<tr>
<td>agq</td>
<td>Aghem</td>
<td>brx</td>
<td>Bodo</td>
</tr>
<tr>
<td>ak</td>
<td>Akan</td>
<td>bs-Cyr</td>
<td>Bosnian-l</td>
</tr>
<tr>
<td>am</td>
<td>Amharic-ul</td>
<td>bs-Latn</td>
<td>Bosnian-ul</td>
</tr>
<tr>
<td>ar</td>
<td>Arabic-ul</td>
<td>bs</td>
<td>Bosnian-ul</td>
</tr>
<tr>
<td>ar-DZ</td>
<td>Arabic-ul</td>
<td>ca</td>
<td>Catalan-ul</td>
</tr>
<tr>
<td>ar-MA</td>
<td>Arabic-ul</td>
<td>ce</td>
<td>Chechen</td>
</tr>
<tr>
<td>ar-SY</td>
<td>Arabic-ul</td>
<td>cgg</td>
<td>Chiga</td>
</tr>
<tr>
<td>as</td>
<td>Assamese</td>
<td>chr</td>
<td>Cherokee</td>
</tr>
<tr>
<td>asa</td>
<td>Asu</td>
<td>ckb</td>
<td>Central Kurdish</td>
</tr>
<tr>
<td>ast</td>
<td>Asturian-ul</td>
<td>cs</td>
<td>Czech-ul</td>
</tr>
<tr>
<td>az-Cyril</td>
<td>Azerbaijani</td>
<td>cy</td>
<td>Welsh-ul</td>
</tr>
<tr>
<td>az-Latn</td>
<td>Azerbaijani</td>
<td>da</td>
<td>Danish-ul</td>
</tr>
<tr>
<td>az</td>
<td>Azerbaijani-ul</td>
<td>dav</td>
<td>Taita</td>
</tr>
<tr>
<td>bas</td>
<td>Basaa</td>
<td>de-AT</td>
<td>German-ul</td>
</tr>
<tr>
<td>be</td>
<td>Belarusian-ul</td>
<td>de-CH</td>
<td>German-ul</td>
</tr>
<tr>
<td>bem</td>
<td>Bemba</td>
<td>de</td>
<td>German-ul</td>
</tr>
<tr>
<td>bez</td>
<td>Bena</td>
<td>dje</td>
<td>Zarma</td>
</tr>
<tr>
<td>bg</td>
<td>Bulgarian-ul</td>
<td>dsb</td>
<td>Lower Sorbian-ul</td>
</tr>
<tr>
<td>bm</td>
<td>Bambara</td>
<td>dua</td>
<td>Duala</td>
</tr>
<tr>
<td>bn</td>
<td>Bangla-ul</td>
<td>dyo</td>
<td>Jola-Fonyi</td>
</tr>
<tr>
<td>Code</td>
<td>Language</td>
<td>Code</td>
<td>Language</td>
</tr>
<tr>
<td>------</td>
<td>--------------</td>
<td>------</td>
<td>--------------</td>
</tr>
<tr>
<td>dz</td>
<td>Dzongkha</td>
<td>ka</td>
<td>Georgian</td>
</tr>
<tr>
<td>ebu</td>
<td>Embu</td>
<td>kab</td>
<td>Kabyle</td>
</tr>
<tr>
<td>ee</td>
<td>Ewe</td>
<td>kam</td>
<td>Kamba</td>
</tr>
<tr>
<td>el</td>
<td>Greek</td>
<td>kde</td>
<td>Makonde</td>
</tr>
<tr>
<td>en-AU</td>
<td>English</td>
<td>kea</td>
<td>Kabuverdianu</td>
</tr>
<tr>
<td>en-CA</td>
<td>English</td>
<td>khq</td>
<td>Koyra Chini</td>
</tr>
<tr>
<td>en-GB</td>
<td>English</td>
<td>ki</td>
<td>Kikuyu</td>
</tr>
<tr>
<td>en-NZ</td>
<td>English</td>
<td>kk</td>
<td>Kazakh</td>
</tr>
<tr>
<td>en-US</td>
<td>English</td>
<td>kkj</td>
<td>Kako</td>
</tr>
<tr>
<td>eo</td>
<td>Esperanto</td>
<td>kl</td>
<td>Kalaallisut</td>
</tr>
<tr>
<td>es-MX</td>
<td>Spanish</td>
<td>kln</td>
<td>Kalenjin</td>
</tr>
<tr>
<td>es</td>
<td>Spanish</td>
<td>km</td>
<td>Khmer</td>
</tr>
<tr>
<td>et</td>
<td>Estonian</td>
<td>kn</td>
<td>Kannada</td>
</tr>
<tr>
<td>eu</td>
<td>Basque</td>
<td>ko</td>
<td>Korean</td>
</tr>
<tr>
<td>ewo</td>
<td>Ewondo</td>
<td>kok</td>
<td>Konkani</td>
</tr>
<tr>
<td>fa</td>
<td>Persian</td>
<td>ks</td>
<td>Kashmiri</td>
</tr>
<tr>
<td>ff</td>
<td>Fulah</td>
<td>ksb</td>
<td>Shambala</td>
</tr>
<tr>
<td>fi</td>
<td>Finnish</td>
<td>ksF</td>
<td>Bafia</td>
</tr>
<tr>
<td>fil</td>
<td>Filipino</td>
<td>ksh</td>
<td>Colognian</td>
</tr>
<tr>
<td>fo</td>
<td>Faroese</td>
<td>kw</td>
<td>Cornish</td>
</tr>
<tr>
<td>fr</td>
<td>French</td>
<td>ky</td>
<td>Kyrgyz</td>
</tr>
<tr>
<td>fr-BE</td>
<td>French</td>
<td>lag</td>
<td>Langi</td>
</tr>
<tr>
<td>fr-CA</td>
<td>French</td>
<td>lb</td>
<td>Luxembourghis</td>
</tr>
<tr>
<td>fr-CH</td>
<td>French</td>
<td>lg</td>
<td>Ganda</td>
</tr>
<tr>
<td>fr-LU</td>
<td>French</td>
<td>lkt</td>
<td>Lakota</td>
</tr>
<tr>
<td>fur</td>
<td>Friulian</td>
<td>ln</td>
<td>Lingala</td>
</tr>
<tr>
<td>fy</td>
<td>Western Frisian</td>
<td>lo</td>
<td>Lao</td>
</tr>
<tr>
<td>ga</td>
<td>Irish</td>
<td>lrc</td>
<td>Northern Luri</td>
</tr>
<tr>
<td>gd</td>
<td>Scottish Gaelic</td>
<td>lt</td>
<td>Lithuanian</td>
</tr>
<tr>
<td>gl</td>
<td>Galician</td>
<td>lu</td>
<td>Luba-Katanga</td>
</tr>
<tr>
<td>gsw</td>
<td>Swiss German</td>
<td>luo</td>
<td>Luo</td>
</tr>
<tr>
<td>gu</td>
<td>Gujarati</td>
<td>luy</td>
<td>Luyia</td>
</tr>
<tr>
<td>guz</td>
<td>Gusii</td>
<td>lv</td>
<td>Latvian</td>
</tr>
<tr>
<td>gv</td>
<td>Manx</td>
<td>mas</td>
<td>Masai</td>
</tr>
<tr>
<td>ha-GH</td>
<td>Hausa</td>
<td>mer</td>
<td>Meru</td>
</tr>
<tr>
<td>ha-NE</td>
<td>Hausa</td>
<td>mfe</td>
<td>Morisyen</td>
</tr>
<tr>
<td>ha</td>
<td>Hausa</td>
<td>mg</td>
<td>Malagasy</td>
</tr>
<tr>
<td>haw</td>
<td>Hawaiian</td>
<td>mgh</td>
<td>Makuwa-Meeto</td>
</tr>
<tr>
<td>he</td>
<td>Hebrew</td>
<td>mgo</td>
<td>Meta</td>
</tr>
<tr>
<td>hi</td>
<td>Hindi</td>
<td>mk</td>
<td>Macedonian</td>
</tr>
<tr>
<td>hr</td>
<td>Croatian</td>
<td>ml</td>
<td>Malayalam</td>
</tr>
<tr>
<td>hsb</td>
<td>Upper Sorbian</td>
<td>mn</td>
<td>Mongolian</td>
</tr>
<tr>
<td>hu</td>
<td>Hungarian</td>
<td>mr</td>
<td>Marathi</td>
</tr>
<tr>
<td>hy</td>
<td>Armenian</td>
<td>ms-BN</td>
<td>Malay</td>
</tr>
<tr>
<td>ia</td>
<td>Interlingua</td>
<td>ms-SG</td>
<td>Malay</td>
</tr>
<tr>
<td>id</td>
<td>Indonesian</td>
<td>ms</td>
<td>Malay</td>
</tr>
<tr>
<td>ig</td>
<td>Igbo</td>
<td>mt</td>
<td>Maltese</td>
</tr>
<tr>
<td>ii</td>
<td>Sichuan Yi</td>
<td>mua</td>
<td>Mundang</td>
</tr>
<tr>
<td>is</td>
<td>Icelandic</td>
<td>my</td>
<td>Burmese</td>
</tr>
<tr>
<td>it</td>
<td>Italian</td>
<td>mzn</td>
<td>Mazanderani</td>
</tr>
<tr>
<td>ja</td>
<td>Japanese</td>
<td>naq</td>
<td>Nama</td>
</tr>
<tr>
<td>jgo</td>
<td>Ngomba</td>
<td>nb</td>
<td>Norwegian Bokmål</td>
</tr>
<tr>
<td>jmc</td>
<td>Machame</td>
<td>nd</td>
<td>North Ndebele</td>
</tr>
<tr>
<td>jmc</td>
<td>Machame</td>
<td>ne</td>
<td>Nepali</td>
</tr>
<tr>
<td>Code</td>
<td>Name</td>
<td>Code</td>
<td>Name</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------</td>
<td>------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>nl</td>
<td>Dutchul</td>
<td>sr-Cyril-BA</td>
<td>Serbianul</td>
</tr>
<tr>
<td>nmg</td>
<td>Kwasio</td>
<td>sr-Cyril-ME</td>
<td>Serbianul</td>
</tr>
<tr>
<td>nn</td>
<td>Norwegian Nynorskul</td>
<td>sr-Cyril-XK</td>
<td>Serbianul</td>
</tr>
<tr>
<td>nnh</td>
<td>Ngiemboon</td>
<td>sr-Latin-BA</td>
<td>Serbianul</td>
</tr>
<tr>
<td>nus</td>
<td>Nuer</td>
<td>sr-Latin-ME</td>
<td>Serbianul</td>
</tr>
<tr>
<td>nyn</td>
<td>Nyankole</td>
<td>sr-Latin-XK</td>
<td>Serbianul</td>
</tr>
<tr>
<td>om</td>
<td>Oromo</td>
<td>sr-Latn</td>
<td>Serbianul</td>
</tr>
<tr>
<td>or</td>
<td>Odia</td>
<td>sr</td>
<td>Serbianul</td>
</tr>
<tr>
<td>os</td>
<td>Ossetic</td>
<td>sv</td>
<td>Swedishul</td>
</tr>
<tr>
<td>pa-Arab</td>
<td>Punjabi</td>
<td>sw</td>
<td>Swahili</td>
</tr>
<tr>
<td>pa-Guru</td>
<td>Punjabi</td>
<td>ta</td>
<td>Tamilul</td>
</tr>
<tr>
<td>pa</td>
<td>Punjabi</td>
<td>te</td>
<td>Teluguul</td>
</tr>
<tr>
<td>pl</td>
<td>Polishul</td>
<td>teo</td>
<td>Teso</td>
</tr>
<tr>
<td>plms</td>
<td>Piedmonteseul</td>
<td>th</td>
<td>Thaiul</td>
</tr>
<tr>
<td>ps</td>
<td>Pashto</td>
<td>ti</td>
<td>Tigrinya</td>
</tr>
<tr>
<td>pt-BR</td>
<td>Portugueseul</td>
<td>tk</td>
<td>Turkmenul</td>
</tr>
<tr>
<td>pt-PT</td>
<td>Portugueseul</td>
<td>to</td>
<td>Tongan</td>
</tr>
<tr>
<td>pt</td>
<td>Portugueseul</td>
<td>tr</td>
<td>Turkishul</td>
</tr>
<tr>
<td>qu</td>
<td>Quechua</td>
<td>twq</td>
<td>Tasawaq</td>
</tr>
<tr>
<td>rm</td>
<td>Romanshul</td>
<td>ug</td>
<td>Uyghur</td>
</tr>
<tr>
<td>rn</td>
<td>Rundi</td>
<td>uk</td>
<td>Ukrainianul</td>
</tr>
<tr>
<td>ro</td>
<td>Romanianul</td>
<td>ur</td>
<td>Urduul</td>
</tr>
<tr>
<td>rof</td>
<td>Rombo</td>
<td>uz-Arab</td>
<td>Uzbek</td>
</tr>
<tr>
<td>ru</td>
<td>Russianul</td>
<td>wae</td>
<td>Walser</td>
</tr>
<tr>
<td>rw</td>
<td>Kinyarwanda</td>
<td>xog</td>
<td>Soga</td>
</tr>
<tr>
<td>rwk</td>
<td>Rwa</td>
<td>yav</td>
<td>Yangben</td>
</tr>
<tr>
<td>sa-Beng</td>
<td>Sanskrit</td>
<td>yi</td>
<td>Yiddish</td>
</tr>
<tr>
<td>sa-Deva</td>
<td>Sanskrit</td>
<td>yo</td>
<td>Yoruba</td>
</tr>
<tr>
<td>sa-Gujr</td>
<td>Sanskrit</td>
<td>yue</td>
<td>Cantonese</td>
</tr>
<tr>
<td>sa-Knda</td>
<td>Sanskrit</td>
<td>zgh</td>
<td>Standard Moroccan</td>
</tr>
<tr>
<td>sa-Mlym</td>
<td>Sanskrit</td>
<td>vun</td>
<td>Vunjo</td>
</tr>
<tr>
<td>sa-Telu</td>
<td>Sanskrit</td>
<td>wae</td>
<td>Walser</td>
</tr>
<tr>
<td>sa</td>
<td>Sanskrit</td>
<td>xog</td>
<td>Soga</td>
</tr>
<tr>
<td>sa</td>
<td>Sakha</td>
<td>yav</td>
<td>Yangben</td>
</tr>
<tr>
<td>saq</td>
<td>Samburu</td>
<td>yi</td>
<td>Yiddish</td>
</tr>
<tr>
<td>sbp</td>
<td>Sangu</td>
<td>yo</td>
<td>Yoruba</td>
</tr>
<tr>
<td>se</td>
<td>Northern Samiul</td>
<td>yue</td>
<td>Cantonese</td>
</tr>
<tr>
<td>seh</td>
<td>Sena</td>
<td>zgh</td>
<td>Standard Moroccan</td>
</tr>
<tr>
<td>ses</td>
<td>Koyraboro Senni</td>
<td>zh-Hans-HK</td>
<td>Chinese</td>
</tr>
<tr>
<td>sg</td>
<td>Sango</td>
<td>zh-Hans-MO</td>
<td>Chinese</td>
</tr>
<tr>
<td>shi-Latn</td>
<td>Tachelhit</td>
<td>zh-Hans-SG</td>
<td>Chinese</td>
</tr>
<tr>
<td>shi-Tfng</td>
<td>Tachelhit</td>
<td>zh-Hans</td>
<td>Chinese</td>
</tr>
<tr>
<td>shi</td>
<td>Tachelhit</td>
<td>zh-Hant-HK</td>
<td>Chinese</td>
</tr>
<tr>
<td>si</td>
<td>Sinhala</td>
<td>zh-Hant-MO</td>
<td>Chinese</td>
</tr>
<tr>
<td>sk</td>
<td>Slovakul</td>
<td>zh-Hant</td>
<td>Chinese</td>
</tr>
<tr>
<td>sl</td>
<td>Slovenianul</td>
<td>zh-Hant-MO</td>
<td>Chinese</td>
</tr>
<tr>
<td>smn</td>
<td>Inari Sami</td>
<td>zh-Han</td>
<td>Chinese</td>
</tr>
<tr>
<td>sn</td>
<td>Shona</td>
<td>zh</td>
<td>Chinese</td>
</tr>
<tr>
<td>so</td>
<td>Somali</td>
<td>zu</td>
<td>Zulu</td>
</tr>
<tr>
<td>sq</td>
<td>Albanianul</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Language</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>aghem</td>
<td>chinese-hans-hk</td>
</tr>
<tr>
<td>akan</td>
<td>chinese-hans-mo</td>
</tr>
<tr>
<td>albanian</td>
<td>chinese-hans-sg</td>
</tr>
<tr>
<td>american</td>
<td>chinese-hans</td>
</tr>
<tr>
<td>amharic</td>
<td>chinese-hant-hk</td>
</tr>
<tr>
<td>arabic</td>
<td>chinese-hant-mo</td>
</tr>
<tr>
<td>arabic-algeria</td>
<td>chinese-hant</td>
</tr>
<tr>
<td>arabic-DZ</td>
<td>chinese-simplified-hongkongsarchina</td>
</tr>
<tr>
<td>arabic-morocco</td>
<td>chinese-simplified-macausarchina</td>
</tr>
<tr>
<td>arabic-MA</td>
<td>chinese-simplified-singapore</td>
</tr>
<tr>
<td>arabic-syria</td>
<td>chinese-simplified</td>
</tr>
<tr>
<td>arabic-SY</td>
<td>chinese-traditional-hongkongsarchina</td>
</tr>
<tr>
<td>armenian</td>
<td>chinese-traditional-macausarchina</td>
</tr>
<tr>
<td>assamese</td>
<td>chinese-traditional</td>
</tr>
<tr>
<td>asturian</td>
<td>chinese</td>
</tr>
<tr>
<td>asu</td>
<td>colognian</td>
</tr>
<tr>
<td>australian</td>
<td>cornish</td>
</tr>
<tr>
<td>austrian</td>
<td>croatian</td>
</tr>
<tr>
<td>azerbaijani-cyrillic</td>
<td>czech</td>
</tr>
<tr>
<td>azerbaijani-cyrl</td>
<td>danish</td>
</tr>
<tr>
<td>azerbaijani-latin</td>
<td>duala</td>
</tr>
<tr>
<td>azerbaijani-latn</td>
<td>dutch</td>
</tr>
<tr>
<td>azerbaijani</td>
<td>dzongkha</td>
</tr>
<tr>
<td>bafia</td>
<td>embu</td>
</tr>
<tr>
<td>bambara</td>
<td>english-au</td>
</tr>
<tr>
<td>basaa</td>
<td>english-australia</td>
</tr>
<tr>
<td>basque</td>
<td>english-ca</td>
</tr>
<tr>
<td>belarusian</td>
<td>english-canada</td>
</tr>
<tr>
<td>bemba</td>
<td>english-gb</td>
</tr>
<tr>
<td>bena</td>
<td>english-newzealand</td>
</tr>
<tr>
<td>bengali</td>
<td>english-nz</td>
</tr>
<tr>
<td>bodo</td>
<td>english-unitedkingdom</td>
</tr>
<tr>
<td>bosnian-cyrillic</td>
<td>english-unitedstates</td>
</tr>
<tr>
<td>bosnian-cyrl</td>
<td>english-us</td>
</tr>
<tr>
<td>bosnian-latin</td>
<td>english</td>
</tr>
<tr>
<td>bosnian-latn</td>
<td>esperanto</td>
</tr>
<tr>
<td>bosnian</td>
<td>estonian</td>
</tr>
<tr>
<td>brazilian</td>
<td>ewe</td>
</tr>
<tr>
<td>breton</td>
<td>ewondo</td>
</tr>
<tr>
<td>british</td>
<td>faroese</td>
</tr>
<tr>
<td>bulgarian</td>
<td>filipino</td>
</tr>
<tr>
<td>burmese</td>
<td>finnish</td>
</tr>
<tr>
<td>canadian</td>
<td>french-be</td>
</tr>
<tr>
<td>cantonese</td>
<td>french-belgium</td>
</tr>
<tr>
<td>catalan</td>
<td>french-ca</td>
</tr>
<tr>
<td>centralatlastamazight</td>
<td>french-canada</td>
</tr>
<tr>
<td>centralkurdish</td>
<td>french-ch</td>
</tr>
<tr>
<td>chechen</td>
<td>french-lu</td>
</tr>
<tr>
<td>cherokee</td>
<td>french-luxembourg</td>
</tr>
<tr>
<td>chiga</td>
<td>french-switzerland</td>
</tr>
</tbody>
</table>
french  
friulian  
fulah  
galician  
ganda  
georgian  
german-at  
german-austria  
german-ch  
german-switzerland  
german  
greek  
gujarati  
gusii  
hausa-gh  
hausa-ghana  
hausa-ne  
hausa-niger  
hausa  
hawaiian  
hebrew  
hindi  
hungarian  
icelandic  
igbo  
inari  
indonesian  
interlingua  
irish  
italian  
japanese  
jolafonyi  
kabuverdianu  
kabyle  
kako  
kalaallisut  
kalenjin  
kamba  
kannada  
kashmiri  
kazakh  
khmer  
kikuyu  
kinyarwanda  
konkani  
korean  
koyraborosenni  
koyrachiini  
kwasio  
kyrgyz  
lakota  
langi  
lao  
latvian  
lingala  
lithuanian  
lowersorbian  
lower sorbian  
lubakatanga  
luo  
luxembourgish  
luyia  
macedonian  
machame  
makhuwameetto  
makonde  
malagasy  
malay-brunei  
malay-sg  
malay-singapore  
malay  
malayalam  
maltese  
manx  
marathi  
masai  
mazanderani  
meru  
meta  
mexican  
mongolian  
morisyen  
mundang  
nama  
nepali  
northernluri  
northernsami  
northern ndebele  
norwegianbokmal  
norwegiannynorsk  
nswissgerman  
nuer  
nyankole  
nynorsk  
occitan  
oiriya  
orom  
ossetic  
pashto  
persian  
piedmontese  
 polish  
portuguese-br
1.14 Selecting fonts

\newcommand{\babelfont}{\langle\text{language-list}\rangle\{\langle\text{font-family}\rangle\}[\langle\text{font-options}\rangle]\{\langle\text{font-name}\rangle\}}

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

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

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

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

\begin{verbatim}
\documentclass{article}
\usepackage[swedish, bidi=default]{babel}
\babelfont{import}{hebrew}
\babelfont{rm}{FreeSerif}
\begin{document}
Svenska \foreignlanguage{hebrew}{תיירבע} svenska.
\end{document}
\end{verbatim}

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

\begin{verbatim}
\babelfont{rm}{Iwona}
\babelfont{hebrew}{rm}{FreeSerif}
\end{verbatim}

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

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

\footnote{See also the package combofont for a complementary approach.}
Now, \kaifamily and \kaidefault, as well as \textkai are at your disposal.

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

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

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

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

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

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

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

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

### 1.15 Modifying a language

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

\(^{14}\)And even with the correct code some fonts could be rendered incorrectly by fontspec, so double check the results. \texttt{xetex} fares better, but some font are still problematic.
• The old way, still valid for many languages, to redefine a caption is the following:

\addto\captionsenglish{%
  \renewcommand\contentsname{Foo}%
}\n
As of 3.15, there is no need to hide spaces with % (babel removes them), but it is advisable to do it.

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

\renewcommand\spanishchaptername{Foo}\n
• Macros to be run when a language is selected can be add to \extras{\langle lang\rangle}:

\addto\extrasrussian{\mymacro}\n
There is a counterpart for code to be run when a language is unselected: \noextras{\langle lang\rangle}.

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

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

\usepackage[danish]{babel}
\babelprovide[captions=da,hyphenrules=nohyphenation]{danish}\n
first loads danish.ldf, and then redefines the captions for danish (as provided by the ini file) and prevents hyphenation. The rest of the language definitions are not touched.

### 1.16 Creating a language

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

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

```
Package babel Warning: \mylangchaptername not set. Please, define (babel) it in the preamble with something like:
(babel) \renewcommand\mylangchaptername{..}
(babel) Reported on input line 18.
```
In most cases, you will only need to define a few macros.

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

\usepackage{danish}{babel}
\ babelprovide{arinish}
\ renewcommand\arinishchaptername{Chapitula}
\ renewcommand\arinishrefname{Refirenke}
\ renewcommand\arinishhyphenmins{22}

The main language is not changed (danish in this example). So, you must add \selectlanguage{arinish} or other selectors where necessary.

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

\import= \langle language-tag \rangle

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

\babelprovide[import=hu]{hungarian}

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

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

\babelprovide[import]{hungarian}

There are about 200 ini files, with data taken from the ldf files and the CLDR provided by Unicode. Not all languages in the latter are complete, and therefore neither are the ini files. A few languages will show a warning about the current lack of suitability of the date format (hindi, french, breton, and occitan). Besides \today, this option defines an additional command for dates: \langle language\rangle date, which takes three arguments, namely, year, month and day numbers. In fact, \today calls \langle language\rangle today, which in turn calls \langle language\rangle date{\the\year}{\the\month}{\the\day}.

\caption= \langle language-tag \rangle

Loads only the strings. For example:

\babelprovide[captions=hu]{hungarian}

\hyphenrules= \langle language-list \rangle

With this option, with a space-separated list of hyphenation rules, babel assigns to the language the first valid hyphenation rules in the list. For example:
If none of the listed hyphenrules exist, the default behavior applies. Note in this example we set chavacano as first option – without it, it would select spanish even if chavacano exists.
A special value is +, which allocates a new language (in the \TeX sense). It only makes sense as the last value (or the only one; the subsequent ones are silently ignored). It is mostly useful with luatex, because you can add some patterns with \belpatterns, as for example:

\belpatterns[neo]{a1 e1 i1 o1 u1}

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

\begin{itemize}
  \item \texttt{main} This valueless option makes the language the main one. Only in newly defined languages.
  \item \texttt{script=} \langle \texttt{script-name} \rangle
    \begin{itemize}
      \item \texttt{New 3.15} Sets the script name to be used by fontspec (eg, Devanagari). Overrides the value in the ini file. If fontspec does not define it, then babel sets its tag to that provided by the ini file. This value is particularly important because it sets the writing direction, so you must use it if for some reason the default value is wrong.
    \end{itemize}
  \item \texttt{language=} \langle \texttt{language-name} \rangle
    \begin{itemize}
      \item \texttt{New 3.15} Sets the language name to be used by fontspec (eg, Hindi). Overrides the value in the ini file. If fontspec does not define it, then babel sets its tag to that provided by the ini file. Not so important, but sometimes still relevant.
    \end{itemize}
  \item \texttt{mapfont=} direction
    \begin{itemize}
      \item Assigns the font for the writing direction of this language (only with \texttt{bidi=basic}). More precisely, what \texttt{mapfont=direction} means is, ‘when a character has the same direction as the script for the “provided” language, then change its font to that set for this language’. There are 3 directions, following the \texttt{bidi} Unicode algorithm, namely, Arabic-like, Hebrew-like and left to right. So, there should be at most 3 directives of this kind.
    \end{itemize}
  \item \texttt{intraspase=} \langle \texttt{base} \rangle \langle \texttt{shrink} \rangle \langle \texttt{stretch} \rangle
    \begin{itemize}
      \item Sets the interword space for the writing system of the language, in em units (so, 0 .1 0 is \texttt{0em plus .1em}). Like \texttt{\spaceskip}, the em unit applied is that of the current text (more precisely, the previous glyph). Currently used only in Southeast Asian scripts, like Thai. Requires \texttt{import}.
    \end{itemize}
  \item \texttt{intrapenalty=} \langle \texttt{penalty} \rangle
    \begin{itemize}
      \item There will be another value, \texttt{language}, not yet implemented.
      \item In future releases an new value (\texttt{script}) will be added.
    \end{itemize}
\end{itemize}
Sets the interword penalty for the writing system of this language. Currently used only in Southeast Asian scripts, like Thai. Ignored if 0 (which is the default value). Requires \import.

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

### 1.17 Digits

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

For example:

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

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

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

### 1.18 Getting the current language name

\texttt{\languagename} \texttt{\iflanguage{\langle language\rangle}{\langle true\rangle}{\langle false\rangle}}

The control sequence \languagename contains the name of the current language.

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

\begin{verbatim}
\iflanguage{\langle language\rangle}{\langle true\rangle}{\langle false\rangle}
\end{verbatim}

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

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

### 1.19 Hyphenation and line breaking
New 3.9a It is customary to classify hyphens in two types: (1) explicit or hard hyphens, which in \TeX{} are entered as -, and (2) optional or soft hyphens, which are entered as \-.

Strictly, a soft hyphen is not a hyphen, but just a breaking opportunity or, in \TeX{} terms, a “discretionary”; a hard hyphen is a hyphen with a breaking opportunity after it. A further type is a non-breaking hyphen, a hyphen without a breaking opportunity.

In \TeX{}, - and \- forbid further breaking opportunities in the word. This is the desired behavior very often, but not always, and therefore many languages provide shorthands for these cases. Unfortunately, this has not been done consistently: for example, " - in Dutch, Portuguese, Catalan or Danish is a hard hyphen, while in German, Spanish, Norwegian, Slovak or Russian is a soft hyphen. Furthermore, some of them even redefine \-, so that you cannot insert a soft hyphen without breaking opportunities in the rest of the word. Therefore, some macros are 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, Portuguese and Spanish.
- \babelhyphen{nobreak} inserts a hard hyphen without a break after it (even if a space follows).
- \babelhyphen{empty} inserts a break opportunity without a hyphen at all.
- \babelhyphen{⟨text⟩} is a hard “hyphen” using ⟨text⟩ instead. A typical case is \babelhyphen{⟨/⟩}.

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

\babelhyphenation{⟨language⟩,⟨language⟩,...}{⟨exceptions⟩}

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

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

Listed words are saved expanded and therefore it relies on the LICR. Of course, it also works without the LICR if the input and the font encodings are the same, like in Unicode based engines.
NOTE Using \babelhyphenation with Southeast Asian scripts is mostly pointless. But with \babelpatterns (below) you may fine-tune line breaking (only luatex). Even if there are no pattern for the language, you can add at least some typical cases.

\babelpatterns \{\langle\text{language}\rangle, \langle\text{language}\rangle, \ldots\}\{\langle\text{patterns}\rangle\}

**New 3.9m** In luatex only\[^{17}\] 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{\langle\text{lang}\rangle} as well as the language specific encoding (not set in the preamble by default). Multiple \babelpatterns's are allowed.

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

**New 3.31** (Only luatex.) With \babelprovide and imported CJK languages, a simple generic line breaking algorithm (push-out-first) is applied, based on a selection of the Unicode rules.

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

### 1.20 Selecting scripts

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

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 is 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\[^{19}\].

\ensureasci {\langle\text{text}\rangle}

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

If non-ASCII encodings are not loaded (or no encoding at all), it is no-op (also \TeX and \LaTeX are not redefined); otherwise, \ensureasci 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 T51, T3, and T53 are not taken into account, since they are not used

\[^{17}\]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.

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

\[^{19}\]But still defined for backwards compatibility.
for “ordinary” text (they are stored in \BabelNonText, used in some special cases when no Latin encoding is explicitly set).
The foregoing rules (which are applied “at begin document”) cover most of cases. No assumption is made on characters above 127, which may not follow the LICR conventions – the goal is just to ensure most of the ASCII letters and symbols are the right ones.

1.21 Selecting directions

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

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

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

There are some package options controlling bidi writing.

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

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

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

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

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

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

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

Remember basic-r is available in luatex only.

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

---

20 At the time of this writing some Arabic fonts are not rendered correctly by the default luatex font loader, with misplaced kerns inside some words, so double check the resulting text. Have a look at the workaround available on GitHub, under /required/babel/samples
EXAMPLE With bidi=basic both L and R text can be mixed without explicit markup (the latter will be only necessary in some special cases where the Unicode algorithm fails). It is used much like bidi=basic-r, but with R text inside L text you may want to map the font so that the correct features are in force. This is accomplished with an option in \babelprovide, as illustrated:

\documentclass{book}
\usepackage[english, bidi=basic]{babel}
\babelprovide[mapfont=direction]{arabic}
\babelfont{rm}{Crimson}
\babelfont[*arabic]{rm}{FreeSerif}
\begin{document}

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

\end{document}

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

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

\newcommand\refrange[2]{\babelsubr{\textthe{\ref{#1}}-\textthe{\ref{#2}}}}

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

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

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

With counters, \arabic is not only considered \text always (with \babelsublr, see below), but also an “isolated” block which does not interact with the surrounding chars. So, while 1.2 in \text is rendered in that order with bidi=basic (as a decimal number), in \arabic{c1}.\arabic{c2} the visual order is c2.c1. Of course, you may always adjust the order by changing the language, if necessary.[21]

lists required in \tex and \pdftex, but only in bidirectional (with both R and L paragraphs) documents in \latex.

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

contents required in \tex and \pdftex; in \latex toc entries are R by default if the main language is R.

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

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

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

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

extras is used for miscellaneous readjustments which do not fit into the previous groups. Currently redefines in \latex \underline and \LaTeXe New 3.19.

EXAMPLE Typically, in an Arabic document you would need:

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

---

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

\begin{quote}
RTL A ltr text \thechapter{} and still ltr RTL B
\end{quote}

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

\begin{quote}
RTL A \foreignlanguage{english}{ltr text \thechapter{} and still ltr} RTL B
\end{quote}

\texttt{\BabelPatchSection \langle section-name \rangle}

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

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

\begin{quote}
New 3.17 Something like:

\BabelFootnote{\parsfootnote}{\languagename}{()}{()}
\end{quote}

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

\begin{quote}
\footnote{\foreignlanguage{\languagename}{\languagename}{note}}
\end{quote}

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

\begin{quote}
\BabelFootnote{\footnote}{\languagename}{()}{()}%
\BabelFootnote{\localfootnote}{\languagename}{()}{()}%
\BabelFootnote{\mainfootnote}{()}{()}
\end{quote}

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

\textbf{EXAMPLE} If you want to preserve directionality in footnotes and there are many footnotes entirely in English, you can define:
It adds a period outside the English part, so that it is placed at the left in the last line. This means the dot the end of the footnote text should be omitted.

1.22 Language attributes

\languageattribute

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

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

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

1.23 Hooks

\AddBabelHook

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

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

Names containing the string babel are reserved (they are used, for example, by \useshortands* to add a hook for the event afterextras). Current events are the following; in some of them you can use one to three \TeX parameters (#1, #2, #3), with the meaning given:

- adddialect (language name, dialect name) Used by luababel1.def to load the patterns if not preloaded.
- patterns (language name, language with encoding) Executed just after the \language has been set. The second argument has the patterns name actually selected (in the form of either lang:ENC or lang).
- hyphenation (language name, language with encoding) Executed locally just before exceptions given in \babelhyphenation are actually set.
- defaultcommands Used (locally) in \StartBabelCommands.
- encodedcommands (input, font encodings) Used (locally) in \StartBabelCommands. Both xetex and luatex make sure the encoded text is read correctly.
- stopcommands Used to reset the above, if necessary.
- write This event comes just after the switching commands are written to the aux file.
- beforeextras Just before executing \extras\{language\}. This event and the next one should not contain language-dependent code (for that, add it to \extras\{language\}).
- afterextras Just after executing \extras\{language\}. For example, the following deactivates shorthands in all languages:

\AddBabelHook\{noshort\}\{afterextras\}\{\languageshorthands\{none\}\}
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 \caption{language} and \date{language}.

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

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

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

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

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

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

### 1.24 Languages supported by babel with ldf files

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

- **Afrikaans** afrikaans
- **Azerbaijani** azerbaijani
- **Basque** basque
- **Breton** breton
- **Bulgarian** bulgarian
- **Catalan** catalan
- **Croatian** croatian
- **Czech** czech
- **Danish** danish
- **Dutch** dutch
- **English** english, USenglish, american, UKenglish, british, canadian, australian, newzealand
- **Esperanto** esperanto
- **Estonian** estonian
- **Finnish** finnish
- **French** french, francais, canadien, acadian
- **Galician** galician
- **German** austrian, german, germanb, ngerman, naustrian
- **Greek** greek, polutonikogreek
- **Hebrew** hebrew
There are more languages not listed above, including hindi, thai, thai-cjk, latvian, turkmen, magyar, mongolian, romansh, lithuanian, spanglish, vietnamese, japanese, pinyin, arabic, farsi, ibygreek, bgreek, serbianc, frechle, 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.

**NOTE** Please, for info about the support in luatex for some complex scripts, see the wiki, on [https://github.com/latex3/latex2e/wiki/Babel:-Remarks-on-the-luatex-support-for-some-scripts](https://github.com/latex3/latex2e/wiki/Babel:-Remarks-on-the-luatex-support-for-some-scripts).

### 1.25 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 hhline 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 hhline (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\extras{\inputencoding{latin1}}
\addto\extras{\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. 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 \useshortands to activate ' and \defineshortand, or redefine \textquoteright{} (the latter is called by the non-ASCII right quote).

- \bibitem{} is out of sync with \selectlanguage{} in the .aux file. The reason is \bibitem uses \immediate{} (and others, in fact), while \selectlanguage doesn't. There is no known workaround.

- Babel does not take into account \normalsf{} and (non-)French spacing is not always properly (un)set by languages. However, problems are unlikely to happen and therefore this part remains untouched in version 3.9 (but it is in the 'to do' list).

- Using a character mathematically active (ie, with math code "8000) as a shorthand can make \TeX{} enter in an infinite loop in some rare cases. (Another issue in the 'to do' list, although there is a partial solution.)

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

- \csquotes Logical markup for quotes.
- \iflang Tests correctly the current language.
- \hyphsubst Selects a different set of patterns for a language.
- \translator An open platform for packages that need to be localized.
- \siunitx Typesetting of numbers and physical quantities.
- \biblatex Programmable bibliographies and citations.
- \bicaption Bilingual captions.
- \babelbib Multilingual bibliographies.
- \microtype Adjusts the typesetting according to some languages (kerning and spacing).
  - Ligatures can be disabled.
- \substitutefont Combines fonts in several encodings.
- \mkpattern Generates hyphenation patterns.
- \tracklang Tracks which languages have been requested.
- \ucharclasses (xetex) Switches fonts when you switch from one Unicode block to another.
- \zhspacing Spacing for CJK documents in xetex.

\footnote{This explains why \LaTeX{} assumes the lowercase mapping of T1 and does not provide a tool for multiple mappings. Unfortunately, \savinghyphcodes is not a solution either, because lccodes for hyphenation are frozen in the format and cannot be changed.}
1.26 Current and future work

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

Useful additions would be, for example, time, currency, addresses and personal names. But that is the easy part, because they don't require modifying the \TeX{} internals.

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

Also interesting are differences in the sentence structure or related to it. For example, in Basque the number precedes the name (including chapters), in Hungarian “from (1)” is “(1)-ből”, but “from (3)” is “(3)-ből”, in Spanish an item labelled “3.” may be referred to as either “item 3.” or “3.\textsuperscript{er} item”, and so on.

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

1.27 Tentative and experimental code

See the code section for \texttt{\textbackslash foreignlanguage*} (a new starred version of \texttt{\textbackslash foreignlanguage}).

Old stuff

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

- \texttt{\textbackslash babelFSstore\{}\{\texttt{babel-language}\}\}\texttt{\}} sets the current three basic families (rm, sf, tt) as the default for the language given.

- \texttt{\textbackslash babelFSdefault\{}\{\texttt{babel-language}\}\}{\{\texttt{\textbackslash fontscape-features}\}\}} patches \texttt{\textbackslash fontscape} so that the given features are always passed as the optional argument or added to it (not an ideal solution).

So, for example:

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

2 Loading languages with \texttt{language.dat}

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

**New 3.9q** With \texttt{luatex}, however, patterns are loaded on the fly when requested by the language (except the “0th” language, typically english, which is preloaded always). Until

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

\footnote{This feature was added to 3.9o, but it was buggy. Both 3.9o and 3.9p are deprecated.}
3.9n, this task was delegated to the package luatex-hyphen, by Khaled Hosny, Élie Roux, and Manuel Pégourié-Gonnard, and required an extra file named language.dat.lua, but now a new mechanism has been devised based solely on language.dat. **You must rebuild the formats** if upgrading from a previous version. You may want to have a local language.dat for a particular project (for example, a book on Chemistry).\footnote{The loader for lua(e)tex is slightly different as it's not based on babel but on etex.src. Until 3.9p it just didn't work, but thanks to the new code it works by reloading the data in the babel way, i.e., with language.dat.}

### 2.1 Format

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. 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 \TeX{} that the hyphenation patterns just processed have to be known under an alternative name. Here is an example:

```plaintext
% File     : language.dat
% Purpose   : tell in\TeX{} 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.\footnote{This is because different operating systems sometimes use very different file-naming conventions.} For example:

```plaintext
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 \texttt{\textbackslash extras(lang)}).

A typical error when using babel is the following:

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

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

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

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

The following assumptions are made:

- Some of the language-specific definitions might be used by plain \TeX users, so the files
  have to be coded so that they can be read by both \LaTeX and plain \TeX. The current
  format can be checked by looking at the value of the macro \f@name.

- 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\hphantom{mins},
  \langle captions\rangle\langle lang\rangle, \langle date\rangle\langle lang\rangle, \langle extras\rangle\langle lang\rangle
  and \langle noextras\rangle\langle lang\rangle (the last two may
  be left empty); where \langle lang\rangle is either the name of the language definition file or the
  name of the \LaTeX option that is to be used. These macros and their functions are
  discussed below. You must define all or none for a language (or a dialect); defining, say,
  \langle date\rangle\langle lang\rangle but not \langle captions\rangle\langle lang\rangle does not raise an error but can lead to
  unexpected results.

- When a language definition file is loaded, it can define \l@\langle lang\rangle to be a dialect of
  \l@ when \l@\langle lang\rangle is undefined.

- Language names must be all lowercase. If an unknown language is selected, babel will
  attempt setting it after lowercasing its name.

- The semantics of modifiers is not defined (on purpose). In most cases, they will just be
  simple separated options (eg, \texttt{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 \texttt{/}).

Some recommendations:

- The preferred shorthand is \texttt{"}, which is not used in \LaTeX (quotes are entered as \texttt{``}
  and \texttt{''}). Other good choices are characters which are not used in a certain context (eg, \texttt{=} in
  an ancient language). Note however \texttt{=}, \texttt{<}, \texttt{>}, 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 \langle noextras\rangle\langle lang\rangle except for umlauthigh and friends,
  \bbl@deactivate, \bbl@nonfrenchspacing, 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 hardcodded value.
  Never. Instead save its value in \langle extras\rangle\langle lang\rangle.

- Do not switch scripts. If you want to make sure a set of glyphs is used, switch either the
  font encoding (low level) or the language (high level, which in turn may switch the font
  encoding). Usage of things like \langle latintext\rangle is deprecated\textsuperscript{28}

- Please, for “private” internal macros do not use the \bbl@ prefix. It is used by babel and
  it can lead to incompatibilities.

\textsuperscript{28} But not removed, for backward compatibility.
There are no special requirements for documenting your language files. Now they are not included in the base babel manual, so provide a standalone document suited for your needs, as well as other files you think can be useful. A PDF and a “readme” are strongly recommended.

3.1 Guidelines for contributed languages

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

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

- Do not hesitate stating on the file heads you are the author and the maintainer, if you actually are. There is no need to state the babel maintainer(s) as authors if they have not contributed significantly to your language files.

- Fonts are not strictly part of a language, so they are best placed in the corresponding TeX tree. This includes not only tfm, vf, ps1, otf, mf files and the like, but also fd ones.

- Font and input encodings are usually best placed in the corresponding tree, too, but sometimes they belong more naturally to the babel style. Note you may also need to define a LICR.

- Babel ldf files may just interface a framework, as it happens often with Oriental languages/scripts. This framework is best placed in its own directory.

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

3.2 Basic macros

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

\addlanguage

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

\adddialect

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

\langhyphenmins

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

\renewcommand\spanishhyphenmins(34)

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

\providehyphenmins

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

\captions{lang}
The macro \captions{lang} defines the macros that hold the texts to replace the original hard-wired texts.
\date{lang}
The macro \date{lang} defines \today.
\extras{lang}
The macro \extras{lang} contains all the extra definitions needed for a specific language. This macro, like the following, is a hook – you can add things to it, but it must not be used directly.
\noextras{lang}
Because we want to let the user switch between languages, but we do not know what state \TeX{} might be in after the execution of \extras{lang}, a macro that brings \TeX{} into a predefined state is needed. It will be no surprise that the name of this macro is \noextras{lang}.
\bbldeclareattribute
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.
\mainlanguage
To postpone the activation of the definitions needed for a language until the beginning of a document, all language definition files should use \mainlanguage 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 \ProvidesPackage{} command.
\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.
\ldfquit
The macro \ldfquit 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.
\ldffinish
The macro \ldffinish 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, \TeX{} 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 \ldffinish.
\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 \TeX{} to use a font from the second family when a font from the first family in the given encoding seems to be needed.

## 3.3 Skeleton

Here is the basic structure of an .ldf file, with a language, a dialect and an attribute. Strings are best defined using the method explained in in sec. 3.8 (babel 3.9 and later).
\bbl@declare@attribute{<language>}{<attrib>}{%
  \expandafter\addto\expandafter\extras<language>
  \let\captions<language>\captions<attrib><language>}
\providehyphenmins{<language>}{\tw@\thr@@}
\StartBabelCommands*{<language>}{captions}
  \SetString{chaptername}{<chapter name>}
  \% More strings
\StartBabelCommands*{<language>}{date}
  \SetString{monthname}{<name of first month>}
  \% More strings
\StartBabelCommands*{<dialect>}{captions}
  \SetString{chaptername}{<chapter name>}
  \% More strings
\StartBabelCommands*{<dialect>}{date}
  \SetString{monthname}{<name of first month>}
  \% More strings
\EndBabelCommands
\addto{extras<language>}{
  \addto{noextras<language>}
  \let{extras<dialect>}{extras<language>}
  \let{noextras<dialect>}{noextras<language>}
\ldf@finish{<language>}

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

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

3.4 Support for active characters

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

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

The command \bbl@activate is used to change the way an active character expands. \bbl@activate ‘switches on’ the active behavior of the character. \bbl@deactivate lets the active character expand to its former (mostly) non-active self.
\declare@shorthand The macro \declare@shorthand is used to define the various shorthands. It takes three arguments: the name for the collection of shorthands this definition belongs to; the character (sequence) that makes up the shorthand, i.e. ~ or “a; and the code to be executed when the shorthand is encountered. (It does not raise an error if the shorthand character has not been “initiated”.)

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

3.5 Support for saving macro definitions

Language definition files may want to redefine macros that already exist. Therefore a mechanism for saving (and restoring) the original definition of those macros is provided. We provide two macros for this:[29]
\bbl@save To save the current meaning of any control sequence, the macro \bbl@save is provided. It takes one argument, ⟨csname⟩, the control sequence for which the meaning has to be saved.
\bbl@savevariable A second macro is provided to save the current value of a variable. In this context, anything that is allowed after the \the primitive is considered to be a variable. The macro takes one argument, the ⟨variable⟩.

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

3.6 Support for extending macros
\addto The macro \addto{⟨control sequence⟩}{⟨\TeX code⟩} can be used to extend the definition of a macro. The macro need not be defined (i.e., it can be undefined or \relax). This macro can, for instance, be used in adding instructions to a macro like \extrasenglish.

Be careful when using this macro, because depending on the case the assignment could be either global (usually) or local (sometimes). That does not seem very consistent, but this behavior is preserved for backward compatibility. If you are using etoolbox, by Philipp Lehman, consider using the tools provided by this package instead of \addto.

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

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

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

\set@low@box For some languages, quotes need to be lowered to the baseline. For this purpose the macro

---

[29] This mechanism was introduced by Bernd Raichle.
\texttt{\set@low@box} is available. It takes one argument and puts that argument in an \texttt{\hbox}, at the baseline. The result is available in \texttt{\box0} for further processing.

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

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

### 3.8 Encoding-dependent strings

#### New 3.9\textsuperscript{a}

Babel 3.9 provides a way of defining strings in several encodings, intended mainly for \texttt{luatex} and \texttt{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 \texttt{strings}. If there is no \texttt{strings}, these blocks are ignored, except \texttt{\SetCases} (and except if forced as described below). In other words, the old way of defining/switching strings still works and it’s used by default.

It consists of a series of blocks started with \texttt{\StartBabelCommands}. The last block is closed with \texttt{\EndBabelCommands}. Each block is a single group (ie, local declarations apply until the next \texttt{\StartBabelCommands} or \texttt{\EndBabelCommands}). An \texttt{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 \texttt{\addto}. If the language is french, just redefine \texttt{\frenchchaptername}.

\texttt{\StartBabelCommands} \{(\texttt{language-list})\}{(\texttt{category})}\{\texttt{(selector)}\}

The \texttt{\language-list} specifies which languages the block is intended for. A block is taken into account only if the \texttt{\CurrentOption} is listed here. Alternatively, you can define \texttt{\BabelLanguages} to a comma-separated list of languages to be defined (if undefined, \texttt{\StartBabelCommands} sets it to \texttt{\CurrentOption}). You may write \texttt{\CurrentOption} as the language, but this is discouraged – a explicit name (or names) is much better and clearer.

A “\texttt{selector}” is a name to be used as value in package option \texttt{strings}, optionally followed by extra info about the encodings to be used. The name \texttt{\unicode} must be used for \texttt{xetex} and \texttt{luatex} (the \texttt{\key} \texttt{\strings} has also other two special values: \texttt{\generic} and \texttt{\encoded}). If a string is set several times (because several blocks are read), the first one take precedence (ie, it works much like \texttt{\providecommand}).

Encoding info is \texttt{\ charset=} followed by a charset, which if given sets how the strings should be translated to the internal representation used by the engine, typically \texttt{\utf8}, which is the only value supported currently (default is no translations). Note \texttt{\ charset} is applied by \texttt{luatex} and \texttt{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 \texttt{\ 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 \texttt{\strings=encoded}.

Blocks without a selector are read always if the key \texttt{\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 \texttt{LGC} scripts using the \texttt{LICR}). Blocks without a selector can be activated explicitly with \texttt{\strings=\generic} (no block is taken into account except those). With \texttt{\strings=\encoded}, strings in those blocks are set as default (internally, ?). With \texttt{\strings=\encoded} strings are protected, but they are correctly expanded in \texttt{\MakeUppercase} and the like. If there is no key \texttt{\strings}, string definitions are ignored, but \texttt{\SetCases} are still honored (in a encoded way).
The ⟨category⟩ is either captions, date or extras. You must stick to these three categories, even if no error is raised when using other name. It may be empty, too, but in such a case using \SetString is an error (but not \SetCase).

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

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

A real example is:

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

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

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

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

\StartBabelCommands{german,austrian}{date}  
  \SetString\monthiname{Februar}  
  \SetString\monthiiiname{M"{a}rz}  
  \SetString\monthvname{April}  
  \SetString\monthname{Mai}  
  \SetString\monthiiiname{Juni}  
  \SetString\monthvname{Juli}  
  \SetString\monthiiiname{August}  
  \SetString\monthixname{September}  
  \SetString\monthxname{Oktober}  
  \SetString\monthixname{November}  
  \SetString\monthiiiname{Dezember}  
  \SetString\today{\number\day,\number\year\ %  
  \cnsname \monthromannumeral\monthname\endcnsname\space  
  \number\year}
\EndBabelCommands

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

When used in 1df files, previous values of \⟨category⟩⟨language⟩ 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, i.e., if \date⟨language⟩ exists).

\footnote{In future releases further categories may be added.}
\StartBabelCommands\{\langle language-list\rangle\}\{\langle category\rangle\}\{\langle selector\rangle\}

The starred version just forces strings to take a value – if not set as package option, then the default for the engine is used. This is not done by default to prevent backward incompatibilities, but if you are creating a new language this version is better. It’s up to the maintainers of the current languages to decide if using it is appropriate.\footnote{This replaces in 3.9g a short-lived \UseStrings which has been removed because it did not work.}

\EndBabelCommands

Marks the end of the series of blocks.

\AfterBabelCommands\{\langle code\rangle\}

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

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

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

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

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

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

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

#1 is replaced by the roman numeral.

\SetCase\{\langle map-list\rangle\}\{\langle toupper-code\rangle\}\{\langle tolower-code\rangle\}

Sets globally code to be executed at \MakeUppercase and \MakeLowercase. The code would 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 \langle map-list\rangle is a series of macros using the internal format of \@uc\lclist (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 \LaTeX, we could set for Turkish:

\StartBabelCommands{turkish}\{ot1enc, fontenc=OT1\}
\SetCase
  {\uccode'^10='I\relax}
  {\lccode'^10='I\relax}
\StartBabelCommands{turkish}\{unicode, fontenc=TU EU1 EU2, charset=utf8\}
\SetCase
  {\uccode'i='I\relax
   \uccode'i='I\relax}
  {\lccode'i='I\relax
   \lccode'i='I\relax}
\StartBabelCommands{turkish}{}
\SetCase{
  \uccode`i="9D\relax
  \uccode"19="I\relax
  \llcode"9D="I\relax
  \llcode`I="19\relax
}
\EndBabelCommands

(Note the mapping for OT1 is not complete.)

\SetHyphenMap{(to-lower-macros)}

\SetHyphenMap

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

- \texttt{\BabelLower\{\uccode\}\{\llcode\}} is similar to \llcode but it's ignored if the char has been set and saves the original llcode to restore it when switching the language (except with hyphenmap=first).

- \texttt{\BabelLowerMM\{\uccode-from\}\{\uccode-to\}\{\step\}\{\llcode-from\}} loops though the given uppercase codes, using the step, and assigns them the llcode, which is also increased (MM stands for many-to-many).

- \texttt{\BabelLowerMO\{\uccode-from\}\{\uccode-to\}\{\step\}\{\llcode\}} loops though the given uppercase codes, using the step, and assigns them the llcode, which is fixed (MO stands for many-to-one).

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

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

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

4 Changes

4.1 Changes in babel version 3.9

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

- \texttt{\select@language} did not set \texttt{\language@name}. This meant the language in force when auxiliary files were loaded was the one used in, for example, shorthands – if the language was german, a \texttt{\select@language{spanish}} had no effect.

- \texttt{\foreignlanguage and otherlanguage* messed up \texttt{\extras<language>}}. Scripts, encodings and many other things were not switched correctly.
• The :ENC mechanism for hyphenation patterns used the encoding of the previous language, not that of the language being selected.

• ' (with activeacute) had the original value when writing to an auxiliary file, and things like an infinite loop could happen. It worked incorrectly with ^ (if activated) and also if deactivated.

• Active chars where not reset at the end of language options, and that lead to incompatibilities between languages.

• \text{tor} math raised and error with a conditional.

• \alias shorthand didn’t work (or only in a few and very specific cases).

• \lenglish was defined incorrectly (using \let instead of \chardef).

• ldf files not bundled with babel were not recognized when called as global options.

Part II

Source code

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

5 Identification and loading of required files

Code documentation is still under revision.
The babel package after unpacking consists of the following files:

switch.def defines macros to set and switch languages.
babel.def defines the rest of macros. It has tow 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.

6 locale directory

A required component of babel is a set of ini files with basic definitions for about 200 languages. They are distributed as a separate zip file, not packed as dtx. With them, babel will fully support Unicode engines.
Most of them are essentially finished (except bugs and mistakes, of course). Some of them are still incomplete (but they will be usable), and there are some omissions (eg. Latin and
polytonic Greek, and there are no geographic areas in Spanish). Hindi, French, Occitan and Breton will show a warning related to dates. Not all include LICR variants. This is a preliminary documentation.

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

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

**version** of the ini file

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

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

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

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

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

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

### 7 Tools

1 ⟨⟨version=3.31⟩⟩
2 ⟨⟨date=2019/05/04⟩⟩

**Do not use the following macros in ldf files. They may change in the future.** This applies mainly to those recently added for replacing, trimming and looping. The older ones, like \bbl@afterfi, will not change.

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

3 ⟨⟨*Basic macros⟩⟩ ≡
4 \bbl@trace{Basic macros}
5 \def\bbl@strip(slash\{\expandafter}@gobble\string
6 \def\bbl@add{\#1\#2}(%
7 \bbl@ifunset{\bbl@strip(slash\#1)%
8 \{\def\#1{\#2}}%
9 \{\expandafter\def\expandafter\#1\expandafter\{\#1\#2}}%
10 \def\bbl@in\{\expandtwoargs\in\}
11 \def\bbl@csarg#1\#2{\expandafter\#1\csname bbl@#2\endcsname}%
12 \def\bbl@cs#1\#2{\csname bbl@#1\endcsname}%
13 \def\bbl@loop#1\#2\#3{\bbl@loop#1\#3,\bbl@loop#1\#2,\@nni,}
14 \def\bbl@loopx#1\#2{\expandafter\bbl@loop\expandafter#1\expandafter\{\#2}}
15 \def\bbl@loop#1\#2\#3,(%
16 \ifx\@nni\relax\else
17 \def\#1(#3)#2\bbl@afterfi\bbl@loop#1(#2)%
18 \fi}
19 \def\bbl@for#1\#2\#3{\bbl@loopx#1\#2}{\ifx#1@empty\else#3\fi}}

\bbl@add@list This internal macro adds its second argument to a comma separated list in its first argument. When the list is not defined yet (or empty), it will be initiated. It presumes
Because the code that is used in the handling of active characters may need to look ahead, we take extra care to ‘throw’ it over the \else and \fi parts of an \if-statement\[32\]. These macros will break if another \if...\fi statement appears in one of the arguments and it is not enclosed in braces.

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

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

This code is based on code presented in TUGboat vol. 12, no2, June 1991 in “An expansion Power Lemma” by Sonja Maus.
A tool from url, by Donald Arseneau, which tests if a string is empty or space.

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

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

Now, just syntactical sugar, but it makes partial expansion of some code a lot more simple and readable. Here \ stands for \noexpand and \<...> for \noexpand applied to a built macro name (the latter does not define the macro if undefined to \relax, because it is created locally). The result may be followed by extra arguments, if necessary.
Two further tools. `\bbl@samestring` first expand its arguments and then compare their expansion (sanitized, so that the catcodes do not matter). `\bbl@engine` takes the following values: 0 is pdfTeX, 1 is luatex, and 2 is xetex. You may use the latter it in your language style if you want.

```latex
\def\bbl@ifsamestring#1#2{% 
  \begingroup
  \protected@edef\bbl@tempb{#1}%
  \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
  \protected@edef\bbl@tempc{#2}%
  \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
  \ifx\bbl@tempb\bbl@tempc
    \aftergroup@firstoftwo
  \else
    \aftergroup@secondoftwo
  \fi
}\endgroup
\chardef\bbl@engine=%
\ifdef\directlua\@undefined
  \ifdef\XeTeXinputencoding\@undefined
    \tw@
  \else
    \ne
  \fi
\else
\ne
\fi
⟨⟨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.

```latex
\iffalse
\edef\ProvidesFile#1[#2 #3 #4]{% 
  \wlog{File: #1 #4 #3 <#2>}%
\let\ProvidesFile\@undefined
\fi
\fi
⟨⟨/Make sure ProvidesFile is defined⟩⟩

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

```latex
\iffalse
\input plain.def\relax
\fi
⟨⟨/Load macros for plain if not \LaTeX⟩⟩

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

```latex
\iffalse
\input plain.def\relax
\fi
⟨⟨/Load macros for plain if not \LaTeX⟩⟩

7.1 Multiple languages

Plain \LaTeX 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.

\ifx\language@undefined
\csname newcount\endcsname\language
\fi
\endmacro

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

To add languages to \TeX's memory plain \TeX version 3.0 supplies \texttt{\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 \texttt{\newlanguage} was defined to be \texttt{\outert}. For a format based on plain version 2.x, the definition of \texttt{\newlanguage} can not be copied because \texttt{\count 19} is used for other purposes in these formats. Therefore \addlanguage is defined using a definition based on the macros used to define \texttt{\newlanguage} in plain \TeX version 3.0.

For formats based on plain version 3.0 the definition of \texttt{\newlanguage} can be simply copied, removing \texttt{\outert}. Plain \TeX version 3.0 uses \texttt{\count 19} for this purpose.

\iffalse
\begin{verbatim}
\csname addlanguage\endcsname{\language}
\end{verbatim}
\fi

Now we make sure all required files are loaded. When the command \texttt{\AtBeginDocument} doesn't exist we assume that we are dealing with a plain-based format or \texttt{\LaTeX2.09}. In that case the file \texttt{plain.def} is needed (which also defines \texttt{\AtBeginDocument}, and therefore it is not loaded twice). We need the first part when the format is created, and \texttt{\orig@dum} is used as a flag. Otherwise, we need to use the second part, so \texttt{\orig@dum} is not defined (\texttt{plain.def} undefines it).

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

8 The Package File (\LaTeX, babel.sty)

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 \texttt{\usepackage} command and defines all the language options whose name is different from that of the \texttt{.ldf} file (like variant spellings). It also takes care of a number of compatibility issues with other packages and defines a few aditional package options.
Apart from all the language options below we also have a few options that influence the behavior of language definition files. Many of the following options don’t do anything themselves, they are just defined in order to make it possible for babel and language definition files to check if one of them was specified by the user.

8.1 base

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

```latex
\NeedsTeXFormat{LaTeX2e}[2005/12/01]
\ProvidesPackage{babel}{{\langle\langle date\rangle\rangle \langle\langle version\rangle\rangle} The Babel package}
\ifpackagewith{babel}{debug}
\providecommand{\bb@trace}[1]{\message{^^J[ #1 ]}}%  
\let\bb@debug\@firstofone
\providecommand{\bb@trace}[1]{%  
\let\bb@debug\@gobble}
\iffx\bb@switchflag\undefined % Prevent double input
\let\bb@switchflag\undefined % Prevent double input
\input switch.def\relax
\input \input switch.def\relax
\fi
\section{Load patterns in luatex}
\section{Basic macros}
\def\AfterBabelLanguage#1{%
\ifodd\bb@languages\undefined\else
\begingroup
\catcode`\^^I=12
\ifpackagewith{babel}{showlanguages}{%
\begingroup
\def\bb@elt#1#2#3#4{%wlog[#2\^\#1\^\#3\^\#4]}%
\wlog{\langle\star\text{languages}\rangle}%
\bb@languages
\wlog{\langle\text{languages}\rangle}%
\bb@languages
\endgroup}{}
%  
\endgroup
\def\bb@elt#1#2#3#4{%  
\ifnum#2=\z@  
\gedef\bb@nulllanguage{#1}%
\def\bb@elt#1#2#3#4{%  
\fi}%
\fi
\bb@languages
\fi
\def\bb@nulllanguage{}%
\let\bb@activate@preotf\relax % only once
\directlua{  
Babel = Babel or {}
%  
function Babel.pre_otload_v(head)
  if Babel.numbers and Babel.digits_mapped then  
    head = Babel.numbers(head)
```
```
Now the base option. With it we can define (and load, with \texttt{luatex}) hyphenation patterns, even if we are not interested in the rest of babel. Useful for old versions of \texttt{polyglossia}, too.
8.2  key=value options and other general option

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

```latex
\bbl@trace{key=value and another general options}
\bbl@csarg\let{tempa}\expandafter\csname opt@babel.sty\endcsname
\def\bbl@tempb#1.#2{% #1\ifx\empty#2\else,\bbl@afterfi\bbl@tempb#2\fi% 
\def\bbl@tempd#1.#2@\@nil{% 
  \edef\bbl@tempc{%\bbl@tempc\ifi\else \bbl@tempc\fi#1%}
\else 
  \@ifx=\@nil\bbl@afterfi% 
  \edef\bbl@tempc{%\bbl@tempc\ifi\else \bbl@tempc\fi#1.\@empty\@@nil}
\else 
  \edef\bbl@tempc{%\bbl@tempc\ifi\else \bbl@tempc\fi#1% \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
\fi

\let\bbl@tempc@empty
\let\bbl@opt@shorthands\@empty
\bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty\@@nil}
\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.

```latex
\DeclareOption{KeepShorthandsActive}{}
\DeclareOption{activeacute}{}
\DeclareOption{activegrave}{}
\DeclareOption{debug}{}
\DeclareOption{noconfigs}{}
\DeclareOption{showlanguages}{}
\DeclareOption{silent}{}
\DeclareOption{mono}{}
\DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
```

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 \texttt{<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.

```latex
\let\bbl@opt@shorthands@nil
\let\bbl@opt@config@nil
```
The following tool is defined temporarily to store the values of options.

```latex
\let\bbl@opt@main@nil
\let\bbl@opt@headfoot@nil
\let\bbl@opt@layout@nil
```

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.

```latex
\let\bbl@language@opts@empty
\DeclareOption*{%
  \bbl@xin{\string=}{\CurrentOption}%
  \ifin%
    \expandafter\bbl@tempa\CurrentOption\bbl@tempa
  \else
    \bbl@addtolist\bbl@language@opts{\CurrentOption}%
  \fi%
\fi}

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

## 8.3 Conditional loading of shorthands

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

```latex
\bbl@trace{Conditional loading of shorthands}
\def\bbl@sh@string#1{%
  \ifx#1@empty\else
    \ifx#1\string-%
      \else\ifx#1c\string,%
      \else\string#1%
    \fi\fi
  \expandafter\bbl@sh@string%
  \fi%
  \ifx\bbl@opt@shorthands@nil
    \def\bbl@ifshorthand#1#2#3{#2}%
  \else\ifx\bbl@opt@shorthands@empty
    \def\bbl@ifshorthand#1#2#3{#3}%
  \else
    \def\bbl@ifshorthand#1{%
      \bbl@xin{\string#1}{\bbl@opt@shorthands}%
    \ifin%
```
Wemakesureallcharsinthestringare‘other’,withthehelpofanauxiliarymacro definedabove(whichalsozapspaces).

Thefollowingisignoredwithshorthands=off,sinceitisintendedtotake聞きadhitional actionsforcertainchars.

Withheadfoot=langwecansetthelanguageusedinheads/foots. Forexample,in babel/3796justaddsheadfoot=english. Itmisuses@resetactivecharsbutseemstowork.

Fortheoption safewee usea different approach – \bbl@opt@safesayswhichmacrosareredefined(BforbibsandRforrefs). Bydefault,bothareaset.

Forlayoutanauxiliarymacroisprovided,availableforpackagesandlanguagesyles.

8.4 Language options

Languagesareloadedwhenprocessingthe correspondingoption exceptifomainlanguage hasbeenset. In suchcase,itisnotloadeduntilalloptionshasbeenprocessed. The
following macro inputs the ldf file and does some additional checks (\input works, too, but possible errors are not caught).

```latex
\bbltrace{Language options}
\let\bbl@afterlang\relax
\let\Babelfi\relax
\let\bbl@loaded\empty
\def\bbl@load@language#1{%
  \InputIfFileExists{#1.ldf}%
  {\edef\bbl@loaded{\CurrentOption}
    \expandafter{\expandafter\bbl@afterlang\CurrentOption.1df-h}@@k\endcsname
    \expandafter{\expandafter\Babelfi\CurrentOption\endcsname}
  }%
  \bbl@error{%
    Unknown option \CurrentOption'. Either you misspelled it\%
    or the language definition file \CurrentOption.1df was not found}{%
    Valid options are: shorthands=, KeepShorthandsActive,\%
    activeacute, activegrave, noconfigs, safe=, main=, math=\%
    headfoot=, strings=, config=, hyphenmap=, or a language name.}}}
```

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

```latex
\def\bbl@try@load@lang#1#2#3{%
  \InputIfFileExists{CurrentOption.1df}%
  {\bbl@load@language{CurrentOption}%
    #1{\bbl@load@language(#2)#3}%
    \DeclareOption{afrikaans}{\bbl@try@load@lang{dutch}{}}%
    \DeclareOption{brazil}{\bbl@try@load@lang{portuguese}{}}%
    \DeclareOption{brazilian}{\bbl@try@load@lang{portuguese}{}}%
    \DeclareOption{hebrew}{%}
      \input{rilabel.def}%
      \bbl@load@language{hebrew}%
    \DeclareOption{hungarian}{\bbl@try@load@lang{magyar}{}}%
    \DeclareOption{lowersorbian}{\bbl@try@load@lang{lsorbian}{}}%
    \DeclareOption{nymorsk}{\bbl@try@load@lang{norsk}{}}%
    \DeclareOption{polutonikogreek}{%}
      \bbl@try@load@lang{\greek}{\languageattribute{greek}{polutonikogreek}}%
    \DeclareOption{portuguese}{\bbl@try@load@lang{portugues}{}}%
    \DeclareOption{russian}{\bbl@try@load@lang{russianb}{}}%
    \DeclareOption{ukrainian}{\bbl@try@load@lang{ukraineb}{}}%
    \DeclareOption{uppersorbian}{\bbl@try@load@lang{usorbian}{}}%
  }%
  \InputIfFileExists{\bbl@opt@config.cfg}%
  {\typeout{***************^J%
    * Local config file bblopts.cfg used^J%
    *}%}
  \}\%}
\else
  \InputIfFileExists{\bbl@opt@config.cfg}%
  {\typeout{***************^J%
    * Local config file \bbl@opt@config.cfg used^J%
```

Another way to extend the list of ‘known’ options for babel was to create the file bblopts.cfg in which one can add option declarations. However, this mechanism is deprecated – if you want an alternative name for a language, just create a new .ldf file loading the actual one. You can also set the name of the file with the package option config=<name>, which will load <name>.cfg instead.

```latex
\ifx\bbl@opt@config\nil
  \ifpackagewith{babel}{noconfigs}{}%
    \InputIfFileExists{bblopts.cfg}%
    {\typeout{***************^J%
      * Local config file bblopts.cfg used^J%
      *}%}
  \}%
\else
  \InputIfFileExists{\bbl@opt@config.cfg}%
  {\typeout{***************^J%
    * Local config file \bbl@opt@config.cfg used^J%
```
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.

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.

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

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

The options have to be processed in the order in which the user specified them (except, of
course, global options, which \LaTeX{} processes before):
\expandafter\bbl@tempa\bbl@loaded,\@nil
\ifx\bbl@tempb\bbl@tempc\else
\bbl@Warning{%
Last declared language option is '\bbl@tempc',\\%
but the last processed one was '\bbl@tempb'.\\%
The main language cannot be set as both a global\\%
and a package option. Use 'main=\bbl@tempc' as\\%
option. Reported}%
\fi
\else
\DeclareOption{\bbl@opt@main}{\bbl@loadmain}
\ExecuteOptions{\bbl@opt@main}
\DeclareOption*{}
\ProcessOptions*
\fi
\def\AfterBabelLanguage{%
\bbl@error
\{Too late for \string\AfterBabelLanguage%
\{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.
\ifx\bbl@main@language\undefined
\bbl@info{%
You haven't specified a language. I'll use 'nil'\\%
as the main language. Reported}
\bbl@load@language(nil)
\fi
⟨⟨/package⟩⟩
⟨∗core⟩⟩

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

The kernel of the babel system is stored in either hyphen.cfg or switch.def and babel.def. The file babel.def contains most of the code, while switch.def defines the language switching commands; both can be read at run time. The file hyphen.cfg is a file that can be loaded into the format, which is necessary when you want to be able to switch hyphenation patterns (by default, it also inputs switch.def, for "historical reasons", but it is not necessary). When babel.def is loaded it checks if the current version of switch.def is in the format; if not, it is loaded. A further file, babel.sty, contains \LaTeX-specific stuff. Because plain \TeX users might want to use some of the features of the babel system too, care has to be taken that plain \TeX can process the files. For this reason the current format will have to be checked in a number of places. Some of the code below is common to plain \TeX and \LaTeX, some of it is for the \LaTeX case only.
Plain formats based on etex (etex, etext, luatex) don't load hyphen.cfg but etex.src, which follows a different naming convention, so we need to define the babel names. It presumes language.def exists and it is the same file used when formats were created.

9.1 Tools
\ifx\ldf@quit\undefined
\else
\expandafter\endinput
\fi
⟨⟨Make sure ProvidesFile is defined⟩⟩

64
The file `babel.def` expects some definitions made in the \LaTeX\ style file. So, in \LaTeX\ 2.09 and Plain we must provide at least some predefined values as well some tools to set them (even if not all options are available). There is no package options, and therefore an alternative mechanism is provided. For the moment, only \verb|\babeloptionstrings| and \verb|\babeloptionmath| are provided, which can be defined before loading `babel`.

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

```
\ifx\bb@ifshorthand!@undefined
  \let\bb@opt@shorthands!@nil
\def\bb@ifshorthand!#1#2#3{#2}%
\let\bb@language@opts!@empty
\ifx\b babeloptionstrings!@undefined
  \let\bb@opt@strings!@nil
\else
  \let\bb@opt@strings!@b babeloptionstrings
\fi
\def\BabelStringsDefault{generic}
\def\bb@tempa!@normal
\ifx\b babeloptionmath!\bb@tempa
  \def\bb@mathnormal{\noexpand\textnormal}
\fi
\def\AfterBabelLanguage!#1#2{}
\ifx\b babelModifiers!@undefined\let\b babelModifiers!@relax\fi
\let\bb@afterlang!@relax
\def\bb@opt!@safe{BR}
\ifx\ucclist!@undefined\let\ucclist!@empty\fi
\ifx\bb@trace!@undefined\def\bb@trace!@{}\fi
\fi
And continue.
```

```
\ifx\bb@switchflag!@undefined \% Prevent double input
  \let\bb@switchflag!@relax
\input switch.def!@relax
\fi
```

```
\f i
\f i
```

\verb|\chardef\l @english\z @\}
\f i

```
\{(Load patterns in \texttt{luatex})
```
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 \textit{(control sequence)} and \TeX-code to be added to the \textit{(control sequence)}. If the \textit{(control sequence)} has not been defined before it is defined now. The control sequence could also expand to \texttt{\relax}, in which case a circular definition results. The net result is a stack overflow. Otherwise the replacement text for the \textit{(control sequence)} is expanded and stored in a token register, together with the \TeX-code to be added. Finally the \textit{(control sequence)} is redefined, using the contents of the token register.

\begin{verbatim}
\def\addto#1#2{\%
  \ifx#1@undefined
    \def#1{#2}\%
  \else
    \ifx#1\relax
      \def#1{#2}\%
    \else
      \{\toks@\expandafter{#1#2}%
      \xdef#1\{\the\toks@}\%
    \fi
    \fi
  \fi}
\end{verbatim}

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

\begin{verbatim}
\def\bbl@withactive#1#2{\%
  \begingroup
  \lccode`\=\#2\relax
  \lowercase{\endgroup#1~}}
\end{verbatim}

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

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

\begin{verbatim}
\def\bbl@redefine#1{\%
  \edef\bbl@tempa{\bbl@stripbash#1}\%
  \expandafter\let\csname org@\bbl@tempa\endcsname#1\%
  \expandafter\def\csname\bbl@tempa\endcsname}
\end{verbatim}

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

\begin{verbatim}
@onlypreamble\bbl@redefine
\end{verbatim}

\textbf{\bbl@redefine@long} This version of \texttt{\bbl@redefine} can be used to redefine \texttt{\long} commands such as \texttt{\ifthenelse}.

\begin{verbatim}
\def\bbl@redefine@long#1{\%
  \edef\bbl@tempa{\bbl@stripbash#1}\%
  \expandafter\let\csname org@\bbl@tempa\endcsname#1\%
  \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
\end{verbatim}

\begin{verbatim}
@onlypreamble\bbl@redefine@long
\end{verbatim}

\textbf{\bbl@redefinerobust} For commands that are redefined, but which \textit{might} be robust we need a slightly more intelligent macro. A robust command \texttt{\foo} is defined to expand to \texttt{\protect\foo}. So it is necessary to check whether \texttt{\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 \texttt{\foo}.

\begin{verbatim}
\def\bbl@redefinerobust#1{\%
  \edef\bbl@tempa{\bbl@stripbash#1}\%
  \expandafter\let\csname org@\bbl@tempa\endcsname#1\%
  \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
\end{verbatim}

\begin{verbatim}
@onlypreamble\bbl@redefinerobust
\end{verbatim}
This command should only be used in the preamble of the document.

\@onlypreamble\bbl@redefinerobust

9.2 Hooks

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

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).

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

The macro \bbl@e@⟨language⟩ contains \bbl@ensure{⟨include⟩}{⟨exclude⟩}{⟨fontenc⟩},
which in in turn loops over the macros names in \bbl@captionslist, excluding (with the
help of \vin@) 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@trace{Defining babelensure}
\newcommand\babelensure[2][]{% TODO - revise test files
\AddBabelHook{babel-ensure}{afterextras}{%
\Ifcase\bbl@select@type
\nameuse{\bbl@@\languagename}\%
\fi%
\begingroup
\let\bbl@ens@include@empty
\let\bbl@ens@exclude@empty
\def\bbl@ens@fontenc{@relax}%
\def\bbl@temp@#1{%
\ifx\empty@#1\else\noexpandafter\bbl@temp\fi%}
\edef\bbl@temp{%\bbl@temp@\empty@}{%
\def\bbl@temp@#1@{@empty}%
\def\bbl@temp@#1@@{\nameuse{\bbl@@\languagename}{#1}}%
\bbl@temp{\bbl@ens}%
\foreach\bbl@temp@{\bbl@temp@}{%
\expandafter{\bbl@ens@include@}}%
\expandafter{\bbl@ens@include@}%
\expandafter{\bbl@ens@exclude@}%
\toks@expandafter{\bbl@temp}%
\bbl@exp{%
\endgroup
\edef\bbl@ens@#1#2#3{% 1: include 2: exclude 3: fontenc%
\def\bbl@temp@#1{\empty@}{%
\in@{#1}#2}{%
\ifin@else\bbl@ifunset{\bbl@ens@\languagename}%
\bbl@exp{%
\\:\DeclareRobustCommand<\bbl@ens@\languagename>[1]{%
\\:\foreignlanguage{\languagename}%
{\ifx\relax@#1\else\fontencoding@3\selectfont\fi

###########1}}}%
}{%
\toks@expandafter{#1}{%
\edef\bbl@csarg{\noexpand\nameuse{\languagename}}%
\nameuse{\toks@}{%}
\expandafter{\bbl@tempb}{%
\if\bbl@csarg@today@empty\fi%
\expandafter{\bbl@tempb@\bbl@captionslist@today@empty}{%
\def\bbl@temp@#1{\empty@}{%
\ifx\empty@#1\else\noexpandafter\bbl@temp\fi%}
\expandafter{\bbl@tempb@\bbl@captionslist@today@empty}{%
\def\bbl@temp@#1{\empty@}{%
\if\bbl@csarg@\in@{\nameuse{\languagename}@expandafter}@expandafter{#1}{}%}
\ifin@else\bbl@temp@#1@empty\fi%
\if\bbl@temp@#1@empty\fi%
\bbl@temp@#1@empty}%
\bbl@exp{%
\edef\bbl@captionslist{%
\prefacename\refname\abstractname\bibname\chaptername\appendixname
\contentsname\listfigurename\listtablename\indexname\figurename
\tablename\partname\enclname\ccname\headtoname\pagename\seename

68
9.3 Setting up language files

\LdfInit

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

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

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

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

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

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

Finally we check \originalTeX.

\bbl@trace{ Macros for setting language files up}
\def\bbl@ldfinit{%
\let\bbl@screset@empty
\let\BabelStrings\bbl@opt@string
\let\BabelOptions@empty
\let\BabelLanguages@relax
\ifx\originalTeX@undefined
\let\originalTeX@empty
\else
\originalTeX
\fi}
\def\LdfInit#1#2{%\chardef@atcatcode=\catcode`@
\catcode`@=11\relax
\chardef@eqcatcode=\catcode`=
\catcode`\=12\relax
\expandafter\if\expandafter\@backslashchar
\expandafter\if\expandafter\csname#2\endcsname\relax\else
\ldf@quit{#1}%
\fi
\else
\expandafter\if\expandafter\csname#2\endcsname\relax\else
\ldf@quit{#1}%
\fi
\fi
\fi}
\def\ldf@quit#1{%\expandafter\main@language\expandafter{#1}%\catcode`@=\atcatcode \let\atcatcode\relax
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.

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

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.

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

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

The macro \bbl@add@special is used to add a new character (or single character control sequence) to the macro \dospecials (and \sanitize if \LaTeX is used). It is used only at one place, namely when \initiate@active@char is called (which is ignored if the char
has been made active before). Because \@sanitize can be undefined, we put the
definition inside a conditional.
Items are added to the lists without checking its existence or the original catcode. It does
not hurt, but should be fixed. It's already done with \nffs@catcodes, added in 3.10.

\bb@trace{Shorthands}
\def\bb@add@special#1{% 1:a macro like \", \?, etc.
  \bb@add@dospecials\{\do#1\}% test @sanitize = \relax, for back. compat.
  \bb@ifunset{\@sanitize}\{\bb@add@sanitize{\@makeother#1}\}\%}
\ifx\nffs@catcodes\@undefined\else % TODO - same for above
  \begingroup
  \catcode`#1='active
  \nffs@catcodes
  \ifnum\catcode`#1='active
  \endgroup
  \bb@add\nffs@catcodes{\@makeother#1}\%
  \else
  \bb@add\nffs@catcodes{\@makeother#1}\
  \endgroup
  \fi
\fi

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

\begin{verbatim}
\def\bb@remove@special#1{%
  \begingroup
  \def\x#1#2\{\ifnum#1=\#2\noexpand\empty
  \else\noexpand#1\noexpand#2\fi\%
  \def\x\do\{\x\do\%
  \def\x\@makeother\{\x\@makeother\%
  \def\x\{}\endgroup
  \noexpand\dospecials{\dospecials}\%
  \expandafter\ifx\csname \@sanitize\endcsname\relax\else
    \def\noexpand@sanitize{\@sanitize}\%
  \fi%
  \}\x\}
\end{verbatim}

\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\langle char \rangle to expand to the character in its ‘normal state’ and it defines the
active character to expand to \normal@char\langle char \rangle by default (\langle char \rangle being the character
to be made active). Later its definition can be changed to expand to \active@char\langle char \rangle
by calling \bb@activate\langle char \rangle.

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 \bb@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, \langle level\rangle@group, \langle level\rangle@active and \langle next-level\rangle@active
(except in system).
\begin{verbatim}
\def\bb@active@def#1#2#3#4{%
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.

\long\begin{def}{#3}{#1}{% 
\expandafter\ifx\csname#2@sh@#1\endcsname\relax 
\bbi@afterelse \bbi@sh@select\#2\{#3@arg\#1\}#4\% 
\else \bbi@afterfi \csname#2@sh@#1\endcsname \fi} %

\initiate@active@char calls \@initiate@active@char with 3 arguments. All of them are the same character with different catcodes: active, other (\string'ed) and the original one. This trick simplifies the code a lot.

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).

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").

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 \texttt{KeepShorthandsActive}). It is re-activate again at \texttt{\begin{document}}. We also need to make sure that the shorthands are active during the processing of the .\texttt{aux} file. Otherwise some citations may give unexpected results in the printout when a shorthand was used in the optional argument of \texttt{\bibitem} for example. Then we make it active (not strictly necessary, but done for backward compatibility).

\begin{verbatim}
\bb@restoreactive[#2]
\AtBeginDocument{%
\catcode'#2=active
\if@files
  \immediate\write\mainaux{\catcode`string#2=active}%
\fi%
\expandafter\bb@addspecial\csname#2\endcsname
\catcode'#2=active
\fi

Now we have set \texttt{\normalchar}, we must define \texttt{\activechar}, to be executed when the character is activated. We define the first level expansion of \texttt{\activechar} to check the status of the \texttt{@safe@actives} flag. If it is set to true we expand to the 'normal' version of this character, otherwise we call \texttt{\userchar} to start the search of a definition in the user, language and system levels (or eventually \texttt{\normalchar}).

\begin{verbatim}
\let\bb@tempa@firstoftwo
\ifstring'#2%
  \def\bb@tempa{\noexpand\textormath}%
\else
  \ifx\bb@mathnormal\undefined\else
    \let\bb@tempa\bb@mathnormal
  \fi
  \fi
\expandafter\edef\csname active\char#2\endcsname%
\bb@tempa
\{\noexpand\if\safe@actives
  \noexpand\expandafter
  \expandafter\noexpand\csname normal\char#2\endcsname
\else
  \noexpand\expandafter
  \expandafter\noexpand\csname bb@doactive#2\endcsname
\fi%
\expandafter\let\expandafter\bb@doactive
\csname normal\char#2\endcsname}
\bb@csargedef{doactive#2}{%
\expandafter\noexpand\csname user\activechar#2\endcsname}
\end{verbatim}

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

\begin{verbatim}
\active@prefix \char \normalchar
\end{verbatim}

(where \texttt{\activechar} is one control sequence!).

\begin{verbatim}
\bb@csargedef{active@#2}{%
\noexpand\active@prefix\noexpand#1%
\expandafter\noexpand\csname active\char#2\endcsname}%
\bb@csargedef{normal@#2}{%
\noexpand\active@prefix\noexpand#1%
\expandafter\noexpand\csname normal\char#2\endcsname}%
\expandafter\let\expandafter\bb@normal@#2\csname bb@normal@#2\endcsname
\end{verbatim}

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.
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.

Finally, a couple of special cases are taken care of. (1) If we are making the right quote (') active we need to change \pr@ms 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.

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

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.

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 \bb@firstcsc or \bb@scndcs. Hence two more arguments need to follow it.
The command `\active@prefix` which is used in the expansion of active characters has a function similar to `\OT1-cmd` in that it protects the active character whenever `\protect` is not `\typeset@protect`.

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).

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⟩`.

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.

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`.

These macros have two arguments. They use one of their arguments to build a control sequence from.

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.
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.

\textormath

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

\usergroup
\languagegroup
\systemgroup

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

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

\defineshorthand

\languageshorthand 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.

\aliasshorthand First the new shorthand needs to be initialized,
\@notshortand
\sorthandon
The first level definition of these macros just passes the argument on to `\bbl@switch@sh`, adding `\@nil` at the end to denote the end of the list of characters.
\sorthandoff
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.

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

\@babelshorthand\{\active@prefix\babelshorthand\bbl@putsh}\}
\def\bbl@putsh@i@#1\@nil{%
You may want to test if a character is a shorthand. Note it does not test whether the shorthand is on or off.

\newcommand{\fbbelshorthand}{\bbll@ifshorthand\{#2}\{\bbll@switch\sh\#1\}\{\bbll@switch\sh\#2\%}

One of the internal macros that are involved in substituting \texttt{\textbackslash prim} for each right quote in math mode is \texttt{\textbackslash 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.

\def{\bbll@prim@s}{\bbll@ifprimes\{#1\#2\}}

Usually the \texttt{~} is active and expands to \texttt{\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 \texttt{~} 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 \texttt{~} is still a non-break space), and in some cases is inconvenient (if \texttt{~} has been redefined); however, for backward compatibility it is maintained (some existing documents may rely on the babel value).

\begin{verbatim}
\bbll@prime\futurelet@\let@token\bbll@pr@m@s
\begin{group}
\catcode`\^=7 \catcode`*=\active \lccode`\^=``
\lccode`\*=`\``
\lowercase{\gdef{\bbll@pr@m@s}{\bbll@ifprimes\%\bbll@if@primem\pr@@t\egroup}}
\end{group}
\end{verbatim}

\begin{verbatim}
\bbl@activate{-}
\bbl@switch{-}
\end{verbatim}
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.

When the macro \f@encoding is undefined (as it is in plain \TeX) we define it here to expand to OT1
	\\\expandafter\def\csname OT1dp\endcsname{127}
	\\\expandafter\def\csname T1dp\endcsname{4}

When the macro \f@encoding is undefined (as it is in plain \TeX) we define it here to expand to OT1
\\\\ifx\f@encoding@undefined
\\\\def\f@encoding{OT1}
\\\\fi

9.5 Language attributes

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

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.

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

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

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

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

This command should only be used in the preamble of a document.
The error text to be issued when an unknown attribute is selected.

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

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}.

```
def\bbl@declare@tribute\#1\#2\#3{% 
  \bl@xin\{,#2,}{\BabelModifiers,}% 
  \ifin@
    \AfterBabelLanguage\#1{\languageattribute\#1\#2}% 
  \fi
  \bbl@add@list\bbl@attributes\#1-\#2%
  \expandafter\def\csname\#1@\#2\endcsname\#3%
}\bbl@ifattributeset
```

This internal macro has 4 arguments. It can be used to interpret \TeX{} code based on whether a certain attribute was set. This command should appear inside the argument to \AtBeginDocument because the attributes are set in the document preamble, \after{} babel is loaded. The first argument is the language, the second argument the attribute being checked, and the third and fourth arguments are the true and false clauses.

```
def\bbl@ifattributeset\#1\#2\#3\#4{% 
  First we need to find out if any attributes were set; if not we're done.
  \ifx\bbl@known@attrs@undefined \in@false \else \fi
  The we need to check the list of known attributes.
  \bl@xin\{,#1-#2,}{\bbl@known@attrs,}% 
  \ifin@
    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’.
    \fiin@
    \bbl@afterelse#3%
  else
    \bbl@afterfi#4%
  \fi
}\bbl@ifknown@tribute
```

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.

```
def\bbl@ifknown@tribute\#1#2{% 
  We first assume the attribute is unknown.
  \let\bbl@tempa@secondoftwo
  Then we loop over the list of known attributes, trying to find a match.
  \bl@loopx\bbl@tempb[#2]{% 
    \expandafter\in@\expandafter{\expandafter{\bbl@tempb,}{,#1,}% 
  \fiin@
When a match is found the definition of `\bbl@tempa` is changed.

\begin{verbatim}
\let\bbl@tempa\@firstoftwo
\else
\fi%

Finally we execute `\bbl@tempa`.
\end{verbatim}

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

\begin{verbatim}
\def\bbl@clear@tribs{%
  \ifexist\bbl@attributes\undefined\else
  \bbl@loopx\bbl@tempa{\bbl@attributes}{%}
  \expandafter\bbl@clear@tribs{\bbl@tempa}.%
  \let\bbl@attributes\undefined
  \fi%
\}
\end{verbatim}

9.6 Support for saving macro definitions

To save the meaning of control sequences using `\bbl@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.

\begin{verbatim}
\bbl@savecnt
\bbl@beginsave

The initialization of a new save cycle: reset the counter to zero.
\end{verbatim}

\begin{verbatim}
\def\bbl@beginsave{\bbl@savecnt\z@}%

Before it's forgotten, allocate the counter and initialize all.
\end{verbatim}

\begin{verbatim}
\newcount\bbl@savecnt
\bbl@beginsave

\bbl@save

The macro `\bbl@save\langle\texttt{csname}\rangle` saves the current meaning of the control sequence `\langle\texttt{csname}\rangle` to `\originalTeX`\footnote{`\originalTeX` has to be expandable, i.e. you shouldn't let it to `\relax`.}. 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.
\end{verbatim}

\begin{verbatim}
\def\bbl@save#1{%
  \expandafter\let\csname bbl@\number\bbl@savecnt\endcsname\1\relax
  \toks@\expandafter{\originalTeX\let\1=}%
  \bbl@exp%
  \def\\originalTeX\langle\the\toks@\bbl@number\bbl@savecnt\rangle\relax}%
  \advance\bbl@savecnt\@ne

\bbl@savevariable

The macro `\bbl@savevariable\langle\texttt{variable}\rangle` saves the value of the variable. `\langle\texttt{variable}\rangle` can be anything allowed after the `\the` primitive.
\end{verbatim}

\begin{verbatim}
\def\bbl@savevariable#1{%
  \toks@\expandafter{\originalTeX\ #1=}%
  \bbl@exp{\def\\originalTeX{\langle\the\toks@\rangle\relax}}}
\end{verbatim}
Some languages need to have \textbackslash frenchspacing in effect. Others don't want that. The command \textbackslash frenchspacing switches it on when it isn't already in effect and \textbackslash nonfrenchspacing switches it off if necessary.

\begin{verbatim}
def\bbl@frenchspacing(%
  \ifnum\the\sfcode`\textbackslash @m
    \let\bbl@nonfrenchspacing\relax
  \else
    \frenchspacing
    \let\bbl@nonfrenchspacing\textbackslash nonfrenchspacing
  \fi}
\end{verbatim}

9.7 Short tags

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

\begin{verbatim}
def\babeltags#1{%
edef\bbl@tempa{\zap@space#1 \@empty}%
def\bbl@tempb#2@@{%
edef\bbl@tempc{%
  \noexpand\newcommand
  \expandafter\noexpand\csname ###1\textbackslash endcsname\textbackslash %
  \noexpand\protect
  \expandafter\noexpand\csname otherlanguage\textbackslash endcsname\textbackslash %
  \noexpand\newcommand
  \expandafter\noexpand\csname text\#1\textbackslash endcsname\textbackslash %
  \noexpand\foreignlanguage\textbackslash %
  \bbl@tempc}%

\bbl@tempb\bbl@tempa{%
  \expandafter\expandafter\expandafter \bbl@tempa}
}
\end{verbatim}

9.8 Hyphens

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

\begin{verbatim}
def\bbl@frenchspacing()
def\babelhyphenation{
  \newcommand\babelhyphenation[2][\@empty]{%\@empty}
  \ifx\bbl@hyphenation@\relax
    \let\bbl@hyphenation@\empty
  \fi
  \ifx@empty#1\textbackslash warning{%
    You must not intermingle \text{\textbackslash selectlanguage} \text{\textbackslash space} and\%\text{\textbackslash babelhyphenation} \text{\textbackslash space} or some exceptions will not\% be taken into account. Reported}\%
  \fi
  \textbackslash foreach=#1{%
}
\end{verbatim}
This macro makes hyphenation possible. Basically its definition is nothing more than \
\nobreak \hskip Opt plus Opt\].
\def\bbl@allowhyphens{\ifvmode\else\nobreak\hskip z@skip\fi}
\def\bbl@@allowhyphens{\if\@encoding\bbl@t@one\else\bbl@allowhyphens\fi}
\bbl@allhyphen

Macros to insert common hyphens. Note the space before @ in \bbl@allhyphen. Instead of
protecting it with \DeclareRobustCommand, which could insert a \relax, we use the same
procedure as shorthands, with \active@prefix.
\newcommand\bbl@nullhyphen{\char\hyphenchar\font}
\def\bbl@hyphen{\active@prefix\bbl@hyphen\bbl@hyphen}
\def\bbl@hyphen@i[]{\@ifstar{}{\bbl@nullhyphen}\empty}
\ifx\@encoding\bbl@t@one\else\bbl@hyphen@i@\fi}
\def\bbl@nullhyphen@i#1#2@empty{}
\bbl@usehyphen@i@#1#2@empty\endcsname\discretionary{#2}{#2}{#2}
\bbl@nullhyphen@i#1#2@empty\endcsname}
\def\bbl@usehyphen{\leavemode\ifdim\lastskip>0\z@mbox{#1}\else\nobreak#1\fi}
\def\bbl@usehyphen@i{\nobreak\hskip z@skip}
\bbl@usehyphen\bbl@hyphen\bbl@usehyphen% The following two commands are used to wrap the “hyphen” and set the behavior of the
rest of the word – the version with a single @ is used when further hyphenation is allowed,
while that with @@ if no more hyphen are allowed. In both cases, if the hyphen is preceded
by a positive space, breaking after the hyphen is disallowed.
There should not be a discretionaty after a hyphen at the beginning of a word, so it is
prevented if preceded by a skip. Unfortunately, this does handle cases like “-suffix”.
\nobreak is always preceded by \leavemode, in case the shorthand starts a paragraph.
\def\bbl@usehyphen@i@{\leavemode\ifdim\lastskip>0\z@mbox{#1}\else\nobreak#1\fi}
\def\bbl@usehyphen{\nobreak\hskip z@skip\bbl@@usehyphen}
\def\bbl@&@usehyphen@i@{\nobreak\hskip z@skip\bbl@&@usehyphen}
\def\bbl@usehyphen@i@{\nobreak\hskip z@skip\bbl@@usehyphen}
\def\bbl@hyphenchar{\char\hyphenchar\font\@ifnum\hyphenchar=\@ne\bbl@nullhyphen\else\bbl@usehyphen\fi}
Finally, we define the hyphen “types”. Their names will not change, so you may use them
in \texttt{lfd}’s. After a space, the \\%box in \bbl@hy@nobreak is redundant.
\def\bbl@@hyphen@soft{\bbl@usehyphen\discretionary{\bbl@hyphenchar}{}}
\def\bbl@@hyphen@hard{\bbl@usehyphen\discretionary{\bbl@hyphenchar}{}}

\TeX\begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.

84
For some languages the macro `\bbl@disc` is used to ease the insertion of discretionaries for letters that behave ‘abnormally’ at a breakpoint.

\bbl@disc

9.9 Multiencoding strings

The aim following commands is to provide a common interface for strings in several encodings. They also contain 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.

The second one. We need to patch `\@ucliclist`, 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 `\@ucliclist` is not a list any more. Therefore a package option is added to ignore it. Instead of gobbling the macro getting the next two elements (usually `\reserved@a`), we pass it as argument to `\bbl@uclic`. The parser is restarted inside `\langle lang\rangle` because we do not know how many expansions are necessary (depends on whether strings are encoded). The last part is tricky – when uppercasing, we have:

\let\bbl@tolower@empty\bbl@toupper@empty

and starts over (and similarly when lowercasing).

\@ifpackagewith{babel}{nocase}\%
{\let\bbl@patchuclic@relax}\%
{\def\bbl@patchuclic{\global\let\bbl@patchuclic@relax
\@addtomacro\@ucliclist\reserved@b\reserved@b\bbl@uclic}}%
{\def\bbl@uclic@#1{\let\bbl@encoded\bbl@encoded@uclic
\bbl@ifunset{\{\languagename \bbl@uclic\} and resumes it
{#1}\%

85
The following package options control the behavior of \SetString.

**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.
only if the current label or font encoding is the value of \texttt{strings}; otherwise (ie, no \texttt{strings} or a block whose label is not in \texttt{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.

```latex
\newcommand\bbl@startcmds@ii[1][\empty]{% \let\SetString@gobbletow
\let\bbl@stringdef@gobbletow
\let\AfterBabelCommands@gobble
\ifx\empty\@1
\def\bbl@sc@label{generic}% \def\bbl@encstring##1##2{% \ProvideTextCommandDefault##1##2{% \bbl@toglobal##1%
\expandafter\bbl@toglobal\csname string?\string##1\endcsname%
\let\bbl@sc@test\in@true
\else
\let\bbl@sc@charset\space % <- zapped below
\let\bbl@sc@fontenc\space % <- " "
\def\bbl@tempa#1##2@nil{% \bbl@csarg\edef\sc@zap@space\@empty\{##2\}}%
\bbl@foreach[\label=]{\bbl@tempa\bbl@sc@fontenc\empty}%
\def\bbl@tempa#1##2{% space -> comma##1%
\ifx\empty\@2\else\ifx,\bbl@afterfi\bbl@tempa\bbl@tempa\fi\fi}
\edef\bbl@sc@fontenc\{\expandafter\bbl@tempa\bbl@sc@fontenc\empty}%
\edef\bbl@sc@charset\{\expandafter\bbl@tempa\bbl@sc@charset\empty}%
\def\bbl@encstring##1##2{%
\bbl@foreach\bbl@sc@charset{\bbl@tempa\bbl@sc@fontenc{\bbl@ifunset{T##1}{{}\bbl@tempa\bbl@sc@fontenc}{\bbl@tempa\bbl@sc@fontenc}}}%
\edef\bbl@sc@test{% \if\bbl@ 자신의 \@empty \def\SetString##1##2{% \bbl@error{Missing group for string \string##1}% (You must assign strings to some category, typically\% captions or extras, but you set none)\}% \fi
\if\bbl@opt@strings\@nil% % ie, no strings key -> defaults
\else\if\bbl@opt@strings\relax% % ie, strings=encoded
\let\AfterBabelCommands\bbl@aftercmds
\let\SetString\bbl@setstring
\let\bbl@stringdef\bbl@encstring
\else % ie, strings=value
\bbl@sc@test
\fi\fi\fi
\bbl@scswitch
\ifx\bbl@G\empty
\def\SetString##1##2{% \bbl@error{Missing group for string \string##1}% (You must assign strings to some category, typically\% captions or extras, but you set none)\}% \fi
\fi\fi
\fi\fi\fi
\fi}
\bbl@sc@charset
\bbl@scswitch
\ifx\bbl@G\empty
\def\SetString##1##2{% \bbl@error{Missing group for string \string##1}% (You must assign strings to some category, typically\% captions or extras, but you set none)\}% \fi
\fi
```
There are two versions of \bbscs\switch. The first version is used when \ldf\s are read, and it makes sure \(\langle\text{group}\rangle\langle\text{language}\rangle\) is reset, but only once (\bbl@reset 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@for\lang loops \bbl@L but its body is executed only if the value is in \bbl@Languages (inside babel) or \date\langle\text{language}\rangle is defined (after babel has been loaded). There are also two version of \bbl@for\lang. The first one skips the current iteration if the language is not in \bbl@Languages (used in \ldf\s), and the second one skips undefined languages (after babel has been loaded).

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 \providecommand). With the event stringprocess you can preprocess the string by manipulating the value of \bbl@String. If there are several hooks assigned to this event, preprocessing is done in the same order as defined. Finally, the string is set.
\bbl@usehooks{stringprocess}\%
\expandafter\bbl@stringdef
\csname\bbl@LC\expandafter\endsname\expandafter{\bbl@stringdef}}

Now, some additional stuff to be used when encoded strings are used. Captions then include \bbl@encoded for string to be expanded in case transformations. It is \relax by default, but in \MakeUppercase and \MakeLowercase its value is a modified expandable \@changed@cmd.

\def\bbl@opt@strings\relax
\def\bbl@scset#1@2\{\def#1@{\bbl@encoded#2@}}
\bbl@patchuclc
\let\bbl@encoded\relax
\def\bbl@encoded@uclc#1%
\@inmathwarn#1%
\expandafter\ifx\csname cf@encoding\string#1\endsname\relax
\expandafter\ifx\csname ?\string#1\endsname\relax
\TextSymbolUnavailable#1%
\else
\csname ?\string#1\endsname
\fi
\else
\csname cf@encoding\string#1\endsname
\fi}
\def\bbl@scset#1@2\{\def#1@{#2@}}
\fi

Define \SetStringLoop, which is actually set inside \StartBabelCommands. The current definition is somewhat complicated because we need a count, but \count@ is not under our control (remember \SetString may call hooks). Instead of defining a dedicated count, we just “pre-expand” its value.

⟨⟨∗Macros local to BabelCommands⟩⟩ ≡
\newcommand\SetStringLoop[#][#][]{% 
\def\bbl@templ###1{\expandafter\noexpand\csname###1\endsname}% 
\count@\z@ 
\bbl@loop\bbl@tempa{\count@}[#2@]{% empty items and spaces are ok
\advance\count@@\one
\toks@\expandafter{\bbl@tempa}% 
\bbl@exp{% 
\\\SetString\bbl@templ{\roman\count@}{\the\toks@}% 
\count@@=\the\count@\relax}}% 
⟨⟨/Macros local to BabelCommands⟩⟩

Delaying code Now the definition of \AfterBabelCommands when it is activated.
\def\bbl@aftercmds1@%
\toks@\expandafter{\bbl@scafter#1@}% 
\edef\bbl@scafter{\the\toks@}

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

⟨⟨∗Macros local to BabelCommands⟩⟩ ≡
\newcommand\SetCase[3][]{% 
\bbl@patchuclc
\bbl@forlang\bbl@tempa{% 
\expandafter\bbl@encstring
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.

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

The following package options control the behavior of hyphenation mapping.
9.10 Macros common to a number of languages

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

\set@low@box \bbl@trace{Macros related to glyphs}
\def\set@low@box#1{\setbox2@hbox{,}\setbox2@box{#1}%
\dimen2@\ht2@ \advance\dimen2@ -\ht\tw@%
\setbox2@hbox{\lower\dimen2@ \box2@\ht2@\ht\tw@ \dp2@\dp\tw@}

\save@sf@q \save@sf@q is used to save and reset the current space factor.
\def\save@sf@q#1{\leavevmode
\begingroup
\edef@SF{\spacefactor the \spacefactor #1 @SF}
\endgroup}

9.11 Making glyphs available

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

9.11.1 Quotation marks

\quoted@bl@base \texttt{in the T1 encoding the opening double quote at the baseline is available as a separate character, accessible via \texttt{\quoted@bl@base}. In the OT1 encoding it is not available, therefore we make it available by lowering the normal open quote character to the baseline.}
\provide@text@command{\quoted@bl@base}{OT1}{
\save@sf@q{\set@low@box{\textquotedblright} %
\box2@\kern-.04em\bbl@allowhyphens}}

Make sure that when an encoding other than OT1 or T1 is used this glyph can still be typeset.
\provide@text@command@default{\quoted@bl@base}{OT1}{\texttt{\quoted@bl@base}}

\quoted@sin@l@base \texttt{we also need the single quote character at the baseline.}
\provide@text@command{\quoted@sin@l@base}{OT1}{
\save@sf@q{\set@low@box{\textquoteleft} %
\box2@\kern-.04em\bbl@allowhyphens}}

Make sure that when an encoding other than OT1 or T1 is used this glyph can still be typeset.
\provide@text@command@default{\quoted@sin@l@base}{OT1}{\texttt{\quoted@sin@l@base}}

\guillemot@l@eft \guillemot@right \texttt{the guillemet characters are not available in OT1 encoding. They are faked.}
\provide@text@command{\guillemot@l@eft}{OT1}{
\ifmmode
\ll 
\else
\save@sf@q{\nobreak
\raise.2ex\hbox{$\scriptscriptstyle\ll$}\bbl@allowhyphens} %
\fi}
\provide@text@command{\guillemot@right}{OT1}{
\ifmmode
\gg 
\else
\end
Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.

\Save\tex\makeatletter\pdf@strcmp{\the\pdfminorversion}{1} %
\ifpdf
  \makeatother
\fi

\DefineTextSymbol{guillemotleft}{OT1}{guillemotleft}%
\DefineTextSymbol{guillemotright}{OT1}{guillemotright}%

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

\IfMmMode
  \edef\Crtic@\hrule height0.1ex width0.3em
\else
  \Save\tex\makeatletter\pdf@strcmp{\the\pdfminorversion}{1} %
  \ifpdf
    \makeatother
  \fi
\fi

\ProvideTextCommandDefault{guillemotleft}{}{OT1}{%
  \ifmmode
    <%
  \else
    \Save\tex\makeatletter\pdf@strcmp{\the\pdfminorversion}{1} %
    \ifpdf
      \makeatother
    \fi
  \fi

\MakeTextCommandDefault{guillemotleft}{}{OT1}%
\MakeTextCommandDefault{guillemotright}{}{OT1}%

\IfMmMode
  \edef\Crtic@\hrule height0.1ex width0.3em
\else
  \Save\tex\makeatletter\pdf@strcmp{\the\pdfminorversion}{1} %
  \ifpdf
    \makeatother
  \fi
\fi

9.11.2 Letters

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

\IJ \DeclaTextCommand{ij}{OT1}{}{%
  \i\kern-0.02em\bb@allowhyphens j}
\DeclaTextCommand{IJ}{OT1}{}{%
  \I\kern-0.02em\bb@allowhyphens J}

\MakeTextCommandDefault{ij}{}{OT1}{%
\MakeTextCommandDefault{IJ}{}{OT1}{%

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

\ProvideTextCommandDefault{ij}{}{OT1}{%
\ProvideTextCommandDefault{IJ}{}{OT1}{%

\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 Stipevic Mario, (stipevic@olimp.ri.hr).

\def\crtic@{\hrule height0.1ex width0.3em}


\def\crttic@{\hrule height0.1ex width0.33em}
\def\ddj@{%}
\setbox0\hbox{d}\dimen@=\ht0
\advance\dimen@.45\dimen@
\leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@\vbox{\crttic@}}}
\def\DDJ@{%}
\setbox0\hbox{D}\dimen@=.55\ht0
\advance\dimen@\i.15ex % correction for the dash position
\leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@\vbox{\crttic@}}}

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

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

\DeclaeTextCommandDefault{\dj}{%}
\UseTextSymbol{OT1}{\dj}
\DeclaeTextCommandDefault{\DJ}{%}
\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.

\DeclaeTextCommand{\SS}{OT1}{SS}
\DeclaeTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}

9.11.3 Shorthands for quotation marks

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

\glq The ‘german’ single quotes.

\grq The ‘german’ single quotes.

\gllq The ‘german’ double quotes.

\grllq The ‘german’ double quotes.
The definition of \grqq depends on the font encoding. With T1 encoding no extra kerning is needed.

1558 \ProvideTextCommand{\grqq}{T1}{%
1559 \textormath{\textquotedblleft}\mbox{\textquotedblleft}}
1560 \ProvideTextCommand{\grqq}{TU}{%
1561 \textormath{\textquotedblleft}\mbox{\textquotedblleft}}
1562 \ProvideTextCommand{\grqq}{OT1}{%
1563 \save@sf@q{\kern-.07em
1564 \textormath{\textquotedblleft}\mbox{\textquotedblleft}}%
1565 \kern.07em\relax}
1566 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}

\flq The 'french' single guillemets.
\frq

1567 \ProvideTextCommandDefault{\flq}{%  
1568 \textormath{\guillemotleft}\mbox{\guillemotleft}}
1569 \ProvideTextCommandDefault{\frq}{%  
1570 \textormath{\guillemotright}\mbox{\guillemotright}}

\flqq The 'french' double guillemets.
\frqq

1571 \ProvideTextCommandDefault{\flqq}{%  
1572 \textormath{\guillemotleft\guillemotleft}\mbox{\guillemotleft\guillemotleft}}
1573 \ProvideTextCommandDefault{\frqq}{%  
1574 \textormath{\guillemotright\guillemotright}\mbox{\guillemotright\guillemotright}}

9.11.4 Umlauts and tremas

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

\umlauthigh
\umlautlow

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

1575 \def\umlauthigh{%  
1576 \def\bb@umlauta##1{\leavevmode\bgROUP%  
1577 \expandafter\accent\csname f@encoding dpos\endcsname
1578 ##1\bb@allowhyphens\egroup%  
1579 \let\bb@umlauta\bb@umlauta}
1580 \def\umlautlow{%  
1581 \def\bb@umlauta{\protect\lower@umlauta}}
1582 \def\umlautelow{%  
1583 \def\bb@umlauta{\protect\lower@umlauta}}
1584 \umlauthigh

\lower@umlaut

The command \lower@umlaut is used to position the \" closer to the letter.

We want the umlaut character lowered, nearer to the letter. To do this we need an extra \langle dimen \rangle register.

1585 \expandafter\ifx\csname U@D\endcsname\relax
1586 \csname newdimen\endcsname\U@D
1587 \fi

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

Then we compute the new x-height in such a way that the umlaut character is lowered to the base character. The value of \texttt{.45ex} depends on the \texttt{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.

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

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

9.12 Layout

Work in progress.

Layout is mainly intended to set bidi documents, but there is at least a tool useful in general.
9.13 Load engine specific macros

\bbl@trace{Input engine specific macros}
\ifcase\bbl@engine
\input txtbabel.def
\or
\input luababel.def
\or
\input xebabel.def
\fi

9.14 Creating languages

\bbl@trace{Creating languages and reading ini files}
\newcommand\bblprovide[2][]{%
\let\bbl@savelangname\language
\edef\bbl@savelocaleid{\the\localeid}%
% Set name and locale id
\def\language@name{#2}%
\bbl@id@assign
\chardef\localeid\nameuse{bbl@id@\language}
\let\bbl@KVP@captions@nil
\let\bbl@KVP@import@nil
\let\bbl@KVP@main@nil
\let\bbl@KVP@script@nil
\let\bbl@KVP@language@nil
\let\bbl@KVP@dir@nil
\let\bbl@KVP@hyphenrules@nil
\let\bbl@KVP@mapfont@nil
%
\let\bbl@KVP@maparabic@nil
\let\bbl@KVP@mapdigits@nil
\let\bbl@KVP@intraspacing@nil
\bbl@forkv{#1}\{\bbl@csargs\dedef{KVP@#1}{#2}\}% TODO - error handling
\ifx\bbl@KVP@import@nil\else
\bbl@exp{\bbbl@fblank\{\bbl@KVP@import\}}%
\beginpgroup
\def\BabelBeforeIni###1###2{\dedef\bbl@KVP@import{###1}\endinput}%
\inputIfFileExists{babe\-###2.tex}{(){}}%
\endpgroup%
\}{%
\fi
\ifx\bbl@KVP@captions@nil
\let\bbl@KVP@captions\bbl@KVP@import
\fi
\%
\% Load ini
\bb@funset{date###2}%
\{\bbl@provide@new{###2}\%
\{\bbl@fblank{###1}\%
\{\bbl@error%
\{If you want to modify `###2' you must tell how in\%
\the optional argument. See the manual for the\%
\available options.\}%
\{Use this macro as documented\}%
\{\bbl@provide@renew{###2}\%
\%
\% Post tasks
\bb@exp{\bbbl@b@l@ensu@r@e[ex@clu@de=\today]{###2}\%}
\bb@funset{\bbbl@ensure@language@name}\%
\{\bb@e@x@p%}
\{\bbbl@decla@re@robust@com@ma@nd<\bbbl@ensure@language@name>[1]{%}
\{foreign@lan@gu@age{\language@name}%
\{###1\}%%%}
\}%
\%
\% At this point all parameters are defined if `import'. Now we
\% execute some code depending on them. But what about if nothing was
\% imported? We just load the very basic parameters: ids and a few
\% more.
\bb@funset{\bbbl@l@name@###2}%
\{\def\BabelBeforeIni###1###2{%
\beginpgroup
\catcode`\=[12 \catcode`\]=12 \catcode`\=-12 %
\let\bbl@ini@captions@aux@gobbletwo
\dedef\bbl@ini@date ###1.###2.###3.###4\relax ###5###6{%%%}
\bb@read@ini{###1}%
\bb@export@key{chrng}{characters.ranges}{}%
\bb@export@key{dgnat}{numbers.digits.native}{}%
\endgroup% boxed, to avoid extra spaces:
\setboxz@h@box{\inputIfFileExists{babe\-###2.tex}{(){}}}%
\}%
\%
\% -
\%
\% Override script and language names with script= and language=
\ifx\bbl@KVP@script@nil\else
\bbl@csargs\edef{sname=###2}\{\bbl@KVP@script\%
\fi
\ifx\bbl@KVP@language@nil\else
\bbl@csargs\edef{lname=###2}\{\bbl@KVP@language\%
\fi
\%
\% For bidi texts, to switch the language based on direction
\ifx\bbl@KVP@mapfont@nil\else
  \bbl@ifsamestring{"\bbl@KVP@mapfont" direction}\{
    \bbl@error{Option \bbl@KVP@mapfont' unknown for \%
      mapfont. Use 'direction'\%}
  \}
  \bbl@ifunset{\bbl@lsys@language@name}{\bbl@provide@lsys{\language@name}}\{
  \bbl@ifunset{\bbl@wdir@language@name}{\bbl@provide@dirs{\language@name}}\{
    \ifx\bbl@mapselect@undefined
      \ATBeginDocument{%
        \expandafter{\bbl@add\csname selectfont \endcsname{\bbl@mapselect}}%
        \selectfont}%
      \def\bbl@mapselect{%
        \let\bbl@mapselect@relax
        \edef\bbl@prefontid{\fontid\font}%
        \def\bbl@mapdir##1{%
          \def\language@name{##1}\%
          \let\bbl@ifrestoring@firstoftwo % avoid font warning
          \bbl@switchfont
          \directlua{Babel\fontmap
            [\language@name \bbl@wdir@##1\endcsname]%
            [\bbl@prefontid+:\fontid\font]}%}
        \fi
      }%
      \bbl@exp{\\bbl@add\\bbl@mapselect\\bbl@mapdir{\language@name}}%
      \fi
    \fi
  \}
  \else
    \% For Southeast Asian, if interspace in ini -- TODO: as hook
    \ifcase\bbl@engine\or
      \bbl@ifsame{\bbl@intsp@language@name}{}
      \expandafter{\bbl@name{\intsp@language@name@endcsname\empty@else
        \bbl@xin\bbl@cs{sbcp@\language@name}\{Hant,Hans,\japanese,Korean\}}%
      }%
      \bbl@cjk
      \directlua{Babel\fontmap
        [\language@name \bbl@wdir@##1\endcsname]%
        [\bbl@prefontid+:\fontid\font]}%}
    \fi
  \fi
\fi
\fii
\ifx\bbl@KVP@intrapenalty@nil
  \bbl@exp{%
    \bbl@intrapenalty 0 %}
  \fi
\ifx\bbl@KVP@intrapenalty 0
  \bbl@cjk
  \directlua{
    Babel = Babel or {}
    Babel.locale\props = Babel.locale\props or {}
    Babel.locale\props[\locale\id].linebreak = 'c'
  }%
  \ifx\bbl@KVP@intrapenalty@nil
    \bbl@exp{%
      \bbl@intrapenalty 0 %}
    \fi
  \else
    \bbl@seaintrapenalty
    \ifx\bbl@KVP@intrapenalty@nil
      \bbl@exp{%
        \bbl@intrapenalty 0 %}
    \fi
  \fi
\directlua{
  Babel = Babel or {}
  Babel.sea\ranges = Babel.sea\ranges or {}
  Babel.set\chranges{\cs{sbcp@\language@name'},
     \cs{chrng@\language@name'}}
}%
\fi
\ife\bbl@KVP@intraspacenil\else % We may override the ini
\expandafter\bbl@KVP@intraspacenil@@
\fi
\ife\bbl@KVP@intrapenaltynil\else
\expandafter\bbl@KVP@intrapenaltynil@@
\fi\}%
\or
\bbl@xin@\{\bbl@cs{sbcp@\languagesename}\}{Thai,Laoo,Khmr}\
\fin@
\bbl@ifunset{\bbl@intsp@\languagesename}{}%
{\expandafter\iffx\csname bbl@intsp@\languagesename\endcsnameempty\else
\iffx\bbl@KVP@intraspacenil
\bbl@exp{%
\\bbl@intraspacenil@cs{intsp@\languagesename}\\@@}%
\fi
\iffx\bbl@KVP@intrapenaltynil
\bbl@intrapenaltynil@@
\fi
\iffx\bbl@KVP@intrapenaltynil\else % We may override the ini
\expandafter\bbl@KVP@intrapenaltynil@@
\fi
\iffx\bbl@KVP@intrapenaltynil\else
\expandafter\bbl@KVP@intrapenaltynil@@
\fi
\iffx\bbl@ispacesizeundefined
\AtBeginDocument{%
\expandafter\bbl@add
\csname selectfont \endcsname{\bbl@ispacesize}%
\def\bbl@ispacesize{\bbl@cs{xesp@\bbl@cs{sbcp@\languagesename}}}%
\fi}%
\fi
% Native digits, if provided in ini (TeX level, xe and lua)
\ifcase\bbl@engine
\bbl@ifunset{\bbl@dgnat@\languagesename}{}%
{\expandafter\iffx\csname bbl@dgnat@\languagesename\endcsnameempty\else
\expandafter\expandafter\expandafter
\bbl@setdigits\csname bbl@dgnat@\languagesename\endcsname
\iffx\bbl@KVP@maparabicnil\else
\iffx\bbl@latinarabicundefined
\expandafter\let\expandafter\arabic
\csname bbl@counter@\languagesename\endcsname
\else % ie, if layout=counters, which redefines \arabic
\expandafter\let\expandafter\bbl@latinarabic
\csname bbl@counter@\languagesename\endcsname
\fi
\fi
\fi}%
\fi
% Native digits (lua level).
\ifodd\bbl@engine
\bbl@ifunset{\bbl@dgnat@\languagesename}{}%
{\RequirePackage{luatexbase}%
\bbl@activate@preotf
\directlua{
Babel = Babel or {} %% -> presets in luabel
A tool to define the macros for native digits from the list provided in the ini file. Somewhat convoluted because there are 10 digits, but only 9 arguments in \TeX.

\begin{verbatim}
\def\bbl@setdigits#1#2#3#4#5#6#7#8#9#10{
  \def<\languagename digits>####1(%  ie, \langdigits
  \bbl@digits<\languagename>####1\nil)%
  \def<\languagename counter>####1(%  ie, \langcounter
  %\expandon\bbl@counter<\languagename>%
  %\expandon\csname c####1\endcsname)%
  \def<\bbl@counter<\languagename>####1(%  ie, \bbl@counter<\lang
  %\expandon\bbl@digits<\languagename>%
  %\expandon\csname number####1\endcsname)%
  \def\bbl@tempa####1#2#3####5{%
  \bbl@exp(%  Wow, quite a lot of hashes! :-(
  \def<\bbl@digits<\languagename>####1{%
    \ifx####1\nil % ie, \bbl@digits@lang
    \else
      \ifx0#####1\nil
    \end{verbatim}
Depending on whether or not the language exists, we define two macros.

\begin{verbatim}
\def\tempb\provide@new#1{%
  \name\def#1\{\% marks lang exists - required by \StartBabelCommands
  \name\def#1\name\{\% \name\def#1\name\{\% \name\def#1\name\{\% \StartBabelCommands*{#1}(captions)\% \name\def#1\tempb#1\{\nil \% and also if import, implicit \def#1\tempb#1\{\nil \% elt for \tempb\captionss\list \name\def#1\exp\{\% \SetString\#1\% \expafter\tempb\tempb\captionss\list@\empty \else \tempb\read@ini\{\tempb\KVP\captionss\} \% Here all letters cat = 11 \tempb\read@ini \tempb
  \tempb\savestrings \fi} \expafter\tempb\tempb\captionss\list@\empty \else \fi \name\def#1\provide@new\#1{%
  \name\def#1\KVP\captionss\{\nil \% \name\def#1\KVP\captionss\{\nil \% \name\def#1\KVP\captionss\{\nil \% \StartBabelCommands*{#1}(date)\% \name\def#1\KVP\import@\nil \name\def#1\exp\{\% \SetString\today\{\\name\nocaption\today\{#1\today\}\}\expafter\tempb\tempb \else \tempb\savetoday \tempb\savedate \fi \tempb \EndBabelCommands
  \name\def#1\exp\{% \def#$\#1\hyphenmins\{% {\@unexpanded{\name\lftthm@#1}\{2\}{\@unexpanded{\name\lftthm@#1}\}}\% {\@unexpanded{\name\rgthm@#1}\{3\}{\@unexpanded{\name\rgthm@#1}\}}\% \name\provide@hyphens#1\% \name\def#1\KVP\main@\nil\else \expafter\main@\language\expandafter\#1\% \fi}
\end{verbatim}
The hyphen rules option is handled with an auxiliary macro.

```latex
\def\bbl@provide@hyphens#1{% 
\let\bbl@tempa\relax 
\if\bbl@KV@hyphenrules@nil\else 
\bbl@replace\bbl@KV@hyphenrules{}{,}% 
\bbl@foreach\bbl@KV@hyphenrules{% 
\if\bbl@tempa\relax % if not yet found 
\bbl@if@array{\addlanguage<#1>}% 
\}% 
\bbl@ifunset{#1}% 
\}% 
\bbl@exp{\let\bbl@tempa<#1>}% 
\}% 
\bbl@ifunset{#1}% % if no opt or no language in opt found 
\if\bbl@KV@import@nil\else % if importing 
\bbl@exp{% and hyphen rules is not empty 
\bbl@ifblank{@nameuse{\bbl@hyphr@#1}}% 
\}% 
\let\bbl@tempa<#1>@nameuse{\bbl@hyphr@language}% 
\}% 
\bbl@ifunset{#1}% % no hyphen rules found - fallback 
\bbl@exp{\adddialect<#1>@language}% 
\}% 
\let#1<#1>@nameuse{\bbl@tempa}% % found in opt list or ini 
\}% 
\fi
\fi
\fi
\bbl@ifunset{#1}% % ie, relax or undefined 
\bbl@for\bbl@KV@hyphenrules{% 
\bbl@exp{\adddialect<#1>@language}% 
}% 
\let#1<#1>@nameuse{\bbl@tempa}% % found in ini file or undefined 
\}% 
\fi
\if\bbl@KV@import@nil\else % if importing 
\bbl@exp{% and hyphen rules is not empty 
\bbl@ifblank{@nameuse{\bbl@hyphr@#1}}% 
\}% 
\let\bbl@tempa<#1>@nameuse{\bbl@hyphr@language}% 
\}% 
\bbl@ifunset{#1}% % no hyphen rules found - fallback 
\bbl@exp{\adddialect<#1>@language}% 
\}% 
\let#1<#1>@nameuse{\bbl@tempa}% % found in opt list or ini 
\}% 
\fi
\bbl@error% (\#1). Perhaps you misspelled it or your installation is not complete.}% 
\fi
\Fix the name or reinstall babel.}% 
\else
\let\bbl@section\empty
\let\bbl@savestrings\empty
\let\bbl@savetoday\empty
\let\bbl@savetode\empty
\let\bbl@ini@reader\bbl@iniskip
\bbl@info{Importing data from babel-\#1.ini for \language}% 
\loop
\if\\\\figure\fi % relax % Trick, because inside \loop 
\endlinechar\max
The special cases for comment lines and sections are handled by the two following commands. In sections, we provide the possibility to take extra actions at the end or at the start (TODO - but note the last section is not ended). By default, key=val pairs are ignored.

Key-value pairs are treated differently depending on the section in the ini file. The following macros are the readers for identification and typography.
Now captions and captions.licr, depending on the engine. And below also for dates. They rely on a few auxiliary macros. It is expected the ini file provides the complete set in Unicode and LICR, in that order.

```
\ifcase\bbl@engine
  \bbl@csarg\def\ini\kv@captions.licr#1=#2\@\{%
  \bbl@ini@captions@aux{#1}{#2}
\else
  \def\bbl@ini@captions#1=#2\@\{%
  \bbl@ini@captions@aux{#1}{#2}
\fi
```

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

```
\def\bbl@ini@captions@aux#1#2{%\bbl@trim\def\bbl@tempa#1%\bbl@ifblank#2{%\bbl@exp{%\toks@{\bbl@nocaption\bbl@tempa\languagename\bbl@tempa name}}}%\bbl@exp{%\\\bbl@add\bbl@savestrings{%\\SetString\bbl@tempa name>\toks@}}}
```

But dates are more complex. The full date format is stores in date.gregorian, so we must read it in non-Unicode engines, too (saved months are just discarded when the LICR section is reached).

```
\ifcase\bbl@engine
  \bbl@csarg\def\ini\kv@date.gregorian#1=#2\@\{%\bbl@ini@date#1...\relax#2\}%
  \bbl@csarg\def\ini\kv@date.islamic#1=#2\@\{%\bbl@ini@date#1...\relax#2\}islamic\%
  \bbl@csarg\def\ini\kv@date.hebrew#1=#2\@\{%\bbl@ini@date#1...\relax#2\}hebrew\%
  \bbl@csarg\def\ini\kv@date.persian#1=#2\@\{%\bbl@ini@date#1...\relax#2\}persian\%
  \bbl@csarg\def\ini\kv@date.indian#1=#2\@\{%\bbl@ini@date#1...\relax#2\}indian\%
\fi
```

TODO. Remove copypaste pattern.
\ifcase\bbl@engine\let\bbl@savedate@empty\fi
\fi
\edef\bbl@initdate{.1.2.3.4}\relax\#5\#6{% TODO - ignore with 'captions'
\bbl@trim\bbl@initdate{#1.2.3.4}{1.2.3.4}{.1.2.3}}%
\bbl@ifsamething{\bbl@tempa}{months.wide}% to savedate
\bbl@trim\bbl@tempa{#3}%
\bbl@exp{%
\bbl@add\bbl@savedate{%
Set String < months. wide name > { the toks@ } } }%
\bbl@ifsamestring{\bbl@tempa}{months. wide} % to savedate
\bbl@trim\bbl@tempa{#3}%
\bbl@trim\bbl@toks@{#5}%
\bbl@exp{%
\bbl@add\bbl@savedate{%
Set String < the today > { the month } { the day } } } %
\bbl@ifsamestring{\bbl@tempa}{date.long} % defined now
\bbl@trim\bbl@tempa{#5}%
\bbl@TG@date
\global\bbl@csarg@let{date@language name}\bbl@toreplace
\bbl@exp{%
\gdef<language name date>{\protect<language name date>}%
\gdef<language name date>#####1#####2#####3{%
\bbl@usedategrouptrue
\bbl@ensure@language name{%
\bbl@date@language name={#####1}{#####2}{#####3}}%
\bbl@add\bbl@savedate today{%
\bbl@setstring\today{%
<language name date>\{\\the\year}{\\the\month}{\\the\day}}}%
}}%
}
}

Dates will require some macros for the basic formatting. They may be redefined by language, so “semi-public” names (camel case) are used. Oddly enough, the CLDR places particles like “de” inconsistently in either in the date or in the month name.
\let\bbl@calendar@empty
\newcommand\BabelDateSpace{\nobreakspace}
\newcommand\BabelDateDot{.}
\newcommand\BabelDate[]{{\number#1}}
\newcommand\BabelDateDD[]{{\ifnum#1<10 0 \fi\number#1}}
\newcommand\BabelDateM[]{{\number#1}}
\newcommand\BabelDateMMM[]{{\number#1}}
\csname month\romannumeral\bbl@calendar name\endcsname}%
\newcommand\BabelDateYY[]{{\number#1}}%
\newcommand\BabelDateYYYY[]{{\number#1}}%
\ifnum#1<10 0\number#1 %
\else\ifnum#1<100 \expandafter@ gobble\number#1 %
\else\ifnum#1<1000 \expandafter@ gobbletwo\number#1 %
\else\error{Currently two-digit years are restricted to the\}
range 0-9999.}%
\fi}
\newcommand\BabelDateyyy[]{{\number#1}} % FIXME - add leading 0
\def\bbl@replace@finish@iii@number#1{1}%
\bbl@exp{\def\#1#####1#####2#####3{\the\toks@}}%
\edef\bbl@TG@date{%
\bbl@replace\bbl@toreplace[ ]{\BabelDateSpace}{%}
\bbl@replace\bbl@toreplace[ ]{\BabelDateDot}{%}
\bbl@replace\bbl@toreplace[ ]{\BabelDate}{%}
\bbl@replace\bbl@toreplace[ ]{\BabelDateDD}{%}
\bbl@replace\bbl@toreplace[ ]{\BabelDateM}{%}
\bbl@replace\bbl@toreplace[ ]{\BabelDateMM}{%}
\bbl@replace\bbl@toreplace[ ]{\BabelDateYYYY}{%}
Language and Script values to be used when defining a font or setting the direction are set with the following macros.

```latex
\def\bbl@provide@lsys#1{%nn\bbl@ifunset{\bbl@lname@#1}{}\%nn\bbl@csarg\let{\lsys@#1}@empty\%nn\bbl@ifunset{\bbl@name@#1}{\bbl@csarg\gdef{\name@#1}{Default}}{}\%nn\bbl@ifunset{\bbl@soft@#1}{\bbl@csarg\gdef{\soft@#1}{DFLT}}{}\%nn\bbl@csarg\bbl@addtolist{\lsys@#1}{Script=\bbl@cs{\name@#1}}\%nn\bbl@ifunset{\bbl@lname@#1}{}\%nn\{\bbl@csarg\bbl@addtolist{\lsys@#1}{Language=\bbl@cs{\lname@#1}}\%nn\bbl@csarg\bbl@toglobal{\lsys@#1}\}
```

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

```latex
\def\bbl@ini@ids#1{%nn\def\BabelBeforeIni##1##2{%nn\begingroup\%nn\bbl@add\bbl@secpost@identification{\closein1}\%nn\\catcode`\izontal=12 \catcode`\vbar=12 \catcode`\<=12 \%nn\bbl@read@ini{##1}\%nn\endgroup\%nn boxed, to avoid extra spaces:\%nn{\setbox12@\hbox{\InputIfFileExists{\bbl@-#1.tex}\{}\}}}\}
```

## 10 The kernel of Babel (babel.def, only \LaTeX)

### 10.1 The redefinition of the style commands

The rest of the code in this file can only be processed by \LaTeX, so we check the current format. If it is plain \TeX, processing should stop here. But, because of the need to limit the scope of the definition of `format`, a macro that is used locally in the following `if` statement, this comparison is done inside a group. To prevent \TeX from complaining about an unclosed group, the processing of the command `\endinput` is deferred until after the group is closed. This is accomplished by the command `\aftergroup`.

```latex
\def\format{lplain}\%nn\ifx\fmtname\format\%nn\else\%nn\def\format{LaTeXe}\%nn\ifsymmetric\%nn\else\%nn\aftergroup\endinput\%nn\fi\%nn\fi
```
10.2 Cross referencing macros

The \LaTeX\ book states:

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

When the above quote should still be true when a document is typeset in a language  
that has active characters, special care has to be taken of the category codes of these  
characters when they appear in an argument of the cross referencing macros.  
When a cross referencing command processes its argument, all tokens in this argument  
should be character tokens with category 'letter' or 'other'.  
The only way to accomplish this in most cases is to use the trick described in the  
\TeX\ book \cite{2} (Appendix D, page 382). The primitive \texttt{\meaning{A}}  
with \texttt{A} defined as  
\texttt{\def\A{	exttt{B}}} expands to the characters \texttt{macro:\#1->\texttt{B}}  
with all category codes set to  
‘other’ or ‘space’.

\texttt{\newlabel} The macro \texttt{\label} writes a line with a \texttt{\newlabel}  
command into the \texttt{aux} file to define labels.

\begin{verbatim}
\bb@redefine\newlabel#1#2{%  
\@safe@activetrue\org@newlabel{#1}{#2}\@safe@activesfalse
\end{verbatim}

\texttt{\@newlabel} We need to change the definition of the \LaTeX\-internal macro  
\texttt{\newlabel}. 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.

\begin{verbatim}
⟨⟨∗More package options⟩⟩ ≡
\DeclareOption{safe=none}{\let\bb@opt@safe\empty}
\DeclareOption{safe=bib}{\def\bb@opt@safe{B}}
\DeclareOption{safe=ref}{\def\bb@opt@safe{R}}
⟨⟨/More package options⟩⟩
\end{verbatim}

First we open a new group to keep the changed setting of \texttt{\protect}  
local and then we set  
the \texttt{@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.

\begin{verbatim}
\bb@trace{Cross referencing macros}
\ifx\bb@opt@safe\empty\else
\def\@newlabel#1#2#3{%
\@ifunset{#1@#2}
\relax
{\gdef@multiplelabels{%
\latex@warning@no@line{There were multiply-defined labels}}%
\latex@warning@no@line{Label `#2' multiply defined}}%
\global\@namedef{#1@#2}{#3}}
\end{verbatim}

\texttt{\@testdef} An internal \TeX\ macro used to test if the labels that have been written on the \texttt{aux} file  
have changed. It is called by the \texttt{\enddocument} macro. This macro needs to be completely  
rewritten, using \texttt{\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  
\TeX\ keeps reporting that the labels may have changed.

\begin{verbatim}
\CheckCommand*\@testdef[3]{%
\def\reserved@a{#3}%
\expandafter\ifx\csname#1@#2\endcsname\reserved@a
\else
\@tempswatrue
\fi
\end{verbatim}
Now that we made sure that \@testdef still has the same definition we can rewrite it. First we make the shorthands ‘safe’.

\def\@testdef#1#2#3{%  
@safe@activestrue
}

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

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

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

\def\bbl@tempb{%  
@safe@activesfalse
}

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

\ife\bbl@tempa\relax
  \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
\else
  \fi

We do the same for \bbl@tempb.

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

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

\ife\bbl@tempa\bbl@tempb
  \else
    \tempswatrue
  \fi

\ref 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.

\bbl@xinn@{R}\bbl@opt@safe
\ifin@
  \bbl@redefineroobust\ref#1{%  
  @safe@activestrue\org@ref#1}@safe@activesfalse
  \bbl@redefineroobust\pageref#1{%  
    @safe@activestrue\org@pageref#1}@safe@activesfalse
  \else
    \let\org@ref\ref
    \let\org@pageref\pageref
  \fi

\@citet The macro used to cite from a bibliography, \cite, uses an internal macro, \@citet. 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.

\bbl@xinn@{B}\bbl@opt@safe
\ifin@
  \bbl@redefine\@citet#1#2{%  
    @safe@activestrue\edef\tempa#2}@safe@activesfalse
    \org@\@citet#1\{\@tempa\}
\else
\fi

Unfortunately, the packages natbib and cite need a different definition of \@citet. To begin with, natbib has a definition for \@citet with three arguments. We only know that a package is loaded when \begin{document} is executed, so we need to postpone the different redefinition.
Notice that we use `\def` here instead of `\bbl@rerefine` 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.)

```
\def@citex[#1][#2][#3]{
  \@safe@activate\true\edef@tempa[#3]\@safe@activate\false\org@citex[#1][#2]\@tempa}
\def@citex{\@ifpackageloaded{natbib}{\@citex}{\@citex}}
```

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

```
\AtBeginDocument{%
  \@ifpackageloaded{cite}{%
    \def@citex[#1][#2]{
      \@safe@activate\true\org@citex[#1][#2]\@safe@activate\false\@tempa}
    }{}}
```

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

```
\bbbl@rerefine\nocite]{%\@ifpackageloaded{cite}{%
  \def@citex[#1][#2]{
    \@safe@activate\true\org@nocite[#1]\@safe@activate\false}}}
```

\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@activate` is in effect. This switch needs to be reset inside the `\ hbox` which contains the citation label. In order to determine during `.aux` file processing which definition of `\bibcite` is needed we define `\bibcite` in such a way that it redefines itself with the proper definition. We call `\bbbl@cite@choice` to select the proper definition for `\bibcite`. This new definition is then activated.

```
\bbbl@rerefine\bibcite{%
  \bbbl@cite@choice
  \bibcite}
```

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

```
\bbbl@rerefine\bibcite{%
  \def\bbbl@bibcite[#1][#2]{%\org@bibcite[#1]{\@safe@activatefalse#2}}
```

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

```
\bbbl@rerefine\cite@choice{%
  \global\let\bibcite\bbbl@bibcite
  }
```

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

```
\bbbl@rerefine\bibcite{%
  \global\let\bibcite\org@bibcite}
```

Make sure this only happens once.

```
\global\let\bbbl@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.

\AtBeginDocument{\bbl@cite@choice}

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

\bbl@redefine\bibitem#1{%
\@safe@activestruemotion@\bibitem{#1}\@safe@activestruefalse}
\else
\let\org@nocite\nocite
\let\org@citex@citex
\let\org@bibcite@bibcite
\let\org@@bibitem@bibitem
\fi

10.3 Marks

\markright Because the output routine is asynchronous, we must pass the current language attribute to the head lines, together with the text that is put into them. To achieve this we need to adapt the definition of \markright and \markboth somewhat.

We check whether the argument is empty; if it is, we just make sure the scratch token register is empty. Next, we store the argument to \markright in the scratch token register. This way these commands will not be expanded later, and we make sure that the text is typeset using the correct language settings. While doing so, we make sure that active characters that may end up in the mark are not disabled by the output routine kicking in while \@safe@activestruemotion is in effect.

\bbl@trace{Marks}
\IfBabelLayout{sectioning}
{\ifx\bbl@opt@headfoot@nil
\g@addto@macro\resetactivechars{%
\set@typeset@protect
\expandafter\selectlanguage@x\expandafter{\bbl@main@language}%
\let\protect\noexpand
\edef\thepage{%
\noexpand\babelsubl{\unexpanded}\expandafter{\thepage}}}%
\fi}
{\bbl@redefine\markright#1{%
\bbl@ifblank{#1}{%
{\org@markright{}}%
{\toks@{#1}}%
{\bbl@exp%}
{\org@markright{}}%
{\toks@{}}%
{\bbl@exp%
{\makeroom}\\protect{\foreignlanguage{language}{%}
{\makeroom}}}%
\bbl@reset@activestruemotion{\makeroom}{\toks@{}}}}%
\bbl@redefine\markboth#1{%
\bbl@ifblank{#1}{%
{\org@markright{}}%
{\bbl@tempc}{\let@markboth@markboth}
\else
{\bbl@tempc{}}
\fi

\markboth The definition of \markboth is equivalent to that of \markright, except that we need two token registers. The documentclasses report and book define and set the headings for the page. While doing so they also store a copy of \markboth in \@mkboth. Therefore we need to check whether \@mkboth has already been set. If so we need to do that again with the new definition of \markboth.

\ifx\@mkboth\markboth
{\def\bbl@tempc{\let@markboth@markboth}}
\else
{\def\bbl@tempc{}}
\fi
Now we can start the new definition of \markboth

\bbl@redefine\markboth#1#2{% 
\protected@edef\bbl@tempb#1{% 
\protect\foreignlanguage
\{\language\name\}\protect\bbl@restore@actives\#1}% 
\bbl@ifblank\#1{% 
\toks@{}% 
\toks@expandafter\{\bbl@tempb\#1}% 
\bbl@ifblank\#2{% 
\@temptokena{}% 
\@temp@expandafter\{\bbl@tempb\#2}% 
\bbl@exp\{\org@markboth\{\the@toks\}\{\the@temp@tokena\}\}}% 
and copy it to @mkboth if necessary.
\bbl@tempc} % end \IfBabelLayout

10.4 Preventing clashes with other packages

10.4.1 \ifthen

\ifthenelse Sometimes a document writer wants to create a special effect depending on the page a certain fragment of text appears on. This can be achieved by the following piece of code:

\ifthenelse{\isodd{\pageref{some:label}}} 
{code for odd pages} 
{code for even pages}

In order for this to work the argument of \isodd needs to be fully expandable. With the above redefinition of \pageref it is not in the case of this example. To overcome that, we add some code to the definition of \ifthenelse to make things work.

The first thing we need to do is check if the package \ifthen is loaded. This should be done at \begin{document} time.
\bbl@trace{Preventing clashes with other packages}
\bbl@xin@{(R)\bbl@Opt@safe
\bbl@fin@ 
\AtBeginDocument{% 
@@ifpackageloaded@ifthen{% 

Then we can redefine \ifthenelse:
\bbl@redefine@long\ifthenelse#1#2#3{% 

We want to revert the definition of \pageref and \ref to their original definition for the first argument of \ifthenelse, so we first need to store their current meanings.
\let\bbl@temp@pref\pageref 
\let\pageref@org@pageref 
\let\bbl@temp@ref\ref 
\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.
\letsafe@activestrue 
\org@ifthenelse(#1)% 
{\let\pageref\bbl@temp@pref 
\let\ref\bbl@temp@ref
10.4.2 varioref

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

\begin{verbatim}
\AtBeginDocument{%
  \ifpackageloaded{varioref}{%
    \bbi@redefine\@vpageref\#1\#2\#3{%
    \@safe@activestrue
    \org@\vpageref\#1\#2\#3{%
    \@safe@activesfalse}%
  }
}

The same needs to happen for \vrefpagenum.
\begin{verbatim}
\AtBeginDocument{%
  \ifpackageloaded{varioref}{%
    \bbi@redefine\vrefpagenum\#1\#2{%
    \@safe@activestrue
    \org@\vrefpagenum\#1\#2{%
    \@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 exandable 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.

\begin{verbatim}
\expandafter\def\csname Ref \endcsname\#1{%
  \protected@edef\@tempa{\org@ref\#1}\expandafter\MakeUppercase\@tempa}
}\fi
\end{verbatim}

10.4.3 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.

\begin{verbatim}
\AtBeginDocument{%
  \ifpackageloaded{hhline}{%
    \makeatletter
    \def\@currrname{hhline}\input{hhline.sty}\makeatother
\end{verbatim}
10.4.4 hyperref

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 removed for the moment because hyperref is expecting it.

\begin{verbatim}
\AtBeginDocument{% 
  \ifx\pdfstringdefDisableCommands\undefined\else
    \pdfstringdefDisableCommands{\languageshorthands{system}}%
  \fi
}
\end{verbatim}

10.4.5 fancyhdr

\FOREIGNLANGUAGE

The package fancyhdr treats the running head and foot lines somewhat differently as the standard classes. A symptom of this is that the command \texttt{foreignlanguage} which babel adds to the marks can end up inside the argument of \texttt{MakeUppercase}. To prevent unexpected results we need to define \texttt{\FOREIGNLANGUAGE} here.

\begin{verbatim}
\DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{% 
  \lowercase{\immediate\openout15=#1\fd\relax}%
  \immediate\write15{%
    \string\ProvidesFile{\#1.f\d\y\m\d\y\d\y\d\y}
    \space generated font description file^^J
    \string\DeclareFontFamily{#1}{#2}{}^^J
    \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{^^J
    \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{^^J
    \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{^^J
    \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{^^J
    \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{^^J
    \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{^^J
    \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{^^J
    \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{^^J
  }%
  \closeout15
}
\end{verbatim}

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

\begin{verbatim}
\@onlypreamble\substitutefontfamily
\end{verbatim}

10.5 Encoding and fonts

Because documents may use non-ASCII font encodings, we make sure that the logos of \TeX{} and \LaTeX{} always come out in the right encoding. There is a list of non-ASCII encodings. Unfortunately, fontenc deletes its package options, so we must guess which encodings has been loaded by traversing \texttt{\@filelist} to search for \texttt{⟨enc⟩enc\def}. If a non-ASCII has been loaded, we define versions of \texttt{\TeX} and \texttt{\LaTeX} for them using \texttt{vensureasci}. The default ASCII encoding is set, too (in reverse order): the “main” encoding (when the document begins), the last loaded, or OT1.
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. 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.

\begin{verbatim}
\ensureascci
\newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LwN,LMA,LMC,LMS,LMU}
\newcommand\BabelNonText{TS1,T3,TS3}
\let\orgTeX\TeX
\let\orgLaTeX\LaTeX
\let\ensureascci@firstofone
\AtBeginDocument{%
  \in@false
  \bbl@foreach\BabelNonASCII{% is there a text non-ascii enc?
    \in@else
      \lowercase{\bbl@xin{\#1.enc.def},\@filelist,}%
      \fi}%
    \ifin% if a text non-ascii has been loaded
      \def\ensureascci#1{%\fontencoding{#1}\selectfont#1}%
      \DeclareTextCommandDefault{\TeX}{\orgTeX}%
      \DeclareTextCommandDefault{\LaTeX}{\orgLaTeX}%
      \def\bbl@tempb#1@@%{\uppercase{\bbl@tempc#1}ENC.DEF:\empty@@%}
      \def\bbl@tempc#1ENC.DEF/#2@@%{\ifx\empty#2}else
      \edef\ensureascci#1{%\fontencoding{\#1}\selectfont#1}%
      \fi%
      \fi%
    }%
  }%
  \bbl@ifunset{T@#1}%
  {}%}
  \bbl@xin{%,
      \BabelNonASCII,\BabelNonText,}%
\fi%
\bbl@foreach% \bbl@filelist{%\bbl@tempb#1@@% TODO - @@ de mas??
  \bbl@xin{%,\cf@encoding,}%\BabelNonASCII,\BabelNonText,}%
\fi%
\edef\ensureascci{%\fontencoding{\cf@encoding}\noexpand\selectfont#1}%
\fi}
\end{verbatim}

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. 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.

\begin{verbatim}
\AtEndOfPackage{\edef\latinencoding{\cf@encoding}}
\AtBeginDocument{%
  @ifpackageloaded{fontspec}%
  \xdef\latinencoding{%
    \ifx\UTFencode\undefined
      EU\ifcase\bbl@engine\or2\or1\fi
    \else
      \UTFencode
    \fi
  }
\end{verbatim}
Then we can define the command \textlatin which is a declarative switch to a Latin font-encoding. Usage of this macro is deprecated.

\DeclaresRobustCommand{\textlatin}{%}
\fontencoding{\latinencoding}\selectfont
\edef\encodingdefault{\latinencoding}

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.

\ifx@undefined\DeclareTextFontCommand
\DeclaresRobustCommand{\textlatin}{[1]{\leavemode{\textlatin #1}}}
\else
\DeclareTextFontCommand{\textlatin}{\textlatin}
\fi

10.6 Basic bidi support

Work in progress. This code is currently placed here for practical reasons. It is loosely based on rlbabel1.def, but most of it has been developed from scratch. This babel module (by Johannes Braams and Boris Lavva) has served the purpose of typesetting R documents for two decades, and despite its flaws I think it is still a good starting point (some parts have been copied here almost verbatim), partly thanks to its simplicity. I've also looked at ARABI (by Youssef Jabri), which is compatible with babel.

There are two ways of modifying macros to make them “bidi”, namely, by patching the internal low level macros (which is what I have done with lists, columns, counters, tocs, much like rlbabel1 did), and by introducing a “middle layer” just below the user interface (sectioning, footnotes).

- pdftex provides a minimal support for bidi text, and it must be done by hand. Vertical typesetting is not possible.
- xetex is somewhat better, thanks to its font engine (even if not always reliable) and a few additional tools. However, very little is done at the paragraph level. Another challenging problem is text direction does not honour \TeX grouping.
- luatex can provide the most complete solution, as we can manipulate almost freely the node list, the generated lines, and so on, but bidi text does not work out of the box and some development is necessary. It also provides tools to properly set left-to-right and right-to-left page layouts. As Lua\TeX-ja shows, vertical typesetting is possible, too. Its main drawback is font handling is often considered to be less mature than xetex, mainly in Indic scripts (but there are steps to make HarfBuzz, the xetex font engine, available in luatex; see <https://github.com/tatzetwerk/luatex-harbuzz>.

\usepackage{bbl@	ext{trace}{Basic (internal) bidi support}}
\def\bbl@alscripts{,Arabic,Syriac,Thaana,}
\def\bbl@rscripts{,%
Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaean,%
Manichaean,Meroitic Cursive,Meroitic,Old North Arabian,%
Nabataean,N'Ko,Orkhon,Palmyrene,Inscriptional Pahlavi,}%
Sadlly, we have to deal with boxes in math with basic. Activated every math with the package option bidi=:

```
def bbl@mathboxdir(%
    if case bbl@thetexto\\relax
        every hbox\\textdir TLT\\relax%
    else
        every hbox\\textdir TRT\\relax%
    fi}
```

```
def bbl@papdir#1{%
    chardef bbl@the\par\dir #1 %
```

```bash
new count bbl@dirlevel
chardef bbl@the\textdir @z@
chardef bbl@the\par\dir @z@
def bbl@\textdir #1{%
    if case #1 \relax
    chardef bbl@\the\textdir @z@
    bbl@\textdir @i \beginL \endL
    else
    chardef bbl@\the\textdir @me
    bbl@\textdir @i \beginR \endR
    fi}
def bbl@\textdir @i #1 #2{%
    if hmode
        \ifnum\current\group\level > @z@
        \ifnum\current\group\level = bbl@dirlevel
            bbl@error{Multiple bidi settings inside a group}%
            {I'll insert a new group, but expect wrong results.}%
            \begingroup\after\group #2 \after\group\egroup
        \else
            \ifcase\current\group\type or % 0 bottom
                \after\group #2 1 simple {}
                \or
                \after\group #2 \after\group #2 \after\group \egroup % 2 hbox
                \or
                \after\group #2 \after\group \egroup % 3 adj hbox
                \or\or\or\or\or\or\or\or\or % vbox vtop align
                \or
                \after\group #2 \after\group \egroup % 7 noalign
                \or\or\or\or\or\or\or\or\or % output math disc insert vcent mathchoice
                \or
                \after\group #2 14 \begingroup
                \else
                \after\group #2 \after\group \egroup % 15 adj
                \fi
            \fi
            \bbl@dirlevel \current\group\level
        \fi
    #1%
    \fi}
def bbl@pardir#1{\chardef bbl@the\par\dir #1 \relax}
let bbl@bodydir @ gobble
let bbl@pagedir @ gobble
def bbl@dirparastext{\chardef bbl@the\par\dir bbl@the\textdir @thetextdir}
The following command is executed only if there is a right-to-left script (once). It activates the \everypar hack for xetex, to properly handle the par direction. Note text and par dirs are decoupled to some extent (although not completely).

\def\bb@xebidipar{% 
\let\bb@xebidipar\relax 
\TeXXeTstate@one 
\def\bb@xeverypar{% 
 \ifcase\bb@thepar\dir 
 \ifcase\bb@the\textdir\else\beginR\fi 
 \else 
 {\setbox2@\lastbox\beginR\box2@}% 
 \fi}% 
\let\bb@severypar\everypar 
\newtoks\everypar 
\everypar=\bb@severypar 
\bb@severypar={\bb@xeverypar\the\everypar}% 
@ifpackagewith{babel}{\bidi=bidi}% 
 {\let\bb@textdir@i@empty 
 \let\bb@sebidipar@empty 
 \AddBabelHook{\bidi}{\foreign}{% 
 \def\bb@tempa{\def\BabelText####1}% 
 \ifcase\bb@the\textdir 
 \expandafter\bb@tempa\expandafter{\BabelText{\LR{####1}}}% 
 \else 
 \expandafter\bb@tempa\expandafter{\BabelText{\RL{####1}}}% 
 \fi}% 
 \def\bb@par\dir{\{\ifcase\relax\setLR\else\setRL\fi}% 
 \fi}% 
\bb@trace{Local Language Configuration} 
\if\loadlocalcfg\undefined 
 @ifpackagewith{babel}{\noconfigs}% 
 {\let\loadlocalcfg@goobble}% 
 {\def\loadlocalcfg#{% 
 \InputIfFileExists{#1.cfg}% 
 \{/typeout{***********************************************\^J% 
 \* Local config file #1.cfg used\^J% 
 \}}% 
 \@empty}}}

\section*{10.7 Local Language Configuration}

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.

\if\loadlocalcfg\undefined 
 \@ifpackagewith{babel}{\noconfigs}% 
 {\let\loadlocalcfg@goobble}% 
 {\def\loadlocalcfg#{% 
 \InputIfFileExists{#1.cfg}% 
 \{/typeout{***********************************************\^J% 
 \* Local config file #1.cfg used\^J% 
 \}}% 
 \@empty}}

118
Just to be compatible with \texttt{\LaTeX} 2.09 we add a few more lines of code:

\begin{verbatim}
if x @unexpandable@protect@undefined
def@unexpandable@protect{\noexpand\protect\noexpand}
long\def\protected@write#1#2#3{%
begingroup
\let\thepage\relax
#2%
\let\protect@unexpandable@protect
\edef\reserved@a{\write#1{#3}}%
\reserved@a
\endgroup
\if@nobreak\if\vmode\nobreak\fi\fi
\end{verbatim}

\langle\langle Make sure ProvidesFile is defined\rangle\rangle

\texttt{ProvidesFile\{switch.def\}}\langle\langle date\rangle\rangle \langle\langle version\rangle\rangle Babel switching mechanism

\texttt{Load macros for plain if not \LaTeX\rangle\rangle \langle\langle Define core switching macros\rangle\rangle

\texttt{adddialect} The macro \texttt{adddialect} can be used to add the name of a dialect or variant language, for which an already defined hyphenation table can be used.

\begin{verbatim}
def\bbl@version{\langle\langle version\rangle\rangle}
def\bbl@date{\langle\langle date\rangle\rangle}
def\adddialect#1#2{%
global\chardef#1#2\relax
\bbl@usehooks{adddialect}{{#1}{#2}}%
\wlog{\string#1 = a dialect from \string\language#2}

\bbl@iflanguage executes code only if the language \texttt{l@} exists. Otherwise raises and error. The argument of \texttt{\bbl@ifname} 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 \texttt{foreign}language and the like appear in a \texttt{MakeXXX} case. However, a lowercase form is not imposed to improve backward compatibility (perhaps you defined a language named MYLANG, but unfortunately mixed case names cannot be trapped). Note \texttt{l@} is encapsulated, so that its case does not change.

\begin{verbatim}
def\bbl@ifname#1{%
begingroup
    \def\bbl@tempe{\l@}%
    \edef\bbl@tempe{\noexpand\ifundefined{\noexpand\bbl@tempe#1}{}{}}%
    \bbl@tempe{\lowercase\expandafter{\bbl@tempe}}%
    {\uppercase\expandafter{\bbl@tempe}}%
    \empty%
    {\edef\bbl@tempe{\noexpand\\if\noexpand#1{}{}}%
    \uppercase\expandafter{\bbl@tempe}}%
    {\edef\bbl@tempe{\noexpand\\if\noexpand#1{}{}}%
    \lowercase\expandafter{\bbl@tempe}}%
\endgroup
\end{verbatim}

11 Multiple languages (switch.def)

Plain \texttt{\LaTeX} version 3.0 provides the primitive \texttt{\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.
\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.

\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.

\ifx\documentclass@undefined

As \LaTeX 2.09 writes to files expanded whereas \LaTeXe 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.

11.1 Selecting the language

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.

As \LaTeX 2.09 writes to files expanded whereas \LaTeXe 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.
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 \texttt{after group} mechanism to help us. The command \texttt{after group} 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.

\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 language names, separated with a ‘+’ sign; the push function can be simple:

\bbl@push@language{}  \edef\bbl@language@stack{\languagename+\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 \languagename. For this we first define a helper function.

\bbl@pop@language  This macro stores its first element (which is delimited by the ‘+’ sign) in the \languagename and stores the rest of the string (delimited by ‘-’) in its third argument.

\bbl@pop@language{}  \edef\languagename{\string\bbl@language@stack\string-}\edef#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 \texttt{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.

\let\bbl@ifrestoring@secondoftwo  Once the name of the previous language is retrieved from the stack, it is fed to \bbl@set@language to do the actual work of switching everything that needs switching.

An alternative way to identify languages (in the babel sense) with a numerical value is introduced in 3.30. This is one of the first steps for a new interface based on the concept of locale, which explains the name of \localename. This means \@... will be reserved for hyphenation patterns.

\chardef\localename\z@  An alternative way to identify languages (in the babel sense) with a numerical value is introduced in 3.30. This is one of the first steps for a new interface based on the concept of locale, which explains the name of \localename. This means \@... will be reserved for hyphenation patterns.

\chardef\localename\z@  % No real need for a new counter
\edef\bbl@id@last{0}
\edef\bbl@id@assign{%
\bbl@ifunset{\bbl@id@\languagename}%
\{count@bbl@id@last\} \relax
\advance\count@@@ne
\bbl@csarg\chardef{id@@\language}\count@
\edef\bbl@id@last{\the\count@}%
\ifcase\bbl@iengine\or
  \directlua{
    Babel = Babel or {}
    Babel.locale.props = Babel.locale.props or {}
    Babel.locale.props\{\bbl@id@last\} = {}
  }%
  \if\fi}
{ }

The unprotected part of \selectlanguage.

\expandafter\def\csname selectlanguage \endcsname#1{%
  \ifnum\bbl@hymapsel = \@cclv\let\bbl@hymapsel\tw@fi
  \bbl@push@language
  \aftergroup\bbl@pop@language
  \bbl@set@language{#1}}

The macro \bbl@set@language takes care of switching the language environment and of writing entries on the auxiliary files. For historical reasons, language names can be either language of \language. To catch either form a trick is used, but unfortunately as a side effect the catcodes of letters in \language name are messed up. This is a bug, but preserved for backwards compatibility. The list of auxiliary files can be extended by redefining \BabelContentsFiles, but make sure they are loaded inside a group (as aux, toc, lof, and lot do) or the last language of the document will remain active afterwards. We also write a command to change the current language in the auxiliary files.

\def\BabelContentsFiles{toc,lof,lot}
\def\bbl@set@language#1{% from selectlanguage, pop@
  \edef\language name{%
    \ifnum\escapechar=\expandafter`\string\#1@empty
      \else\string\#1@empty\fi}%
    \selectlanguage{\language name}%
    \% write to auxs
    \expandafter\if\csname date@language\endcsname\relax\else
      \if\files\w@\protected@write\@auxout{\string\bbl@aux{\language name}{}}%
      \bbl@usehooks\write{}%
      \fi
    \fi
  }
  \def\selectlanguage#1{% from set@, babel@aux
    \% set hmap
    \ifnum\bbl@hymapsel = \@cclv\chardef\bbl@hymapsel4\relax\fi
    \% set name
    \edef\language name{#1}%
    \bbl@fixname\language name
    \bbl@iflanguage\language name%
    \expandafter\if\csname date@language\endcsname\relax
      \bbl@error
      \{Unknown language '#1'. Either you have\%
        misspelled its name, it has not been installed,\%
        or you requested it in a previous run. Fix its name,\%
        install it or just rerun the file, respectively. In\%
        some cases, you may need to remove the aux file\%
        \{You may proceed, but expect wrong results\}%
    \else
      \% set type

A bit of optimization. Select in heads/foots the language only if necessary. The real thing is in babel.def.

First, check if the user asks for a known language. If so, update the value of \language and call \originalTeX to bring T\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\langle lang\rangle command at definition time by expanding the \csname primitive.

Now activate the language-specific definitions. This is done by constructing the names of three macros by concatenating three words with the argument of \selectlanguage, and calling these macros.

The switching of the values of \lefthyphenmin and \righthyphenmin is somewhat different. First we save their current values, then we check if \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.

\let\bl@select@type\z
\expandafter\bl@switch\expandafter{\language}%
\f i}
\def\babel@aux#12{%
\expandafter\ifx\csname date#1\endcsname\relax
\expandafter\ifx\csname bb@auxwarn#1\endcsname\relax
\ndef\babel@aux#1{ %
def \expandafter \bb@expandafter \def \bb@expandafter { %
\let \csname #1 \endcsname
\let \originalTeX @empty
\bl@beginsave}%
\bb@usehooks{afterreset}{}%
\languageshorthands{none}%
\% set the locale id
\bb@id@assign
\chardef\localeid@nameuse{bb@id@\language}%
\% switch captions, date
\ifcase\bb@select@type
\% fhmode
\hskip\z@skip % trick to ignore spaces
\csname captions#1\endcsname\relax

\let\select@language@x\select@language
\def\babel@switch#1{%
\from \select@, \foreign
\select@language{#1}%
\foreach\BabelContentsFiles{
\writefile{#1}{\bb@toc{#1}{#2}}}%
% TODO - ok in plain?
\f i}
\def\bb@toc#12{%
\select@language{#1}]

The `otherlanguage` environment can be used as an alternative to using the \selectlanguage declarative command. When you are typesetting a document which mixes left-to-right and right-to-left typesetting you have to use this environment in order to let things work as you expect them to. The `\ignorespaces` command is necessary to hide the environment when it is entered in horizontal mode.
The `\endotherlanguage` part of the environment tries to hide itself when it is called in horizontal mode.

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 `\foreignlanguage`.

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".

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`.

`\beforeforeign` is a trick to fix a bug in bidi texts. `\foreignlanguage` is supposed to be a `text` command, and therefore it must emit a `\leavemode`, but it does not, and therefore the indent is placed on the opposite margin. For backward compatibility, however, it is done only if a right-to-left script is requested; otherwise, it is no-op.

`\foreignlanguage*` is a temporary, experimental macro for a few lines with a different script direction, while preserving the paragraph format (thank the braces around `\par`, things like `\hangindent` are not reset). Do not use it in production, because its semantics and its syntax may change (and very likely will, or even it could be removed altogether). Currently it enters in vmode and then selects the language (which in turn sets the paragraph direction).

(3.11) Also experimental are the hook `\foreign` and `\foreign*`. With them you can redefine `BabelText` which by default does nothing. Its behavior is not well defined yet. So, use it in horizontal mode only if you do not want surprises.

In other words, at the beginning of a paragraph `\foreignlanguage` enters into hmode with the surrounding lang, and with `\foreignlanguage*` with the new lang.
This macro does the work for \foreignlanguage and the other language* environment. First we need to store the name of the language and check that it is a known language. Then it just calls bbl@switch.

\foreignlanguage 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 \bbl@hyphenation, 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.

\bbl@patterns

\edef\bbl@foreign@s#1#2{% TODO - shapemode, \@par, ?@par}\begingroup\let\BabelText@firstofone\\oreignlanguage/#1}%\bbl@usehooks{foreign*}{%\bbl@dirparastext\BabelText{#2}% Still in vertical mode!%}{\par}%\endgroup\foreignlanguage

\bbl@patterns
The environment `hyphenrules` can be used to select just the hyphenation rules. This environment does not change \language name 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*`.

The macro `\providehyphenmins` should be used in the language definition files to provide a default setting for the hyphenation parameters `\lefthyphenmin` and `\righthyphenmin`. If the macro `\lang`/hyphenmins is already defined this command has no effect.

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

The identification code for each file is something that was introduced in \TeX\ 2e. 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, i.e., on if the former is defined, we use a similar definition or not.
\begingroup
\catcode\`\ 10 \\
\@makeother\%
\ifnextchar[
{\@provideslanguage{#1}{#1}}{\@provideslanguage{#1}{#1}}
\edef\provideslanguage[#1]{#2}{%
\wlog{Language: \#1 \#2}
\expandafter\xdef\csname ver@#1.ldf\endcsname{#2}
\endgroup}

\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

\def\LdfInit{%
\chardef\atcatcode=\catcode\@ 
\catcode\@=11\relax 
\input babel.def\relax 
\catcode\@=\atcatcode \let\atcatcode\relax 
\LdfInit}

\originalTeX

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

\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, \begin{save}, is not considered to be undefined.

\iffalse\b babel@begin\@save\@undefined\let\b babel@begin\relax\fi

A few macro names are reserved for future releases of babel, which will use the concept of ‘locale’:

\providecommand\setlocale{%
\bbi@error 
{Not yet available}%
{Find an armchair, sit down and wait}
\let\uselocale\setlocale 
\let\locale\setlocale 
\let\selectlocale\setlocale 
\let\textlocale\setlocale 
\let\textlanguage\setlocale 
\let\language\setlocale 

11.2 Errors

\@nolanerr The babel package will signal an error when a documents tries to select a language that hasn’t been defined earlier. When a user selects a language for which no hyphenation patterns were loaded into the format he will be given a warning about that fact. We revert to the patterns for \language=0 in that case. In most formats that will be (US)english, but it might also be empty.

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

128
When the format knows about \PackageError it must be \ifx\PackageError\undefined, so we can safely use its error handling interface. Otherwise we'll have to ‘keep it simple’.

\def\bb@nulllanguage{\string\language=0}
\if\PackageError\undefined
  \begingroup
  \def\\{\^^J(babel)\}%
  \errhelp{\#2}\errmsg{\\#1}%
  \endgroup
\else
\def\bb@error#1#2{%}
  \begingroup
  \def\\{\^^J(babel)\}%
  \message{\\#1}%
  \endgroup
\else
\def\bb@warning#1{%}
  \begingroup
  \def\\{\^^J(babel)\}%
  \message{\\#1}%
  \endgroup
\fi
\if\packagewith{babel}{silent}
  \let\bb@info@gobble
  \let\bb@warning@gobble
\else
\def\bb@nocaption{
  \protect\bb@nocaption@i
\def\\@nocaption{i\#1\#2}{% : text to be printed 2: caption macro \langXname
  \global\namedef{\#2}{\textbf{?\#1?}}%
  \@nameuse{\#2}%
  \bb@warning{%
    \@backslashchar#2 not set. Please, define\%
    it in the preamble with something like:\%
    \string\renewcommand@backslashchar#2{..}\
    Reported}
  \def\\@tentative{i\#1}{%}
  \@tentative@i\#1{%
    \bb@warning{%
      Some functions for \#1 are tentative.\%
      They might not work as expected and their behavior\%
      could change in the future.\%
}
12 Loading hyphenation patterns

The following code is meant to be read by \ini\TeX{} because it should instruct \TeX{} to read hyphenation patterns. To this end the \texttt{docstrip} option \texttt{patterns} can be used to include this code in the file \texttt{hyphen.cfg}. Code is written with lower level macros.

We want to add a message to the message \MTEX{} 2.09 puts in the \texttt{everyjob} register. This could be done by the following code:

\begin{verbatim}
\let\orgeveryjob\everyjob
\def\everyjob#1{%
  \orgeveryjob{\begingroup\edef\noexpand\everyjob{#1}}%\endgroup%}
\let\orgeveryjob\everyjob
\def\orgeveryjob\expandafter{\the\orgeveryjob\immediate\write16{%
  hyphenation patterns for \the\loaded@patterns loaded.}}%
\let\everyjob\orgeveryjob

\end{verbatim}

The code above redefines the control sequence \texttt{\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 \MTEX{} fills the register.

There are some problems with this approach though.

- When someone wants to use several hyphenation patterns with \MTEX{} the above scheme won’t work. The reason is that \MTEX{} overwrites the contents of the \texttt{\everyjob} register with its own message.

- Plain \TeX{} does not use the \texttt{\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, \texttt{\dump{}}. Therefore the original \texttt{\dump{}} is saved in \texttt{\org\dump{}} and a new definition is supplied.

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


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.

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.

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.

The macro `process@language` is used to process a non-empty line from the ‘configuration file’. It has three arguments, each delimited by white space. The first argument is the ‘name’ of a language; the second is the name of the file that contains the patterns. The optional third argument is the name of a file containing hyphenation exceptions. The first thing to do is call `addlanguage` to allocate a pattern register and to make that register ‘active’. Then the pattern file is read. For some hyphenation patterns it is needed to load them with a specific font encoding selected. This can be specified in the file `language.dat` by adding for instance ‘:T1’ to the
name of the language. The macro \bbl@get@enc extracts the font encoding from the language name and stores it in \bbl@hyph@enc. The latter can be used in hyphenation files if you need to set a behavior depending on the given encoding (it is set to empty if no encoding is given).

Pattern files may contain assignments to \left@hyphen@min and \right@hyphen@min. \TeX{} does not keep track of these assignments. Therefore we try to detect such assignments and store them in the \texttt{(lang)hyphenmins} macro. When no assignments were made we provide a default setting.

Some pattern files contain changes to the \texttt{lc@code en uc@code} arrays. Such changes should remain local to the language; therefore we process the pattern file in a group; the \texttt{patterns} command acts globally so its effect will be remembered. Then we globally store the settings of \left@hyphen@min and \right@hyphen@min and close the group.

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

\bbl@languages saves a snapshot of the loaded languages in the form
\texttt{\bbl@elt\{language-name\}\{number\}\{patterns-file\}\{exceptions-file\}}. Note the last 2 arguments are empty in ‘dialects’ defined in language.dat with \texttt{=}.

Note also the language name can have encoding info.

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

\begin{verbatim}
\def\process@language#1#2#3{%
  \expandafter\addlanguage\csname l@#1\endcsname
  \expandafter\addlanguage\csname l@#1\endcsname
  \edef\language@name{#1}\%
  \ifx\bbl@hook\everylanguage\else
    \expandafter\xdef\csname #1\endcsname{%
      \the\left@hyphen@min\the\right@hyphen@min}%
  \fi
  \bbl@get@enc #1 : @
\endgroup
\def\bbl@elt#1#2#3{%
  \ifnum\left@hyphen@min=#1\relax
    \set@hyphen@min tw thr @ \relax
  \else
    \expandafter\expandafter\expandafter\set@hyphen@min
    \csname #1\endcsname
  \fi
  \the\toks@
  \toks@{}
\}%
\end{verbatim}
The macro `\bbl@get@enc` extracts the font encoding from the language name and stores it in `\bbl@hyph@enc`. It uses delimited arguments to achieve this.

```
def \bbl@get@enc #1 : #2 @ @ @ { \def \bbl@hyph@enc { #2 } }
```

Now, hooks are defined. For efficiency reasons, they are dealt here in a special way. Besides latex, format specific configuration files are taken into account.

```
def \bbl@hook@everylanguage #1 {}
def \bbl@hook@loadpatterns #1 { \input #1 \relax }
let \bbl@hook@loadexceptions \bbl@hook@loadpatterns
let \bbl@hook@loadkernel \bbl@hook@loadpatterns
begingroup
\def \AddBabelHook #1 #2 @ %
\expandafter ifx \csname bbl@hook@#2 \endcsname \relax
\def \next \toks1 @ %
\else
\def \next \expandafter gdef \csname bbl@hook@#2 \endcsname \###1 @ %
\fi
\next
\ifx directlua @ undefined
\ifx XeTeXinputencoding @ undefined \else
\input xebabel.def
\fi
\fi
\else
\input lualatex.def
\fi
\else
\input lualatex.def
\fi
openin1 = babel-bbl@format.cfg
\ifeof1
\message { I couldn’t find the file language.dat, \space I will try the file hyphen.tex }
\input hyphen.tex \relax
\chardef \l@english @
\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 $-1$.

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

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

```
\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.

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 `\bb@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.

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

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

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

Here the code for ini\TeX ends.
13 Font handling with fontspec

Add the bidi handler just before luatolatex, which is loaded by default by LaTeX. Just in

\begin{verbatim}
\ifodd\bbl@engine
\DeclareOption{bidi=basic-r}{}
\ExecuteOptions{bidi=basic}
\DeclareOption{bidi=basic}{}
{\let\bbl@beforeforeignleavevmode
% TODO - to locale_props, not as separate attribute
\newattribute\bbl@attr@dir
% I don't like it, hackish:
\frozen@everymath\expandafter{%}
\expandafter\bbl@mathboxdir\the\frozen@everymath%
\frozen@everydisplay\expandafter{%}
\expandafter\bbl@mathboxdir\the\frozen@everydisplay}
\bbl@exp{\output{\bodir\pagedir\the\output}}%
\AtEndOfPackage{\EnableBabelHook{babel-bidi}}}
\else
\DeclareOption{bidi=basic-r}{}
\ExecuteOptions{bidi=basic}
\DeclareOption{bidi=basic}{}
{\bbl@error
{The bidi method 'basic' is available only in\%
luatex. I'll continue with 'bidi=default', so\%
expect wrong results}%
{See the manual for further details.}%
\let\bbl@beforeforeignleavevmode
\AtEndOfPackage{%
\EnableBabelHook{babel-bidi}%
\bbl@xebidipar}}
\def\bbl@loadxebidi#1{%}
\ifx\RTLfootnotetext@undefined
\AtEndOfPackage{%
\EnableBabelHook{babel-bidi}%
\ifx\fontspec@undefined
\usepackage{fontspec}% bidi needs fontspec
\fi
\usepackage#1{bidi}}%
\fi
\fi}]
\DeclareOption{bidi=bidi}{}
{\bbl@tentative{bidi=bidi}%
{\bbl@loadxebidi{}}}
\DeclareOption{bidi=bidi-r}{}
{\bbl@tentative{bidi=bidi-r}%
{\bbl@loadxebidi{[rldocument]}}}
\DeclareOption{bidi=bidi-l}{}
{\bbl@tentative{bidi=bidi-l}%
{\bbl@loadxebidi{}}}
\fi
\DeclareOption{bidi=default}{}
{\let\bbl@beforeforeignleavevmode
\ifodd\bbl@engine
\newattribute\bbl@attr@dir
\bbl@exp{\output{\bodir\pagedir\the\output}}%
\fi
\end{verbatim}

135
With explicit languages, we could define the font at once, but we don't. Just wait and see if the language is actually activated.

If the family in the previous command does not exist, it must be defined. Here is how:

The following macro is activated when the hook babel-fonspec is enabled.
Now the macros defining the font with fontspec.
When there are repeated keys in fontspec, the last value wins. So, we just place the ini
settings at the beginning, and user settings will take precedence. We must deactivate
temporarily \bbl@mapselect because \selectfont is called internally when a font is
defined.
font@rst and famrst are only used when there is no global settings, to save and restore de previous families. Not really necessary, but done for optimization.

```
def@bbl@font@rst@#1@#2@#3@#4{%
 \bbl@csarg@def{famrst@#4}\{\bbl@font@set{#1}@#2@#3\}
```

The default font families. They are eurocentric, but the list can be expanded easily with \babelfont.

```
def@bbl@font@fams\{rm,sf,tt\}
```

The old tentative way. Short and preserved for compatibility, but deprecated. Note there is no direct alternative for \babelfeatures. The reason is explained in the user guide, but essentially – that was not the way to go :-).

```
def@babel@FS@store#1#2#3#4{%
def@csarg@edef@#2@default@#1@{#3}%
\expandafter@addto@csname@extras@#1@endcsname{%
 \edef@#3@\csname@babel@#2@default@#1@endcsname%\}
 \fontfamily@#3\selectfont
 \else
 \edef@#3@\csname@babel@#2@default@#1@endcsname%
 \fi%
\expandafter@addto@csname@noextras@#1@endcsname{%
 \ift@family
 \fontfamily@#4\selectfont
 \fi
 \let@#3@endcsname{}
\let@babel@FS@features@empty
\def@babelfeatures{% make sure @fontspec is redefined once
 \let@babel@ori@fontspec@fontspec
 \renewcommand@fontspec@#1%}
 \bbl@ori@fontspec\{\bbl@langfeatures@#1\}
\let@babelfeatures@bbl@FS@features
\babelfeatures
```

⟨⟨/Font selection⟩⟩
14 Hooks for XeTeX and LuaTeX

14.1 XeTeX

Unfortunately, the current encoding cannot be retrieved and therefore it is reset always to utf8, which seems as 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(a)tex.ini, so here we just “undo” some of the changes done by \LaTeX. Anyway, for consistency Lua\TeX{} also resets the catcodes.

\begin{verbatim}
\begin{group}
% Reset chars "80-"C0 to category "other", no case mapping:
\catcode`@=11 \count@=128
\loop\ifnum\count@<192
  \global\uccode\count@=0 \global\lccode\count@=0
  \global\ccode\count@=12 \global\sfcode\count@=1000
  \advance\count@ by 1 \repeat
% Other:
\def\0 ### 1 %
  \global\uccode##1=0 \global\lccode##1=0
  \global\ccode##1=12 \global\sfcode##1=1000 %
% Letter:
\def\L ### 1 #2 #3 {
  \global\uccode##1=11
  \global\lccode##1=#2
  \global\ccode##1=#3
% Uppercase letters have sfc ode=999:
  \ifnum##1=##3 \else \global\sfc ode##1=999 \fi %
% Letter without case mappings:
\def\l ### 1 \{ \L ### 1 ### 1 ### 1 \}%
\L 00AA
\L 00B5 039C 00B5
\L 00BA
\L 00D7
\L 00DF
\L 00F7
\L 00FF 0178 00FF
\end{group}
\input #1!relax
\end{verbatim}

Some more common code.

\begin{verbatim}
\begin{group}
% Footnote changes
\bb@lst{Bidi footnotes}
\ifx\bb@beforeforeign\leavevmode
\def\bb@fntexto##1#2#3(%
  \ifnextchar\{
    \bb@fntexto##1#2\{#3}%%%%
    \bb@fntexto##1#2\{#3\
  \def\bb@fntexto##1#2#3#4(%
  \def\bb@fntexto##1#2#3#4%\}
  \selectlanguage\{\bb@main\}
  \bb@fn@footnote##1#2#3#4#5://%
  \bb@fn@footnote##1#2#3#4#5%\}
\end{verbatim}

139
Now, the code.

\def\BabelStringsDefault{unicode}
\let\webbl@stop\relax
\AddBabelHook{xtex}{encodedcommands}{}
\def\webbl@tempa{1}{}
\if\webbl@tempa\empty
\XeTeXinputencoding"bytes"
\else
\XeTeXinputencoding"#1"
\fi
\def\webbl@stop{\XeTeXinputencoding"utf8"}
\AddBabelHook{xtex}{stopcommands}{}
\webbl@stop
\let\webbl@stop\relax
\def\webbl@intraspaces#1 #2 #3\@{\%
\webbl@csarg@gdef{xesp@\webbl@cs{sbcps@\lanugename}#1}{\%
{\XeTeXlinebreakskipskip #1em plus #2em minus #3em}\relax}
\def\webbl@intrapenalty#1\@{\%
\webbl@csarg@gdef{xepn@\webbl@cs{sbcps@\lanugename}}{\%
{\XeTeXlinebreakpenalty #1}\relax}
\AddBabelHook{xtex}{loadkernel}{}
\AddBabelHook{xtex}{\RestoreUnicodecatcodesbeforeloadingpatterns}}
\if\DisableBabelHook\undefined\endinput\fi
\DisableBabelHook{babel-fontspec}{afterextras}{\webbl@switchfont}
\DisableBabelHook{babel-fontspec}
14.2 Layout

In progress.
Note elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titleps, and geometry.
\texttt{\textbackslash{}bbl\textbackslash{}startskip} and \texttt{\textbackslash{}bbl\textbackslash{}endskip} are available to package authors. Thanks to the \TeX\ expansion mechanism the following constructs are valid: \texttt{\textbackslash{}adim\textbackslash{}bbl\textbackslash{}startskip, \textbackslash{}advance\textbackslash{}bbl\textbackslash{}startskip}\texttt{\textbackslash{}adim, \textbackslash{}bbl\textbackslash{}startskip}\texttt{\textbackslash{}adim}.
Consider \texttt{txtbabel} as a shorthand for \texttt{tex-xet babel}, which is the bidi model in both pdftex and xetex.

\begin{verbatim}
\input{txtbاملballs}
14.2 Layout

In progress.
Note elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titleps, and geometry.
\texttt{\textbackslash{}bbl\textbackslash{}startskip} and \texttt{\textbackslash{}bbl\textbackslash{}endskip} are available to package authors. Thanks to the \TeX\ expansion mechanism the following constructs are valid: \texttt{\textbackslash{}adim\textbackslash{}bbl\textbackslash{}startskip, \textbackslash{}advance\textbackslash{}bbl\textbackslash{}startskip}\texttt{\textbackslash{}adim, \textbackslash{}bbl\textbackslash{}startskip}\texttt{\textbackslash{}adim}.
Consider \texttt{txtbabel} as a shorthand for \texttt{tex-xet babel}, which is the bidi model in both pdftex and xetex.

\begin{verbatim}
\input{txtbاملballs}
\end{verbatim}
\end{verbatim}
Implicitly reverses sectioning labels in bidi=basic-r, because the full stop is not in contact with L numbers any more. I think there must be a better way.
14.3 LuaTeX

The new loader for luatex is based solely on language.dat, which is read on the fly. The code shouldn't be executed when the format is build, so we check if \AddBabelHook is defined. Then comes a modified version of the loader in hyphen.cfg (without the hyphenmin stuff, which is under the direct control of babel).

The names \l@<language> are defined and take some value from the beginning because all ld files assume this for the corresponding language to be considered valid, but patterns are not loaded (except the first one). This is done later, when the language is first selected (which usually means when the ld finishes). If a language has been loaded, \bbl@hypendata@<num> exists (with the names of the files read).

The default setup preloads the first language into the format. This is intended mainly for ‘english’, so that it’s available without further intervention from the user. To avoid duplicating it, the following rule applies: if the “0th” language and the first language in language.dat have the same name then just ignore the latter. If there are new synonymous, the are added, but note if the language patterns have not been preloaded they won’t at run time.

Other preloaded languages could be read twice, if they has been preloaded into the format. This is not optimal, but it shouldn’t happen very often – with luatex patterns are best loaded when the document is typeset, and the “0th” language is preloaded just for backwards compatibility.

As of 1.1b, lua(e)tex is taken into account. Formerly, loading of patterns on the fly didn’t work in this format, but with the new loader it does. Unfortunately, the format is not based on babel, and data could be duplicated, because languages are reassigned above those in the format (nothing serious, anyway). Note even with this format language.dat is used (under the principle of a single source), instead of language.def.

Of course, there is room for improvements, like tools to read and reassign languages, which would require modifying the language list, and better error handling.

We need catcode tables, but no format (targeted by babel) provide a command to allocate them (although there are packages like tablestack). For the moment, a dangerous approach is used – just allocate a high random number and cross the fingers. To complicate things, etex.sty changes the way languages are allocated.

```latex
\begin{verbatim}
\ifx\AddBabelHook\undefined
\bbl@trace{Read language.dat}
\begin{group}
\toks={} \count@% 0=start, 1=0th, 2=normal
\def\bbl@process@line#1#2 #3 #4 {%
  \ifx=#1%
  \bbl@process@synonym(#2)%
  \else
  \bbl@process@language(#1#2){#3}{#4}%
  \fi
  \ignorespaces}
\def\bbl@manylang{%
  \ifnum\bbl@last>@ne
  \bbl@info{Non-standard hyphenation setup}%
  \fi
  \let\bbl@manylang\relax}
\def\bbl@process@language#1#2#3{%
  \ifcase\count@
  \ifundefined{zth@#1}{\count@\tw@}{\count@@ne}%
  \or
  \count@\tw@
  \fi
\end{verbatim}
```
\ifnum\count@=\tw@
\expandafter\addlanguage\csname l@#1\endcsname
\language\allocationnumber
\chardef\bbl@last\allocationnumber
\bbl@manylang
\let\bbl@elt\relax
\xdef\bbl@languages{%
\bbl@languages\bbl@elt{#1}{the\language}{#2}{#3}%
\fi
\the\toks@
\toks@{}}
\def\bbl@process@synonym@aux#1#2{%
\global\expandafter\chardef\csname l@#1\endcsname\relax
\let\bbl@elt\relax
\xdef\bbl@languages{%
\bbl@languages\bbl@elt{#1}{#2}{}}
\def\bbl@process@synonym@aux#1{%
\ifcase\count@
\toks@\expandafter{\the\toks@\relax}\bbl@process@synonym@aux{#1}{}% 
\or
\atodef\bbl@process@synonym@aux{#1}{0}{})}{}
\else
\bbl@process@synonym@aux{#1}{the\bbl@last}%
\fi}
\ife\bbl@languages\undefined % Just a (sensible?) guess
\chardef\l@english\z@
\chardef\l@USenglish\z@
\chardef\bbl@last\z@
\global\edef\bbl@hyphendata@0\{\{hyphen.tex\}\{\{hyphen.tex\}\}
\gdef\bbl@languages{%
\bbl@elt\l@english\{0\}{hyphen.tex\}
\bbl@elt\l@USenglish\{0\}\{\}
\global\let\bbl@languages\bbl@languages
\else
\global\let\bbl@languages\format\bbl@languages
\def\bbl@elt#1#2#3#4{% Remove all except language 0
\ifnum#2>1@\else
\noexpand\bbl@elt{#1}{#2}{#3}{#4}%
\fi%
\fi}
\edef\bbl@languages{%
\def\bbl@elt#1#2#3#4{%\ paraphdef\zth@#1\}{} % Define flags
\bbl@languages
\openin1=language.dat
\ifeof!
\bbl@warning(I couldn’t find language.dat. No additional
% patterns loaded. Reported)%
\else
\loop
\read1 to \bbl@line
\endlinechar\mone
\if \vline \fi \vline \relax
\ifx\bbl@line@empty\else
\edef\bbl@line\{\bbl@line\space\space\space}\% 
\expandafter\bbl@line\line\bbl@line\relax
\fi
\repeat
\fi
\endgroup
\bbl@trace{Macros for reading patterns files}
\def\bbl@get@enc#1:#2:#3@{
\bbl@get@hyph@enc{#2}}
\ifx\bbl@catcodetablenum@undefined
\\bbl@catcodetablenum{(5211)}
\fi
\bbl@luapatterns#1#2{%
\bbl@get@enc{#1}:
\setbox\@hbox\bgroup
\beginninggroup
\ifx\catcode@table@undefined
\let\save@cat@code@table\lua@ex@save@cat@code@table
\let\init@cat@code@table\lua@ex@init@cat@code@table
\let\cat@code@table\lua@ex@cat@code@table
\fi
\save@cat@code@table\bbl@cat@code@tablenum\relax
\init@cat@code@table\numexpr\bbl@cat@code@tablenum+1\relax
\catcode`\%=12
\catcode`\'=12
\catcode`\%=12
\catcode`\&=12
\catcode`\%=12
\catcode`\%=12
\catcode`\%=12
\catcode`\%=12
\catcode`\%=12
\catcode`\%=12
\catcode`\%=12
\catcode`\%=12
\catcode`\%=12
\input #1\relax
\catcode@table\bbl@cat@code@tablenum\relax
\endgroup
\def\bbl@tempa{#2}%
\if\bbl@tempa@empty\else
\input #2\relax
\fi
\egroup}%
\def\bbl@patterns@lua#{%
\language=\expandafter\if\csname l@#1:f@encoding\endcsname\relax
\csname l@#1\endcsname
\edef\bbl@tempa{#1}%
\else
\csname l@#1:f@encoding\endcsname
\edef\bbl@tempa{#1:f@encoding}%
\fi\relax
\@namedef{\lu@text@hyphen@loaded@the\language}{}% Temp
\@ifundefined{\bbl@hyphen@data@the\language}%
{\def\bbl@elt@#1@#2@#3@#4@{%
\ifnum#2=\csname l@\bbl@tempa\endcsname % #2=spanish, dutch:OT1...
\def\bbl@tempb{#3}%
\if\bbl@tempb@empty\else % if not a synonymous
\def\bbl@tempc{#3}{}%
\fi
\bbl@csarg@\edef\hyphen@data@#2@{\bbl@tempc}%
\fi}%
\bbl@languages
\@ifundefined{\bbl@hyphen@data@the\language}%
{\{\bbl@info{No hyphenation patterns were set for\%\%language "\bbl@tempa", Reported}\%
\expandafter\expandafter\expandafter\expandafter\bbl@luapatterns
\csname bbl@hyphen@data@the\language@endcsname\}{}}
\endinput\fi
\beginninggroup
\catcode`\%=12
\catcode`\'=12
Babel = Babel or {}

function Babel.bytes(line)
  return line.gsub("(.)", function (chr) return unicode.utf8.char(string.byte(chr)) end)
end

function Babel.begin_process_input()
  if luatexbase and luatexbase.add_to_callback then
    luatexbase.add_to_callback('process_input_buffer',
      Babel.bytes,'Babel.bytes')
  else
    Babel.callback = callback.find('process_input_buffer')
    callback.register('process_input_buffer',Babel.bytes)
  end
end

function Babel.end_process_input()
  if luatexbase and luatexbase.remove_from_callback then
    luatexbase.remove_from_callback('process_input_buffer','Babel.bytes')
  else
    callback.register('process_input_buffer',Babel.callback)
  end
end

function Babel.addpatterns(pp, lg)
  local lg = lang.new(lg)
  local pats = lang.patterns(lg) or ''
  lang.clear_patterns(lg)
  for p in pp:gmatch('[^%s]+') do
    ss = ''
    for i in string.utfcharacters(p:gsub('%d', '')) do
      ss = ss .. '%d?' .. i
    end
    ss = ss:gsub('^[%.]%d[%?]$', '^[%.]')
    pats, n = pats:gsub('%s .. ss .. %s', [[.. p .. ]])
    if n == 0 then
      tex.sprint(
        [[\string\csname space bbl@info\endscncname(New pattern: ]]
        .. p .. [[{}]])
      pats = pats .. ' ' .. p
    else
      tex.sprint(
        [[\string\csname space bbl@info\endscncname(Renew pattern: ]]
        .. p .. [[{}]])
    end
  end
  lang.patterns(lg, pats)
end

endgroup

if\newattribute@undefined\else
\newattribute\bbl@attr@locale
\AddBabelHook{luatex}{beforeextras}{%
  \setattribute\bbl@attr@locale\localeid}
\fi
\def\BabelStringsDefault{unicode}
\let\luabbl@stop\relax
\AddBabelHook{luatex}{encodedcommands}{%
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.


14.4 Southeast Asian scripts

In progress. Replace regular (ie, implicit) discretionary by spaceskips, based on the previous glyph (which I think makes sense, because the hyphen and the previous char go always together). Other discretionary are not touched.
For the moment, only 3 SA languages are activated by default (see Unicode UAX 14).

\directlua{
Babel = Babel or {} Babel.intraspaces = Babel.intraspaces or {}
Babel.intraspaces['\csname bbl@sbcp\languagesname\endcsname'] = %
{b = #1, p = #2, m = #3} Babel.locale_props[the\localeid].intraspace = %
{b = #1, p = #2, m = #3}
}
\directlua{
Babel = Babel or {} Babel.intrapenalties = Babel.intrapenalties or {}
Babel.intrapenalties['\csname bbl@sbcp\languagesname\endcsname'] = #1
Babel.locale_props[the\localeid].intrapenalty = #1
}
\begingroup
\catcode`\%=12 \catcode`\^=14 \catcode`\'=12 \catcode`\-=12 \gdef\bbl@seaintraspace^\let\bbl@seaintraspace\relax
\directlua{
Babel = Babel or {} Babel.sea_enabled = true
}
Babel.sea_ranges = Babel.sea_ranges or {}

function Babel.set_chranges (script, chrng)
    local c = 0
    for s, e in string.gmatch(chrng..' ', '(.%)(.%)') do
        Babel.sea_ranges[script..c] = {tonumber(s,16), tonumber(e,16)}
        c = c + 1
    end
end

function Babel.set_disc_to_space (head)
    local sea_ranges = Babel.sea_ranges
    local last_char = nil
    local quad = 655360
    local pt = 655360 = 10 * 65536
    for item in node.traverse(head) do
        local i = item.id
        if i == node.id'glyph' then
            last_char = item
        elseif i == 7 and item.subtype == 3 and last_char
            and last_char.char > 0xC99 then
            quad = font.getfont(last_char.font).size
            for lg, rg in pairs(sea_ranges) do
                if last_char.char > rg[1] and last_char.char < rg[2] then
                    lg = lg:sub(1, 4)
                    local intraspace = Babel.intraspaces[lg]
                    local intrapenalty = Babel.intrapenalties[lg]
                    local n
                    if intrapenalty ~= 0 then
                        n = node.new(14, 0)  ^^ penalty
                        n.penalty = intrapenalty
                        node.insert_before(head, item, n)
                    end
                    n = node.new(12, 13)  ^^ (glue, spaceskip)
                    node.setglue(n, intraspace.b * quad,
                        intraspace.p * quad,
                        intraspace.m * quad)
                    node.insert_before(head, item, n)
                    node.remove(head, item)
                end
            end
        end
    end
end

bbox @ luahyphenate
\catcode `%=14
\def\bbox@cjkintraspacemap %
\let\bbox@cjkintraspacemap\relax
\directlua{
    Babel = Babel or {}
    require'babel-data-cjk.lua'
    Babel.cjk_enabled = true
    function Babel.cjk_linebreak(head)
        local GLYPH = node.id'glyph'
        local last_char = nil
        local quad = 655360
        local pt = 655360 = 10 * 65536
        local last_class = nil
        for item in node.traverse(head) do
            if item.id == GLYPH then
                break
            end
        end
        return Babel
    end
}
local LOCALE = node.get_attribute(item, 
    luatexbase.registernumber('bbl@attr@locale'))

local props = Babel.locale_props[LOCALE]

class = Babel.cjk_class[item.char].c

if class == 'cp' then class = 'cl' end 
if class == 'id' then class = 'I' end

if class and last_class and Babel.cjk_breaks[last_class][class] then
    br = Babel.cjk_breaks[last_class][class]
else
    br = 0
end

if br == 1 and props.linebreak == 'c' then
    local intrapenalty = props.intrapenalty
    if intrapenalty == 0 then
        local n = node.new(14, 0) % penalty
        n.penalty = intrapenalty
        node.insert_before(head, item, n)
    end

    local intraspace = props.intraspace
    local n = node.new(12, 13) % (glue, spaceskip)
    node.setglue(n, intraspace.b * quad,
                intraspace.p * quad,
                intraspace.m * quad)

    node.insert_before(head, item, n)
end

quad = font.getfont(item.font).size
last_class = class
else % if penalty, glue or anything else
    last_class = nil
end

end

lang.hyphenate(head)
end

\bbl@luahyphenate}
\def\bbl@luahyphenate{%
\let\bbl@luahyphenate\relax
\directlua{
  luatexbase.add_to_callback('hyphenate',
    function (head, tail)
      if Babel.cjk_enabled then
        Babel.cjk_linebreak(head)
      end
      lang.hyphenate(head)
      if Babel.sea_enabled then
        Babel.sea_disc_to_space(head)
      end
    end,
    'Babel.hyphenate')
}
\endgroup
14.5 CJK line breaking

Minimal line breaking for CJK scripts, mainly intended by simple documents and short texts as a secondary language. Only line breaking, with a little stretching for justification, without any attempt to adjust the spacing. It is based (but does not strictly follow) on the Unicode algorithm.

We first need a little table with the corresponding line breaking properties. A few characters have an additional key for the width (fullwidth vs. halfwidth), not yet used. There is a separate file, defined below.

Work in progress.

Common stuff.

Temporary fix for luatex <1.10, which sometimes inserted a spurious closing dir node with a \textdir within \hboxes. This will be eventually removed.

14.6 Layout

Work in progress.

Unlike xetex, luatex requires only minimal changes for right-to-left layouts, particularly in monolingual documents (the engine itself reverses boxes – including column order or headings –, margins, etc.) and with bidi=basic-r, without having to patch almost any macro where text direction is relevant.

\@hangfrom is useful in many contexts and it is redefined always with the layout option. There are, however, a number of issues when the text direction is not the same as the box direction (as set by \bodydir), and when \parbox and \hangindent are involved. Fortunately, latest releases of luatex simplify a lot the solution with \shape{mode}.

4065 \AddBabelHook\{luatex\}{loadkernel}{%
4066 {\langle Restore Unicode catcodes before loading patterns\rangle\} \}
4067 \if\DisableBabelHook\@undefined\endinput\fi
4068 \AddBabelHook\{babel-fontspec\}{afterextras}{\bb@switchfont}
4069 \DisableBabelHook\{babel-fontspec\}
4070 {\langle Font selection\rangle\} \}

Temporary fix for luatex <1.10, which sometimes inserted a spurious closing dir node with a \textdir within \hboxes. This will be eventually removed.

4071 \def\bb@luafixboxdir{%
4072 \setbox2@\hbox{\textdir TLT}\%
4073 \directlua{
4074 function Babel.first_dir(head)
4075 for item in node.traverse_id(node.id'dir', head) do
4076 return item
4077 end
4078 return nil
4079 end
4080 if Babel.first_dir(tex.box[0].head) then
4081 function Babel.fixboxdirs(head)
4082 local fd = Babel.first_dir(head)
4083 if fd and fd.dir:sub(1,1) == '-' then
4084 head = node.remove(head, fd)
4085 end
4086 return head
4087 end
4088 end
4089 }}
4090 \AtBeginDocument{\bb@luafixboxdir}

14.6 Layout

Work in progress.

Unlike xetex, luatex requires only minimal changes for right-to-left layouts, particularly in monolingual documents (the engine itself reverses boxes – including column order or headings –, margins, etc.) and with bidi=basic-r, without having to patch almost any macro where text direction is relevant.

\@hangfrom is useful in many contexts and it is redefined always with the layout option. There are, however, a number of issues when the text direction is not the same as the box direction (as set by \bodydir), and when \parbox and \hangindent are involved. Fortunately, latest releases of luatex simplify a lot the solution with \shape{mode}.
Implicitly reverses sectioning labels in bidi=basic-r, because the full stop is not in contact with L numbers any more. I think there must be a better way. Assumes bidi=basic-r, but there are some additional readjustments for bidi=default.
14.7 Auto bidi with basic and basic-r

The file babel-data-bidi.lua currently only contains data. It is a large and boring file and it's not shown here. See the generated file.

Now the basic-r bidi mode. One of the aims is to implement a fast and simple bidi algorithm, with a single loop. I managed to do it for R texts, with a second smaller loop for a special case. The code is still somewhat chaotic, but its behavior is essentially correct. I cannot resist copying the following text from Emacs bidi.c (which also attempts to implement the bidi algorithm with a single loop):

Arrrgh!! The UAX#9 algorithm is too deeply entrenched in the assumption of batch-style processing [...]. May the fleas of a thousand camels infest the armpits of those who design supposedly general-purpose algorithms by looking at their own implementations, and fail to consider other possible implementations!

Well, it took me some time to guess what the batch rules in UAX#9 actually mean (in other word, what they do and why, and not only how), but I think (or I hope) I've managed to understand them.

In some sense, there are two bidi modes, one for numbers, and the other for text.
Furthermore, setting just the direction in R text is not enough, because there are actually two R modes (set explicitly in Unicode with RLM and ALM). In babel the dir is set by a higher protocol based on the language/script, which in turn sets the correct dir (<l>, <r> or <al>).

From UAX#9: “Where available, markup should be used instead of the explicit formatting characters”. So, this simple version just ignores formatting characters. Actually, most of that annex is devoted to how to handle them.
BD14-BD16 are not implemented. Unicode (and the W3C) are making a great effort to deal with some special problematic cases in “streamed” plain text. I don't think this is the way to go – particular issues should be fixed by a high level interface taking into account the needs of the document. And here is where luatex excels, because everything related to bidi writing is under our control.

```
4176 ⟨∗basic-r⟩
4177 Babel = Babel or {}
4178 Babel bidi_enabled = true
4179 require('babel-data-bidi.lua')
4180 local characters = Babel.characters
4181 local ranges = Babel.ranges
4182 local DIR = node.id("dir")

4188 local function dir_mark(head, from, to, outer)
4189   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
4190   local d = node.new(DIR)
4191   d.dir = '+' .. dir
4192   node.insert_before(head, from, d)
4193   d = node.new(DIR)
4194   d.dir = '-' .. dir
4195   node.insert_after(head, to, d)
4196 end

4198 function Babel bidi(head, ispar)
4199   local first_n, last_n -- first and last char with nums
4200   local last_es -- an auxiliary 'last' used with nums
4201   local first_d, last_d -- first and last char in L/R block
4202   local dir, dir_real

4203   local strong = ('TRT' == text.pardir) and 'r' or 'l'
4204   local strong_lr = (strong == 'l') and 'l' or 'r'
4205   local outer = strong
4206   local new_dir = false
4207   local first_dir = false
4208   local inmath = false
4209
4210   for item in node.traverse(head) do
4211     -- three cases: glyph, dir, otherwise
4212     if item.id == node.id('glyph')
4213       or (item.id == 7 and item.subtype == 2) then
4214       local itemchar
4215       if item.id == 7 and item.subtype == 2 then
4216         itemchar = item.replace.char
4217       else
4218         itemchar
```

itemchar = item.char
end
local chardata = characters[itemchar]
dir = chardata and chardata.d or nil
if not dir then
  for nn, et in ipairs(ranges) do
    if itemchar < et[1] then
      break
    elseif itemchar <= et[2] then
      dir = et[3]
      break
    end
  end
  end
end

if inmath then dir = ('TRT' == tex.mathdir) and 'r' or 'l' end

Next is based on the assumption babel sets the language AND switches the script with its dir. We treat a language block as a separate Unicode sequence. The following piece of code is executed at the first glyph after a ‘dir’ node. We don’t know the current language until then. This is not exactly true, as the math mode may insert explicit dirs in the node list, so, for the moment there is a hack by brute force (just above).

if new_dir then
  attr_dir = 0
  for at in node.traverse(item.attr) do
    if at.number == luatexbase.registernumber'bbl@attr@dir' then
      attr_dir = at.value % 3
    end
  end
  if attr_dir == 1 then
    strong = 'r'
  elseif attr_dir == 2 then
    strong = 'al'
  else
    strong = 'l'
  end
  strong_lr = (strong == 'l') and 'l' or 'r'
  outer = strong_lr
  new_dir = false
end

if dir == 'nsm' then dir = strong end -- W1

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

dir_real = dir -- We need dir_real to set strong below
if dir == 'al' then dir = 'r' end -- W3

By W2, there are no <en> <et> <es> if strong == <al>, only <an>. Therefore, there are not <et en> nor <en et>, W5 can be ignored, and W6 applied:

if strong == 'al' then
  if dir == 'en' then dir = 'an' end -- W2
  if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
  strong_lr = 'r'
  end

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

elseif item.id == node.id'dir' and not inmath then
  new_dir = true

156
 Numbers in R mode. A sequence of <en>, <et>, <an>, <es> and <cs> is typeset (with some rules) in L mode. We store the starting and ending points, and only when anything different is found (including nil, ie, a non-char), the textdir is set. This means you cannot insert, say, a whatsit, but this is what I would expect (with luacolor you may colorize some digits). Anyway, this behavior could be changed with a switch in the future. Note in the first branch only <an> is relevant if <al>.

```c
if dir == 'en' or dir == 'an' or dir == 'et' then
    if dir ~= 'et' then
        type_n = dir
    end
    first_n = first_n or item
    last_n = last_es or item
    last_es = nil
    elseif dir == 'es' and last_n then -- W3+W6
        last_es = item
    elseif dir == 'cs' then -- it's right - do nothing
    elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
        if strong_lr == 'r' and type_n ~= '' then
            dir_mark(head, first_n, last_n, 'r')
        elseif strong_lr == 'l' and first_d and type_n == 'an' then
            dir_mark(head, first_n, last_n, 'r')
            dir_mark(head, first_d, last_d, outer)
            first_d, last_d = nil, nil
        elseif strong_lr == 'l' and type_n ~= '' then
            last_d = last_n
        end
        type_n = ''
    first_n, last_n = nil, nil
end
```

R text in L, or L text in R. Order of dir_mark's are relevant: d goes outside n, and therefore it's emitted after. See dir_mark to understand why (but is the nesting actually necessary or is a flat dir structure enough?). Only L, R (and AL) chars are taken into account – everything else, including spaces, whatsit's, etc., are ignored:

```c
if dir == 'l' or dir == 'r' then
    if dir ~= outer then
        first_d = first_d or item
        last_d = item
    elseif first_d and dir ~= strong_lr then
        dir_mark(head, first_d, last_d, outer)
        first_d, last_d = nil, nil
    end
end
```

**Mirroring.** Each chunk of text in a certain language is considered a “closed” sequence. If <r on r> and <l on l>, it’s clearly <r> and <l>, resp, but with other combinations depends on outer. From all these, we select only those resolving <on> → <r>. At the beginning (when last_lr is nil) of an R text, they are mirrored directly. TODO - numbers in R mode are processed. It doesn't hurt, but should not be done.

```c
if dir and not last_lr and dir == 'l' and outer == 'r' then
    item.char = characters[item.char] and
```
characters[item.char].m or item.char

elseif (dir or new_dir) and last_lr = item then
    local mir = outer .. strong_lr .. (dir or outer)
    if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
        for ch in node.traverse(node.next(last_lr)) do
            if ch == item then break end
            if ch.id == node.id'glyph' then
                ch.char = characters[ch.char].m or ch.char
            end
        end
    end
end
end
end
end
end
end

Save some values for the next iteration. If the current node is ‘dir’, open a new sequence. Since dir could be changed, strong is set with its real value (dir_real).

if dir == 'l' or dir == 'r' then
    last_lr = item
    strong = dir_real
    -- Don't search back - best save now
    strong_lr = (strong == 'l') and 'l' or 'r'
else new_dir then
    last_lr = nil
end
end

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

if last_lr and outer == 'r' then
    for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
       ch.char = characters[ch.char].m or ch.char
    end
end

if first_n then
    dir_mark(head, first_n, last_n, outer)
end
if first_d then
    dir_mark(head, first_d, last_d, outer)
end

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

return node.prev(head) or head
end
end
</basic-r>

And here the Lua code for bidi=basic:

长途(=basic)
Babel = Babel or {}
-- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
Babel.fontmap = Babel.fontmap or {}
Babel.fontmap[0] = {} -- l
Babel.fontmap[1] = {} -- r
Babel.bidi_enabled = true
require('babel-data-bidi.lua')
local characters = Babel.characters
local ranges = Babel.ranges
local DIR = node.id('dir')
local GLYPH = node.id('glyph')
local function insert_implicit(head, state, outer)
    local new_state = state
    if state.sim and state.eim and state.sim ~= state.eim then
        dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
        local d = node.new(DIR)
        d.dir = '+' .. dir
        node.insert_before(head, state.sim, d)
        local d = node.new(DIR)
        d.dir = '-' .. dir
        node.insert_after(head, state.eim, d)
    end
    new_state.sim, new_state.eim = nil, nil
    return head, new_state
end

local function insert_numeric(head, state)
    local new_state = state
    if state.san and state.ean and state.san ~= state.ean then
        local d = node.new(DIR)
        d.dir = '+TLT'
        _, new = node.insert_before(head, state.san, d)
        if state.san == state.sim then state.sim = new end
        local d = node.new(DIR)
        d.dir = '-TLT'
        _, new = node.insert_after(head, state.ean, d)
        if state.ean == state.eim then state.eim = new end
    end
    new_state.san, new_state.ean = nil, nil
    return head, new_state
end

-- TODO - \hbox with an explicit dir can lead to wrong results
-- \<R \hbox dir TLT\{<R}> and \<L \hbox dir TRT\{<L}>. A small attempt
-- was made to improve the situation, but the problem is the 3-dir
-- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
-- well.

function Babel.bidi(head, ispar, hdir)
    local d -- d is used mainly for computations in a loop
    local prev_d = ''
    local new_d = false
    local nodes = {}
    local outer_first = nil
    local inmath = false
    local glue_d = nil
    local glue_i = nil
    local has_en = false
    local first_et = nil
    local ATDIR = luatexbase.registernumber'bbl@attr@dir'
local save_outer
local temp = node.get_attribute(head, ATDIR)
if temp then
    temp = temp % 3
    save_outer = (temp == 0 and 'l') or
        (temp == 1 and 'r') or
        (temp == 2 and 'al')
elseif ispar then  -- Or error? Shouldn't happen
    save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
else  -- Or error? Shouldn't happen
    save_outer = ('TRT' == hdire) and 'r' or 'l'
end
-- when the callback is called, we are just _after_ the box,
-- and the textdir is that of the surrounding text
-- if not ispar and hdire == tex.textdir then
--   save_outer = ('TRT' == hdire) and 'r' or 'l'
-- end
local outer = save_outer
local last = outer
'al' is only taken into account in the first, current loop
if save_outer == 'al' then save_outer = 'r' end
local fontmap = Babel.fontmap
for item in node.traverse(head) do
    -- In what follows, #node is the last (previous) node, because the
    -- current one is not added until we start processing the neutrals.
    -- three cases: glyph, dir, otherwise
    if item.id == GLYPH
        or (item.id == 7 and item.subtype == 2) then
        local d_font = nil
        local item_r
        if item.id == 7 and item.subtype == 2 then
            item_r = item.replace  -- automatic discs have just 1 glyph
        else
            item_r = item
        end
        local chardata = characters[item_r.char]
        d = chardata and chardata.d or nil
        if not d or d == 'nsm' then
            for nn, et in ipairs(ranges) do
                if item_r.char < et[1] then
                    break
                elseif item_r.char <= et[2] then
                    if not d then d = et[3]
                    elseif d == 'nsm' then d_font = et[3]
                    end
                    break
                end
            end
            d = d or 'l'
            d_font = d_font or d
        end
    end
end
d_font = (d_font == 'l' and 0) or
(d_font == 'nsm' and 0) or
(d_font == 'r' and 1) or
(d_font == 'al' and 2) or
(d_font == 'an' and 2) or nil
if d_font and fontmap and fontmap[d_font][item_r.font] then
  item_r.font = fontmap[d_font][item_r.font]
end

if new_d then
  table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
  if inmath then
    attr_d = 0
  else
    attr_d = node.get_attribute(item, ATDIR)
    attr_d = attr_d % 3
  end
  if attr_d == 1 then
    outer_first = 'r'
    last = 'r'
  elseif attr_d == 2 then
    outer_first = 'r'
    last = 'al'
  else
    outer_first = 'l'
    last = 'l'
  end
  outer = last
  has_en = false
  first_et = nil
  new_d = false
end

if glue_d then
  if (d == 'l' and 'l' or 'r') ~= glue_d then
    table.insert(nodes, {glue_i, 'on', nil})
  end
  glue_d = nil
  glue_i = nil
end

elseif item.id == DIR then
  d = nil
  new_d = true
elseif item.id == node.id'glue' and item.subtype == 13 then
  glue_d = d
  glue_i = item
  d = nil
elseif item.id == node.id'math' then
  inmath = (item.subtype == 0)
else
  d = nil
end

-- AL <= EN/ET/ES -- W2 + W3 + W6
if last == 'al' and d == 'en' then
d = 'an' -- W3
elseif last == 'al' and (d == 'et' or d == 'es') then
    d = 'on' -- W6
end

-- EN + CS/ES + EN -- W4
if d == 'en' and #nodes >= 2 then
    if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
        and nodes[#nodes-1][2] == 'en' then
        nodes[#nodes][2] = 'en'
    end
end

-- AN + CS + AN -- W4 too, because uax9 mixes both cases
if d == 'an' and #nodes >= 2 then
    if (nodes[#nodes][2] == 'cs')
        and nodes[#nodes-1][2] == 'an' then
        nodes[#nodes][2] = 'an'
    end
end

-- ET/EN -- W5 + W7->1 / W6->on
if d == 'et' then
    first_et = first_et or (#nodes + 1)
elseif d == 'en' then
    has_en = true
    first_et = first_et or (#nodes + 1)
elseif first_et then -- d may be nil here!
    if has_en then
        if last == 'l' then
            temp = 'l' -- W7
        else
            temp = 'en' -- W5
        end
    else
        temp = 'on' -- W6
    end
end
for e = first_et, #nodes do
    if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
end
first_et = nil
has_en = false
end

if d then
    if d == 'al' then
        d = 'r'
        last = 'al'
    elseif d == 'l' or d == 'r' then
        last = d
    end
prev_d = d
    table.insert(nodes, {item, d, outer_first})
end
outer_first = nil
end
-- TODO -- repeated here in case EN/ET is the last node. Find a
-- better way of doing things:
if first_et then       -- dir may be nil here!
    if has_en then
        if last == 'l' then
            temp = 'l'   -- W7
        else
            temp = 'en'  -- W5
        end
    else
        temp = 'on'   -- W6
    end
for e = first_et, #nodes do
    if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
end
end

-- dummy node, to close things
table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})

------------  NEUTRAL ---------------
outer = save_outer
last = outer

local first_on = nil
for q = 1, #nodes do
    local item
local outer_first = nodes[q][3]
    outer = outer_first or outer
last = outer_first or last
local d = nodes[q][2]
    if d == 'an' or d == 'en' then d = 'r' end
    if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end  --- W6
    if d == 'on' then
        first_on = first_on or q
    elseif first_on then
        if last == d then
            temp = d
        else
            temp = outer
        end
for r = first_on, q - 1 do
    nodes[r][2] = temp
item = nodes[r][1]    -- MIRRORING
    if item.id == GLYPH and temp == 'r' then
        item.char = characters[item.char].m or item.char
    end
end
first_on = nil

end

if d == 'r' or d == 'l' then last = d end
end

163
-- IMPLICIT, REORDER --

```
outer = save_outer
last = outer

local state = {}
state.has_r = false

for q = 1, #nodes do
    local item = nodes[q][1]
    outer = nodes[q][3] or outer
    local d = nodes[q][2]
    if d == 'nsm' then d = last end -- W1
    if d == 'en' then d = 'an' end
    local isdir = (d == 'r' or d == 'l')
    if outer == 'l' and d == 'an' then
        state.sim = state.sim or item
        state.eim = item
        elseif state.sim and state.has_r then
            head, state = insert_implicit(head, state, outer)
    elseif d == 'l' then
        state.sim, state.eim, state.has_r = nil, nil, false
    end
else
    if d == 'an' or d == 'l' then
        if nodes[q][3] then -- nil except after an explicit dir
            state.sim = item -- so we move sim 'inside' the group
        else
            state.sim = state.sim or item
        end
    end
else
    elseif d == 'r' and state.sim then
        head, state = insert_implicit(head, state, outer)
    elseif d == 'r' then
        state.sim, state.eim = nil, nil
end

if isdir then
    last = d -- Don't search back - best save now
else d == 'on' and state.sim then
    state.sim = state.sim or item
    state.eim = item
end
```
15 Data for CJK

It is a boring file and it's not shown here. See the generated file.

16 The 'nil' language

This 'language' does nothing, except setting the hyphenation patterns to nohyphenation. For this language currently no special definitions are needed or available. The macro \LdfInit takes care of preventing that this file is loaded more than once, checking the category code of the @ sign, etc.

\ProvidesLanguage{nil}{\langle⟨date⟩⟩\langle⟨version⟩⟩}{Nil language}

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.

\ifdef\@nohyphenation\@undefined
\@nopatterns{nil}
\adddialect{\@nil0
\else
\let\@nil\@nohyphenation
\fi

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

\providehyphenmins{\CurrentOption}{\@ne\@ne}

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

\captionnil
\datenil
\let\captionsnil@empty
\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.

\ldf@finish{nil}
\end{nil}

17 Support for Plain \TeX\ (plain.def)

17.1 Not renaming hyphen.tex

As Don Knuth has declared that the filename hyphen.tex may only be used to designate his version of the american English hyphenation patterns, a new solution has to be found in order to be able to load hyphenation patterns for other languages in a plain-based \TeX-format. When asked he responded:
That file name is “sacred”, and if anybody changes it they will cause severe upward/downward compatibility headaches.

People can have a file locally by changing the file name or replacing plain.tex (or plain.tex except to preload additional fonts).

The files bplain.tex and blplain.tex can be used as replacement wrappers around plain.tex and lplain.tex to achieve the desired effect, based on the babel package. If you load each of them with \texttt{\LaTeX}, you will get a file called either bplain.fmt or blplain.fmt, which you can use as replacements for plain.fmt and lplain.fmt. As these files are going to be read as the first thing \texttt{\LaTeX} sees, we need to set some category codes just to be able to change the definition of \texttt{\input}.

\begin{verbatim}
\catcode`\{=1 % left brace is begin-group character
\catcode`\}=2 % right brace is end-group character
\catcode`\#=6 % hash mark is macro parameter character
\end{verbatim}

Now let's see if a file called hyphen.cfg can be found somewhere on \texttt{\LaTeX}'s input path by trying to open it for reading...

\begin{verbatim}
\ifempty{0}{hyphen.cfg}
\else
\afalse
\fi
\end{verbatim}

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 \texttt{\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).

\begin{verbatim}
\let\a\input
\end{verbatim}

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

\begin{verbatim}
\def\input #1 {%
\let\input\a
\a \hyphen.cfg
\}
\fi
\end{verbatim}

Once that's done the original meaning of \texttt{\input} can be restored and the definition of \texttt{\a} can be forgotten.

\begin{verbatim}
\let\a\undefined
\fi
\end{verbatim}

Now that we have made sure that hyphen.cfg will be loaded at the right moment it is time to load plain.tex.

\begin{verbatim}
(bplain)\a plain.tex
(bplain)\a lplain.tex
\end{verbatim}

Finally we change the contents of \texttt{\fmtname} to indicate that this is not the plain format, but a format based on plain with the babel package preloaded.

\begin{verbatim}
(bplain)\def\fmtname{babel-plain}
(bplain)\def\fmtname{babel-lplain}
\end{verbatim}

When you are using a different format, based on plain.tex you can make a copy of blplain.tex, rename it and replace plain.tex with the name of your format file.
17.2 Emulating some \LaTeX{} features

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

\begin{verbatim}
\def\empty{}
\def\loadlocalcfg#1{\openin0=#1.cfg
\ifeof0
\closein0
\else
\closein0
{\immediate\write16{******************}}
\immediate\write16{* Local config file #1.cfg used}*
\immediate\write16{*}
}
\input #1.cfg\relax
\endinput}
\end{verbatim}

17.3 General tools

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

\begin{verbatim}
\long\def\firstofone#1{#1}
\long\def\firstoftwo#1#2{#1}
\def\nil{\nil}
\def\gobbletwo#1#2{}
\def\ifstar#1{%\ifnextchar *{\firstoftwo{#1}}}
\def\starorlong#1{%}
\let\@ngrelx\relax
\def\gobble#1{}
\edef\backslashchar{\expandafter\gobble\string\}
\def\strip@prefix#1{}
\def\gobble@macro#1{%}
\def\toks@\expandafter{#1#2}%
\def\#1{\the\toks@}
\def\namedef#1{\expandafter\def\csname #1\endcsname}
\def\nameuse#1{\csname #1\endcsname}
\def\ifdefundefined#1{%}
\expandafter\ifx\csname #1\endcsname\relax
\expandafter\firstoftwo
\else
\expandafter\secondoftwo
\fi
\def\expandafter\secondoftwo
\def\expandafter\reserved@a{\noexpand#1{#2}{#3}}\reserved@a
\def\zap@space#1 #2{%
1%
\ifx#2\empty\else\expandafter\zap@space@fi
\expandafter\zapp@space@fi
\end{verbatim}

167
\TeX\ has the command \texttt{\@onlypreamble} which adds commands to a list of commands that are no longer needed after \texttt{\begin{document}}.

Mimick \TeX\'s \texttt{\AtBeginDocument}; for this to work the user needs to add \texttt{\begindocument} to his file.

We also have to mimick \TeX\'s \texttt{\AtEndOfPackage}. Our replacement macro is much simpler; it stores its argument in \texttt{\@endofldf}.

\TeX\ needs to be able to switch off writing to its auxiliary files; plain doesn't have them by default.

Mimick \TeX\'s commands to define control sequences.
The following little macro \in@ is taken from \texttt{latex.ltx}; it checks whether its first argument is part of its second argument. It uses the boolean \texttt{\in@}; allocating a new boolean inside conditionally executed code is not possible, hence the construct with the temporary definition of \texttt{\bb@tempa}.

\begin{verbatim}
\def\bb@tempa\ifx\csname newif\endcsname\ifin@\fi
\let\bb@tempa@empty\empty
\fi
\end{verbatim}
\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).

\begin{verbatim}
\def@ifpackagewith#1#2#3#4{#3}
The \LaTeX macro \@ifadded 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.
\end{verbatim}

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 \LaTeX2ε versions; just enough to make things work in plain \TeX environments.

\begin{verbatim}
\ife\@tempcnta\undefined
\csname newcount@endcsname@\@tempcnta\relax
\fi
\ife\@tempcntb\undefined
\csname newcount@endcsname@\@tempcntb\relax
\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).
\end{verbatim}

\begin{verbatim}
\ife\bye\undefined
\advance\count10 by -2\relax
\fi
\ife\@ifnextchar\undefined
\def\@ifnextchar#1#2{\}
\let\reserved@d=#1
\def\reserved@a{#2}\def\reserved@b{#3}%
\futurelet@let@token@@ifnch}
\def\@ifnch{\}
\ifx@@let@token@\@sptoken
\let\reserved@c\@xifnch
\else
\ifx@@let@token@\@reserved@d
\let\reserved@c\@reserved@a
\else
\let\reserved@c\@reserved@b
\fi
\fi
\fi
\def\@reserved@c{}
\def\@reserved@b{\@sptoken=}
% this makes \@sptoken a space token
\def\@reserved@c{\@xifnch} \expandafter\def\@reserved@b{\@futurelet@let@token@@ifnch}
\fi
\ife
\def\@testopt#1#2{\}
\if\@ifnextchar{{#1}{{#1[#2]}}}
\def\@protected@testopt#1{\}
\expandafter\@testopt
\else
\@x@protect#1
\fi
\long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@whilenum#1\relax #2\relax}\fi
\end{verbatim}
17.4 Encoding related macros

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

```tex
\long\def@iwhilenum#1\{\ifnum #1\expandafter@iwhilenum
\else\expandafter@gobble\fi[#1]\}
```

17.4 Encoding related macros

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

```tex
\def\DeclareTextCommand{% 
@dec@text@cmd\providecommand
}\def\ProvideTextCommand{% 
@dec@text@cmd\providecommand
}\def\DeclareTextSymbol#1#2#3{% 
@dec@text@cmd#1#2#3\relax
}\def\@dec@text@cmd#1#2#3{% 
\expandafter\def\expandafter#2%
\expandafter{%
\csname#3-cmd\expandafter\endcsname
\expandafter#2%
\csname#3\string#2\endcsname
}%
\let@ifdefinable@rc@ifdefinable
\expandafter#1\csname#3\string#2\endcsname
}\def\@current@cmd#1{% 
\ifx\protect@typeset\protect\else
 \noexpand\expandafter@gobble
 \fi
}\def\@changed@cmd#1#2{% 
\ifx\protect@typeset\protect
 \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
 \expandafter\ifx\csname ?\string#1\endcsname\relax
 \expandafter\ifx\csname ?\string#1\endcsname%
 \@changed@x@err{#1}%
}%
\fi
 \global\expandafter@let
 \csname\cf@encoding \string#1\expandafter\endcsname
 \csname ?\string#1\endcsname
 \fi
 \csname\cf@encoding \string#1%
 \expandafter\endcsname
 \else
 \noexpand#1%
 \fi
}\def\@changed@x@err#1{% 
\errhelp{Your command will be ignored, type \texttt{<return>} to proceed}%
\errmessage{Command \protect#1 undefined in encoding \texttt{\cf@encoding}}
}\def\DeclareTextCommandDefault#1{% 
\DeclareTextCommand#1%
}\def\ProvideTextCommandDefault#1{% 
\ProvideTextCommand#1%
}\expandafter@let\csname OT1-cmd\endcsname@current@cmd
```
4994 \expandafter\let\csname?-cmd\endcsname@changed@cmd
4995 \def\DeclareTextAccent#1#2#3{%
4996 \DeclareTextCommand#1{[#2]{[1]}{#3 ##1} #1}
4997 }
4998 \def\DeclareTextCompositeCommand#1#2#3#4{%
4999 \expandafter\let\reserved@a\csname#2\string#1\endcsname
5000 \edef\reserved@b{\{#1}\%}
5001 \edef\reserved@c{%
5002 \expandafter\strip@args\meaning\reserved@a:\@\strip@args%
5003 \ifx\reserved@b\reserved@c
5004 \expandafter\expandafter\expandafter\ifx
5005 \expandafter\@car\reserved@a\relax\relax@nil
5006 \@text@composite
5007 \else
5008 \edef\reserved@b##1{%
5009 \def\expandafter\noexpand\%
5010 \csname#2\string#1\endcsname###1{%
5011 \noexpand@text@composite
5012 \expandafter\noexpand\csname#2\string#1\endcsname
5013 ###1\noexpand\@empty\noexpand@text@composite
5014 {##1}%
5015 }%
5016 }%
5017 \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
5018 \fi
5019 \expandafter\csname\expandafter\string\csname
5020 #2\endcsname\string#1-\string#2\endcsname{#4}
5021 \else
5022 \errhelp{Your command will be ignored, type <return> to proceed}%
5023 \errmessage{string\DeclareTextCompositeCommand\space used on
5024 inappropriate command \protect#1}%
5025 \fi
5026 }
5027 \def\@text@composite#1#2#3@text@composite{%
5028 \expandafter\@text@composite@x
5029 \csname#2\string#1\endcsname
5030 }
5031 \def\@text@composite@x#1#2{%
5032 \ifx#1\relax
5033 #2%
5034 \else
5035 #1%
5036 \fi
5037 }
5038 %
5039 \def\@strip@args#1:#2-#3@strip@args{#2}
5040 \def\DeclareTextCompositeCommand#1#2#3#4{%
5041 \def\reserved@a{\\csname#2\string#1{#2}{#3}}%
5042 \bgroup
5043 \lccode`@=#4%
5044 \lowercase%
5045 \egroup
5046 \reserved@a @%
5047 }%
5048 }
5049 %
5050 \def\UseTextSymbol#1#2{%
5051 \let\curr@enc\cf@encoding
5052 % \use@text@encoding{#1}%
Currently we only use the \TeX2ε method for accents for those that are known to be made active in some language definition file.

The following control sequences are used in babel.def but are not defined for plain \TeX.

For a couple of languages we need the \TeX-control sequence \texttt{\scriptsize} to be available. Because plain \TeX doesn't have such a sophisticated font mechanism as \TeX has, we just let it to \texttt{\sevenrm}.

18 Acknowledgements

I would like to thank all who volunteered as β-testers for their time. Michel Goossens supplied contributions for most of the other languages. Nico Poppelier helped polish the text of the documentation and supplied parts of the macros for the Dutch language. Paul Wackers and Werenfried Spit helped find and repair bugs.

During the further development of the babel system I received much help from Bernd Raichle, for which I am grateful.
References


