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

This manual describes babel, a package that makes use of the capabilities of \TeX{} version 3 and, to some extent, \texttt{xetex} and \texttt{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 (\texttt{Xe\TeX} and \texttt{Lua\TeX}) and the so-called \textit{complex scripts}. New features related to font selection, bidi writing and the like will be 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 \texttt{xelatex} and \texttt{lualatex} 6
   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 9
   1.10 Shorthands 11
   1.11 Package options 14
   1.12 The base option 16
   1.13 ini files 17
   1.14 Selecting fonts 23
   1.15 Modifying a language 24
   1.16 Creating a language 25
   1.17 Digits 26
   1.18 Getting the current language name 27
   1.19 Hyphenation tools 27
   1.20 Selecting scripts 29
   1.21 Selecting directions 29
   1.22 Language attributes 32
   1.23 Hooks 32
   1.24 Languages supported by babel 33
   1.25 Tips, workarounds, know issues and notes 34
   1.26 Current and future work 36
   1.27 Tentative and experimental code 37

2 Loading languages with \texttt{language.dat} 38
   2.1 Format 38

3 The interface between the core of babel and the language definition files 39
   3.1 Guidelines for contributed languages 40
   3.2 Basic macros 41
   3.3 Skeleton 42
   3.4 Support for active characters 43
   3.5 Support for saving macro definitions 43
   3.6 Support for extending macros 43
   3.7 Macros common to a number of languages 44
   3.8 Encoding-dependent strings 44

4 Changes 48
   4.1 Changes in babel version 3.9 48
   4.2 Changes in babel version 3.7 48

II The code 49

5 Identification and loading of required files 49
14 The 'nil’ language

15 Support for Plain \TeX\ (plain.def)
  15.1 Not renaming hyphen.tex ..................................... 153
  15.2 Emulating some \LaTeX\ features ............................... 155
  15.3 General tools .................................................... 155
  15.4 Encoding related macros ....................................... 159

16 Acknowledgements

Troubleshooting
  Paragraph ended before \UTFviii@three@octets was complete .......... 4
  No hyphenation patterns were preloaded for (babel) the language ‘LANG’ into the
  format ................................................................. 5
  You are loading directly a language style .......................... 7
  Unknown language ‘LANG’ ............................................ 7
  Argument of \language@active@arg” has an extra } ................. 11
Part I

User guide

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

• Changes and new features with relation to version 3.8 are highlighted with \texttt{New XXX}. The most recent features could be still unstable. Please, report any issues you find.

• 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/latex2e/tree/master/required/babel (which provides some sample files, too).

• See section \ref{sec:lang} for contributing a language.

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.

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

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

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

\begin{verbatim}
! Paragraph ended before \UTFviii@three@octets was complete.
\end{verbatim}

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:
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{}, Mik\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}
```

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

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

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

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):  

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

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

1.4 xelatex and lualatex

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

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

\begin{itemize}
  \item \texttt{\LaTeX} \begin{itemize}
    \item \texttt{\LaTeX} \geq 2018-04-01 if the encoding is UTF-8.
  \end{itemize}
\end{itemize}
EXAMPLE The following bilingual, single script document in UTF-8 encoding just prints a couple of ‘captions’ and \today in Danish and Vietnamese. No additional packages are required.

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

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

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

1.5 Troubleshooting

• Loading directly sty files in \LaTeX (i.e., \usepackage{(language)}) 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{quote}
\texttt{! Package babel Error: You are loading directly a language style. (babel) This syntax is deprecated and you must use (babel) \usepackage[language]{babel}.}
\end{quote}

• Another typical error when using babel is the following:\footnote{In old versions the error read “You haven’t loaded the language LANG yet.”}
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):

\begin{verbatim}
\input estonian.sty
\begindocument
\end{verbatim}

**WARNING** Not all languages provide a sty file and some of them are not compatible with Plain.\(^4\)

### 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 $\texttt{selectlanguage}$ and $\texttt{foreignlanguage}$ are necessary. The environments $\texttt{otherlanguage}$, $\texttt{otherlanguage*}$ and $\texttt{hyphenrules}$ are auxiliary, and described in the next section. The main language is selected automatically when the document environment begins.

\begin{itemize}
  \item $\texttt{selectlanguage}$
\end{itemize}

When a user wants to switch from one language to another he can do so using the macro $\texttt{selectlanguage}$. 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:

\begin{verbatim}
\selectlanguage{german}
\end{verbatim}

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, $\texttt{selectlanguage(\german)}$ is equivalent to $\texttt{selectlanguage(\german)}$. 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:

\begin{verbatim}
{\selectlanguage{<inner-language>} \ldots}\selectlanguage{<outer-language>}
\end{verbatim}

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

\(^4\) Even in the babel kernel there were some macros not compatible with plain. Hopefully these issues will be fixed soon.
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).

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

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.

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

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

To set hyphenation exceptions, use \babelhyphenation (see below).

1.9 More on selection
\babeltags \{(tag1) = \langle language1\rangle, (tag2) = \langle language2\rangle, \ldots\}

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

**EXAMPLE** With

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

you can write

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

and

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

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

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

\babelensure \{include=\langle commands\rangle, exclude=\langle commands\rangle, fontenc=\langle encoding\rangle\}\{\langle language\rangle\}

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

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

Of course, \LaTeX{} 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:

---

\footnote{With it encoded string may not work as expected.}

10
They are activated when the language is selected (at the afterextras event), and it makes some assumptions which could not be fulfilled in some languages. Note also you should include only macros defined by the language, not global macros (eg, \TeX 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 discretionary 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 pdfTeX provides \knbccode, 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, ".").

\shorthandoff \langle\{\textit{shorthands-list}\}\rangle

\shorthandon \langle\{\textit{shorthands-list}\}\rangle

It is sometimes necessary to switch a shorthand character off temporarily, because it must be used in an entirely different way. For this purpose, the user commands \shorthandoff and \shorthandon are provided. They each take a list of characters as their arguments. The command \shorthandoff sets the \catcode for each of the characters in its argument to other (12); the command \shorthandon sets the \catcode to active (13). Both commands only work on 'known' shorthand characters. If a character is not known to be a shorthand character its category code will be left unchanged.
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

\useshorthands*{⟨char⟩}

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

\useshorthands

\useshorthands*{⟨char⟩}

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

\defineshorthand*[⟨language⟩,⟨language⟩,...]{⟨shorthand⟩}{⟨code⟩}

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

\defineshorthand

\defineshorthand*[⟨language⟩,⟨language⟩,...]{⟨shorthand⟩}{⟨code⟩}

An optional argument allows to (re)define language and system shorthands (some languages do not activate shorthands, so you may want to add \language{⟨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.

\defineshorthand

\defineshorthand*[⟨language⟩,⟨language⟩,...]{⟨shorthand⟩}{⟨code⟩}

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:

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

\defineshorthand*{"} \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:

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

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

Now, you have a single unified shorthand ("-), with a content-based meaning (“compound word hyphen”) whose visual behavior is that expected in each context.
The command \aliasshorthand can be used to let another character perform the same functions as the default shorthand character. If one prefers for example to use the character / over " in typing Polish texts, this can be achieved by entering \aliasshorthand{"}{/}.

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

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

\aliasshorthand{-}{^}
\AtBeginDocument{\shorthandoff*{-}}

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

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

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

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

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

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

For your records, here is a list of shorthands, but you must double check them, as they may change\(^7\)

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

---

\(^6\) 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.

\(^7\) Thanks to Enrico Gregorio
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.

1.11 Package options

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

KeepShorthandsActive Tells babel not to deactivate shorthands after loading a language file, so that they are also available in the preamble.

activeacute For some languages babel supports this options to set ' as a shorthand in case it is not done by default.

activegrave Same for `.

shorthands= ⟨char⟩⟨char⟩... | off
The only language shorthands activated are those given, like, eg:

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

If ' is included, activeacute is set; if ' is included, activegrave is set. Active characters (like ~) should be preceded by \string (otherwise they will be expanded by \LaTeXX 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

\footnote{This declaration serves to nothing, but it is preserved for backward compatibility.}

14
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 spoiled by an unexpected .cfg file. However, if the key config is set, this file is loaded.

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

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

silent [New 3.9l] No warnings and no infos are written to the log file.

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

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

*You can use alternatively the package silence.
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 other language;
- **other** also sets it at other language* as well as in heads and feet (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= New 3.14 Selects the bidi algorithm to be used in \texttt{luatex} and \texttt{xetex}. See sec. 1.21.

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

1.12 The base option

With this package option babel just loads some basic macros (those in switch.def), defines \AfterBabelLanguage and exits. It also selects the hyphenation patterns for the 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, some reason, incompatible languages. It can be used if you just want to select the hyphenation patterns of a single language, too.

\AfterBabelLanguage \{(option-name)\}{\{code\}}

This command is currently the only provided by base. Executes \{code\} when the file loaded by the corresponding package option is finished (at \ldf@finish). The setting is global. So does ... at the end of french.ldf. It can be used in ldf files, too, but in such a case the code is executed only if \{option-name\} is the same as \CurrentOption (which could not be the same as the option name as set in \usepackage!).

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

\usepackage[base]{babel}
\AfterBabelLanguage{foo}{{\%
\let\macroFoo\macro}}

---

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

**EXAMPLE** Although Georgian has its own ldf file, here is how to declare this language with an ini file in Unicode engines. The \texttt{nil} language is required, because currently babel raises an error if there is no language.

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

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

<table>
<thead>
<tr>
<th>Language Code</th>
<th>Language Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>af</td>
<td>Afrikaans</td>
</tr>
<tr>
<td>agq</td>
<td>Aghem</td>
</tr>
<tr>
<td>ak</td>
<td>Akan</td>
</tr>
<tr>
<td>am</td>
<td>Amharic</td>
</tr>
<tr>
<td>ar</td>
<td>Arabic</td>
</tr>
<tr>
<td>as</td>
<td>Assamese</td>
</tr>
<tr>
<td>asa</td>
<td>Asu</td>
</tr>
<tr>
<td>ast</td>
<td>Asturian</td>
</tr>
<tr>
<td>az-Cyril</td>
<td>Azerbaijani</td>
</tr>
<tr>
<td>az-Latin</td>
<td>Azerbaijani</td>
</tr>
<tr>
<td>az</td>
<td>Azerbaijani</td>
</tr>
<tr>
<td>bas</td>
<td>Basaa</td>
</tr>
<tr>
<td>be</td>
<td>Belarusian</td>
</tr>
<tr>
<td>bem</td>
<td>Bemba</td>
</tr>
<tr>
<td>bez</td>
<td>Bena</td>
</tr>
<tr>
<td>bg</td>
<td>Bulgarian</td>
</tr>
<tr>
<td>bm</td>
<td>Bambara</td>
</tr>
<tr>
<td>bn</td>
<td>Bangla</td>
</tr>
<tr>
<td>bo</td>
<td>Tibetan</td>
</tr>
<tr>
<td>brx</td>
<td>Bodo</td>
</tr>
<tr>
<td>bs-Cyril</td>
<td>Bosnian</td>
</tr>
<tr>
<td>bs-Latin</td>
<td>Bosnian</td>
</tr>
<tr>
<td>bs</td>
<td>Bosnian</td>
</tr>
<tr>
<td>ca</td>
<td>Catalan</td>
</tr>
<tr>
<td>ce</td>
<td>Chechen</td>
</tr>
<tr>
<td>cgg</td>
<td>Chiga</td>
</tr>
<tr>
<td>chr</td>
<td>Cherokee</td>
</tr>
<tr>
<td>ckb</td>
<td>Central Kurdish</td>
</tr>
<tr>
<td>cs</td>
<td>Czech</td>
</tr>
<tr>
<td>cy</td>
<td>Welsh</td>
</tr>
<tr>
<td>da</td>
<td>Danish</td>
</tr>
<tr>
<td>dav</td>
<td>Taita</td>
</tr>
<tr>
<td>Code</td>
<td>Language</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------</td>
</tr>
<tr>
<td>de-AT</td>
<td>German</td>
</tr>
<tr>
<td>de-CH</td>
<td>German</td>
</tr>
<tr>
<td>de</td>
<td>German</td>
</tr>
<tr>
<td>dje</td>
<td>Zarma</td>
</tr>
<tr>
<td>dsb</td>
<td>Lower Sorbian</td>
</tr>
<tr>
<td>dua</td>
<td>Duala</td>
</tr>
<tr>
<td>dyo</td>
<td>Jola-Fonyi</td>
</tr>
<tr>
<td>dz</td>
<td>Dzongkha</td>
</tr>
<tr>
<td>ebu</td>
<td>Embu</td>
</tr>
<tr>
<td>ee</td>
<td>Ewe</td>
</tr>
<tr>
<td>el</td>
<td>Greek</td>
</tr>
<tr>
<td>en-AU</td>
<td>English</td>
</tr>
<tr>
<td>en-CA</td>
<td>English</td>
</tr>
<tr>
<td>en-GB</td>
<td>English</td>
</tr>
<tr>
<td>en-NZ</td>
<td>English</td>
</tr>
<tr>
<td>en-US</td>
<td>English</td>
</tr>
<tr>
<td>en</td>
<td>English</td>
</tr>
<tr>
<td>eo</td>
<td>Esperanto</td>
</tr>
<tr>
<td>es-MX</td>
<td>Spanish</td>
</tr>
<tr>
<td>es</td>
<td>Spanish</td>
</tr>
<tr>
<td>et</td>
<td>Estonian</td>
</tr>
<tr>
<td>eu</td>
<td>Basque</td>
</tr>
<tr>
<td>ewo</td>
<td>Ewondo</td>
</tr>
<tr>
<td>fa</td>
<td>Persian</td>
</tr>
<tr>
<td>ff</td>
<td>Fulah</td>
</tr>
<tr>
<td>fi</td>
<td>Finnish</td>
</tr>
<tr>
<td>fil</td>
<td>Filipino</td>
</tr>
<tr>
<td>fo</td>
<td>Faroese</td>
</tr>
<tr>
<td>fr</td>
<td>French</td>
</tr>
<tr>
<td>fr-BE</td>
<td>French</td>
</tr>
<tr>
<td>fr-CA</td>
<td>French</td>
</tr>
<tr>
<td>fr-CH</td>
<td>French</td>
</tr>
<tr>
<td>fr-LU</td>
<td>French</td>
</tr>
<tr>
<td>fur</td>
<td>Friulian</td>
</tr>
<tr>
<td>fy</td>
<td>Western Frisian</td>
</tr>
<tr>
<td>ga</td>
<td>Irish</td>
</tr>
<tr>
<td>gd</td>
<td>Scottish Gaelic</td>
</tr>
<tr>
<td>gl</td>
<td>Galician</td>
</tr>
<tr>
<td>gsw</td>
<td>Swiss German</td>
</tr>
<tr>
<td>gu</td>
<td>Gujarati</td>
</tr>
<tr>
<td>guz</td>
<td>Gusii</td>
</tr>
<tr>
<td>gv</td>
<td>Manx</td>
</tr>
<tr>
<td>ha-GH</td>
<td>Hausa</td>
</tr>
<tr>
<td>ha-NE</td>
<td>Hausa</td>
</tr>
<tr>
<td>ha</td>
<td>Hausa</td>
</tr>
<tr>
<td>haw</td>
<td>Hawaiian</td>
</tr>
<tr>
<td>he</td>
<td>Hebrew</td>
</tr>
<tr>
<td>hi</td>
<td>Hindi</td>
</tr>
<tr>
<td>hr</td>
<td>Croatian</td>
</tr>
<tr>
<td>hsb</td>
<td>Upper Sorbian</td>
</tr>
<tr>
<td>hu</td>
<td>Hungarian</td>
</tr>
<tr>
<td>hy</td>
<td>Armenian</td>
</tr>
<tr>
<td>ia</td>
<td>Interlingua</td>
</tr>
<tr>
<td>id</td>
<td>Indonesian</td>
</tr>
<tr>
<td>Code</td>
<td>Name</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
</tr>
<tr>
<td>mua</td>
<td>Mundang</td>
</tr>
<tr>
<td>my</td>
<td>Burmese</td>
</tr>
<tr>
<td>mzn</td>
<td>Mazanderani</td>
</tr>
<tr>
<td>naq</td>
<td>Nama</td>
</tr>
<tr>
<td>nb</td>
<td>Norwegian Bokmål</td>
</tr>
<tr>
<td>nd</td>
<td>North Ndebele</td>
</tr>
<tr>
<td>ne</td>
<td>Nepali</td>
</tr>
<tr>
<td>nl</td>
<td>Dutch</td>
</tr>
<tr>
<td>nmg</td>
<td>Kwasio</td>
</tr>
<tr>
<td>nn</td>
<td>Norwegian Nynorsk</td>
</tr>
<tr>
<td>nnh</td>
<td>Ngiemboon</td>
</tr>
<tr>
<td>nus</td>
<td>Nuer</td>
</tr>
<tr>
<td>nyn</td>
<td>Nyankole</td>
</tr>
<tr>
<td>om</td>
<td>Oromo</td>
</tr>
<tr>
<td>or</td>
<td>Odia</td>
</tr>
<tr>
<td>os</td>
<td>Ossetic</td>
</tr>
<tr>
<td>pa-Arab</td>
<td>Punjabi</td>
</tr>
<tr>
<td>pa</td>
<td>Punjabi</td>
</tr>
<tr>
<td>pl</td>
<td>Polish</td>
</tr>
<tr>
<td>pms</td>
<td>Piedmontese</td>
</tr>
<tr>
<td>ps</td>
<td>Pashto</td>
</tr>
<tr>
<td>pt-BR</td>
<td>Portuguese</td>
</tr>
<tr>
<td>pt-PT</td>
<td>Portuguese</td>
</tr>
<tr>
<td>pt</td>
<td>Portuguese</td>
</tr>
<tr>
<td>qu</td>
<td>Quechua</td>
</tr>
<tr>
<td>rm</td>
<td>Romansh</td>
</tr>
<tr>
<td>rn</td>
<td>Rundi</td>
</tr>
<tr>
<td>ro</td>
<td>Romanian</td>
</tr>
<tr>
<td>rof</td>
<td>Rombo</td>
</tr>
<tr>
<td>ru</td>
<td>Russian</td>
</tr>
<tr>
<td>rw</td>
<td>Kinyarwanda</td>
</tr>
<tr>
<td>rwk</td>
<td>Rwa</td>
</tr>
<tr>
<td>sah</td>
<td>Sakha</td>
</tr>
<tr>
<td>saq</td>
<td>Samburu</td>
</tr>
<tr>
<td>sbp</td>
<td>Sangu</td>
</tr>
<tr>
<td>se</td>
<td>Northern Sami</td>
</tr>
<tr>
<td>seh</td>
<td>Sena</td>
</tr>
<tr>
<td>ses</td>
<td>Koyraboro Senni</td>
</tr>
<tr>
<td>sg</td>
<td>Sango</td>
</tr>
<tr>
<td>shi-Ltn</td>
<td>Tachelhit</td>
</tr>
<tr>
<td>shi-Tfng</td>
<td>Tachelhit</td>
</tr>
<tr>
<td>shi</td>
<td>Tachelhit</td>
</tr>
<tr>
<td>si</td>
<td>Sinhala</td>
</tr>
<tr>
<td>sk</td>
<td>Slovenian</td>
</tr>
<tr>
<td>sl</td>
<td>Slovenian</td>
</tr>
<tr>
<td>smn</td>
<td>Inari Sami</td>
</tr>
<tr>
<td>sn</td>
<td>Shona</td>
</tr>
<tr>
<td>so</td>
<td>Somali</td>
</tr>
<tr>
<td>sq</td>
<td>Albanian</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 ldif file).

<table>
<thead>
<tr>
<th>Language</th>
<th>Script</th>
</tr>
</thead>
<tbody>
<tr>
<td>aghem</td>
<td>chinese-hant</td>
</tr>
<tr>
<td>akan</td>
<td>chinese-simplified-hongkongsarchina</td>
</tr>
<tr>
<td>albanian</td>
<td>chinese-simplified-macausarchina</td>
</tr>
<tr>
<td>american</td>
<td>chinese-simplified-singapore</td>
</tr>
<tr>
<td>amharic</td>
<td>chinese-simplified</td>
</tr>
<tr>
<td>arabic</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>australian</td>
<td>cognonian</td>
</tr>
<tr>
<td>austrian</td>
<td>cornish</td>
</tr>
<tr>
<td>azerbajani-cyrillic</td>
<td>czech</td>
</tr>
<tr>
<td>azerbajani-cyrl</td>
<td>danish</td>
</tr>
<tr>
<td>azerbajani-latin</td>
<td>duala</td>
</tr>
<tr>
<td>azerbajani-latn</td>
<td>dutch</td>
</tr>
<tr>
<td>azerbajani</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>belarussian</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>
<tr>
<td>chinese-hans-hk</td>
<td>french</td>
</tr>
<tr>
<td>chinese-hans-mo</td>
<td>friulian</td>
</tr>
<tr>
<td>chinese-hans-sg</td>
<td>fulah</td>
</tr>
<tr>
<td>chinese-hans</td>
<td>galician</td>
</tr>
<tr>
<td>chinese-hant-hk</td>
<td>ganda</td>
</tr>
<tr>
<td>chinese-hant-mo</td>
<td>georgian</td>
</tr>
<tr>
<td>Language</td>
<td>Region</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>german-at</td>
<td></td>
</tr>
<tr>
<td>german-austria</td>
<td></td>
</tr>
<tr>
<td>german-ch</td>
<td></td>
</tr>
<tr>
<td>german-switzerland</td>
<td></td>
</tr>
<tr>
<td>german</td>
<td></td>
</tr>
<tr>
<td>greek</td>
<td></td>
</tr>
<tr>
<td>gujarati</td>
<td></td>
</tr>
<tr>
<td>gusii</td>
<td></td>
</tr>
<tr>
<td>Hausa-Gh</td>
<td></td>
</tr>
<tr>
<td>Hausa-Ghana</td>
<td></td>
</tr>
<tr>
<td>Hausa-Ne</td>
<td></td>
</tr>
<tr>
<td>Hausa-Niger</td>
<td></td>
</tr>
<tr>
<td>Hausa</td>
<td></td>
</tr>
<tr>
<td>Hawaiian</td>
<td></td>
</tr>
<tr>
<td>Hebrew</td>
<td></td>
</tr>
<tr>
<td>Hindi</td>
<td></td>
</tr>
<tr>
<td>Hungarian</td>
<td></td>
</tr>
<tr>
<td>Icelandic</td>
<td></td>
</tr>
<tr>
<td>Igbo</td>
<td></td>
</tr>
<tr>
<td>Inarisi</td>
<td></td>
</tr>
<tr>
<td>Indonesian</td>
<td></td>
</tr>
<tr>
<td>Interlingua</td>
<td></td>
</tr>
<tr>
<td>Irish</td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td></td>
</tr>
<tr>
<td>Jolafonyi</td>
<td></td>
</tr>
<tr>
<td>Kabuverdianu</td>
<td></td>
</tr>
<tr>
<td>Kabyle</td>
<td></td>
</tr>
<tr>
<td>Kako</td>
<td></td>
</tr>
<tr>
<td>Kalaallisut</td>
<td></td>
</tr>
<tr>
<td>Kalenjin</td>
<td></td>
</tr>
<tr>
<td>Kamba</td>
<td></td>
</tr>
<tr>
<td>Kannada</td>
<td></td>
</tr>
<tr>
<td>Kashmiri</td>
<td></td>
</tr>
<tr>
<td>Kazakh</td>
<td></td>
</tr>
<tr>
<td>Khmer</td>
<td></td>
</tr>
<tr>
<td>Kikuyu</td>
<td></td>
</tr>
<tr>
<td>Kinyarwanda</td>
<td></td>
</tr>
<tr>
<td>Konkani</td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td></td>
</tr>
<tr>
<td>Koyraborosenni</td>
<td></td>
</tr>
<tr>
<td>Koyrachiini</td>
<td></td>
</tr>
<tr>
<td>Kwasio</td>
<td></td>
</tr>
<tr>
<td>Kyrgyz</td>
<td></td>
</tr>
<tr>
<td>Lakota</td>
<td></td>
</tr>
<tr>
<td>Langi</td>
<td></td>
</tr>
<tr>
<td>Lao</td>
<td></td>
</tr>
<tr>
<td>Latvian</td>
<td></td>
</tr>
<tr>
<td>Lingala</td>
<td></td>
</tr>
<tr>
<td>Lithuanian</td>
<td></td>
</tr>
<tr>
<td>Lowersorbian</td>
<td></td>
</tr>
<tr>
<td>Lsorbian</td>
<td></td>
</tr>
<tr>
<td>Lubakatanga</td>
<td></td>
</tr>
<tr>
<td>Luo</td>
<td></td>
</tr>
<tr>
<td>Luxembourghish</td>
<td></td>
</tr>
<tr>
<td>Luyia</td>
<td></td>
</tr>
<tr>
<td>Macedonian</td>
<td></td>
</tr>
<tr>
<td>Machame</td>
<td></td>
</tr>
<tr>
<td>Makhuwameetto</td>
<td></td>
</tr>
<tr>
<td>Makonde</td>
<td></td>
</tr>
<tr>
<td>Malagasy</td>
<td></td>
</tr>
<tr>
<td>Malay-Bn</td>
<td></td>
</tr>
<tr>
<td>Malay-Brunei</td>
<td></td>
</tr>
<tr>
<td>Malay-Sg</td>
<td></td>
</tr>
<tr>
<td>Malay-Singapore</td>
<td></td>
</tr>
<tr>
<td>Malay</td>
<td></td>
</tr>
<tr>
<td>Malayalam</td>
<td></td>
</tr>
<tr>
<td>Maltese</td>
<td></td>
</tr>
<tr>
<td>Manx</td>
<td></td>
</tr>
<tr>
<td>Marathi</td>
<td></td>
</tr>
<tr>
<td>Masai</td>
<td></td>
</tr>
<tr>
<td>Mazanderani</td>
<td></td>
</tr>
<tr>
<td>Meru</td>
<td></td>
</tr>
<tr>
<td>Meta</td>
<td></td>
</tr>
<tr>
<td>Mexican</td>
<td></td>
</tr>
<tr>
<td>Mongolian</td>
<td></td>
</tr>
<tr>
<td>Morisyen</td>
<td></td>
</tr>
<tr>
<td>Mundang</td>
<td></td>
</tr>
<tr>
<td>Nama</td>
<td></td>
</tr>
<tr>
<td>Nepali</td>
<td></td>
</tr>
<tr>
<td>Newzealand</td>
<td></td>
</tr>
<tr>
<td>Ngemboon</td>
<td></td>
</tr>
<tr>
<td>Ngomba</td>
<td></td>
</tr>
<tr>
<td>Norvwegian</td>
<td></td>
</tr>
<tr>
<td>Norskie</td>
<td></td>
</tr>
<tr>
<td>Northernlurian</td>
<td></td>
</tr>
<tr>
<td>Northernsamian</td>
<td></td>
</tr>
<tr>
<td>Northendebale</td>
<td></td>
</tr>
<tr>
<td>Norwegianbokmal</td>
<td></td>
</tr>
<tr>
<td>Norwegiannorskie</td>
<td></td>
</tr>
<tr>
<td>Nswissgerman</td>
<td></td>
</tr>
<tr>
<td>Nuer</td>
<td></td>
</tr>
<tr>
<td>Nyankole</td>
<td></td>
</tr>
<tr>
<td>Nynorsk</td>
<td></td>
</tr>
<tr>
<td>Occitan</td>
<td></td>
</tr>
<tr>
<td>Oriya</td>
<td></td>
</tr>
<tr>
<td>Oromo</td>
<td></td>
</tr>
<tr>
<td>Ossetic</td>
<td></td>
</tr>
<tr>
<td>Pashto</td>
<td></td>
</tr>
<tr>
<td>Persian</td>
<td></td>
</tr>
<tr>
<td>Piedmontese</td>
<td></td>
</tr>
<tr>
<td>Polish</td>
<td></td>
</tr>
<tr>
<td>Portuguese-Brazil</td>
<td></td>
</tr>
<tr>
<td>Portuguese-Portugal</td>
<td></td>
</tr>
<tr>
<td>Portuguese-Port</td>
<td></td>
</tr>
<tr>
<td>Portuguese</td>
<td></td>
</tr>
<tr>
<td>Punjabi-Arab</td>
<td></td>
</tr>
<tr>
<td>Punjabi-Arabic</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>Language</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>punjabi-gurmukhi</td>
<td>swedish</td>
</tr>
<tr>
<td>punjabi-guru</td>
<td>swissgerman</td>
</tr>
<tr>
<td>punjabi</td>
<td>tachelhit-latin</td>
</tr>
<tr>
<td>quechua</td>
<td>tachelhit-latn</td>
</tr>
<tr>
<td>romanian</td>
<td>tachelhit-tfng</td>
</tr>
<tr>
<td>romansh</td>
<td>tachelhit-tifinagh</td>
</tr>
<tr>
<td>rombo</td>
<td>tachelhit</td>
</tr>
<tr>
<td>rundi</td>
<td>taita</td>
</tr>
<tr>
<td>russian</td>
<td>tamil</td>
</tr>
<tr>
<td>rwa</td>
<td>tasawaq</td>
</tr>
<tr>
<td>sakha</td>
<td>telugu</td>
</tr>
<tr>
<td>samburu</td>
<td>teso</td>
</tr>
<tr>
<td>samin</td>
<td>thai</td>
</tr>
<tr>
<td>sango</td>
<td>tibetan</td>
</tr>
<tr>
<td>sangu</td>
<td>tigrinya</td>
</tr>
<tr>
<td>scottishgaelic</td>
<td>tongan</td>
</tr>
<tr>
<td>sena</td>
<td>turkish</td>
</tr>
<tr>
<td>serbian-cyrillic-bosniaherzegovina</td>
<td>turkmen</td>
</tr>
<tr>
<td>serbian-cyrillic-kosovo</td>
<td>ukenglish</td>
</tr>
<tr>
<td>serbian-cyrillic-montenegro</td>
<td>ukrainian</td>
</tr>
<tr>
<td>serbian-cyrillic</td>
<td>uppersorbian</td>
</tr>
<tr>
<td>serbian-cyrl-ba</td>
<td>urdu</td>
</tr>
<tr>
<td>serbian-cyrl-me</td>
<td>usenglish</td>
</tr>
<tr>
<td>serbian-cyrl-xk</td>
<td>usorbian</td>
</tr>
<tr>
<td>serbian-cyrl</td>
<td>uygur</td>
</tr>
<tr>
<td>serbian-latin-bosniaherzegovina</td>
<td>uzbek-arab</td>
</tr>
<tr>
<td>serbian-latin-kosovo</td>
<td>uzbek-arabic</td>
</tr>
<tr>
<td>serbian-latin-montenegro</td>
<td>uzbek-cyrillic</td>
</tr>
<tr>
<td>serbian-latin</td>
<td>uzbek-cyrl</td>
</tr>
<tr>
<td>serbian-ltn-ba</td>
<td>uzbek-latin</td>
</tr>
<tr>
<td>serbian-ltn-me</td>
<td>uzbek-latn</td>
</tr>
<tr>
<td>serbian-ltn-xf</td>
<td>uzbek</td>
</tr>
<tr>
<td>serbian-ltn</td>
<td>vai-latin</td>
</tr>
<tr>
<td>serbian</td>
<td>vai-latn</td>
</tr>
<tr>
<td>shambala</td>
<td>vai-vai</td>
</tr>
<tr>
<td>shona</td>
<td>vai-vaii</td>
</tr>
<tr>
<td>sichuanyi</td>
<td>vai</td>
</tr>
<tr>
<td>sinhala</td>
<td>vietnam</td>
</tr>
<tr>
<td>slovak</td>
<td>vietnamese</td>
</tr>
<tr>
<td>slovene</td>
<td>vunjo</td>
</tr>
<tr>
<td>slovenian</td>
<td>walser</td>
</tr>
<tr>
<td>sogal</td>
<td>welsh</td>
</tr>
<tr>
<td>somali</td>
<td>westernfrisian</td>
</tr>
<tr>
<td>spanish-mexico</td>
<td>yangben</td>
</tr>
<tr>
<td>spanish-mx</td>
<td>yiddish</td>
</tr>
<tr>
<td>spanish</td>
<td>yoruba</td>
</tr>
<tr>
<td>standardmoroccantamazight</td>
<td>zarma</td>
</tr>
<tr>
<td>swahili</td>
<td>zulu afrikaans</td>
</tr>
</tbody>
</table>
1.14 Selecting fonts

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

Here font-family is \texttt{rm}, \texttt{sf} or \texttt{tt} (or newly defined ones, as explained below), and font-name is the same as in fontspec and the like. If no language is given, then it is considered the default font for the family, activated when a language is selected. On the other hand, if there is one or more languages in the optional argument, the font will be assigned to them, overriding the default. Alternatively, you may set a font for a script – just precede its name (lowercase) with a star (eg, *devanagari*). Babel takes care of the font language and the font script when languages are selected (as well as the writing direction); see the recognized languages above. In most cases, you will not need font-options, which is the same as in fontspec, but you may add further key/value pairs if necessary.

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

\begin{verbatim}
\documentclass{article}
\usepackage[swedish, bidi=default]{babel}
\babelfont{rm}{FreeSerif}
\begin{document}
Svenska \foreignlanguage{hebrew}{יהודי svensk.}
\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 \texttt{rm}, \texttt{sf} or \texttt{tt}. This is the preferred way to select fonts in addition to the three basic ones.

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

\begin{verbatim}
\babelfont{kai}{FandolKai}
\end{verbatim}

Now, \texttt{kaifamily} and \texttt{kaidefault}, as well as \texttt{textkai} are at your disposal.

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

\[13\] See also the package combofont for a complementary approach.
This makes sure the OpenType script for Devanagari is deva and not dev2.

**NOTE** Directionality is a property affecting margins, intension, 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 \textit{ad hoc} font is active, or you select the font with this command, neither the script nor the language are passed. You must add them by hand. This is by design, for several reasons (for example, each font has its own set of features and a generic setting for several of them could be problematic, and also a "lower level" font selection is useful).

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

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

### 1.15 Modifying a language

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

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

\begin{verbatim}
\addto\captionsenglish{%
  \renewcommand\contentsname{Foo}%
}%
\end{verbatim}

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 \babelfont and its key \import, is:

\begin{verbatim}
\renewcommand\spanishchaptername{Foo}
\end{verbatim}

- Macros to be run when a language is selected can be add to \texttt{\textbackslash extras\langle lang\rangle}:
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.

### 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.

\babelprovide\[\langle options\rangle\]{\langle language-name\rangle}

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

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

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

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

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

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

**import=** \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. 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, there is a \texttt{\language\date} macro with three arguments: year, month and day numbers. In fact, \today calls \texttt{\language\today}, which in turn calls \texttt{\language\date{\the\year}{\the\month}{\the\day}}.

\begin{itemize}
\item \texttt{\language\caption=} \texttt{\language-tag}
\end{itemize}

\texttt{\language\caption=} Loads only the strings. For example:

\begin{verbatim}
\babelprovide[caption=hu]{hungarian}
\end{verbatim}

\begin{itemize}
\item \texttt{\language\hyphenrules=} \texttt{\language-list}
\end{itemize}

\texttt{\language\hyphenrules=} 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:

\begin{verbatim}
\babelprovide[hyphenrules=chavacano spanish italian]{chavacano}
\end{verbatim}

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 \texttt{\belpatterns}, as for example:

\begin{verbatim}
\babelprovide[hyphenrules=+]{neo}
\belpatterns[neo]{a1 e1 i1 o1 u1}
\end{verbatim}

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

\begin{itemize}
\item \texttt{\language\main}
\end{itemize}

\texttt{\language\main} This valueless option makes the language the main one. Only in newly defined languages.

\begin{itemize}
\item \texttt{\language\script=} \texttt{\script-name}
\end{itemize}

\texttt{\language\script=} Sets the script name to be used by fontspec (eg, Devanagari). Overrides the value in the ini file. This value is particularly important because it sets the writing direction.

\begin{itemize}
\item \texttt{\language\language=} \texttt{\language-name}
\end{itemize}

\texttt{\language\language=} Sets the language name to be used by fontspec (eg, Hindi). Overrides the value in the ini file. Not so important, but sometimes still relevant.

\textbf{NOTE} (1) If you need shorthands, you can use \texttt{\useshort} and \texttt{\defineshorthand} as described above. (2) Captions and \today are “ensured” with \texttt{\belsure} (this is be the default in ini-based languages).

\section*{1.17 Digits}

\textbf{New 3.20} A few ini files define a field named \texttt{digits\_native}. When it is present, two macros are created: \texttt{\language\digits} and \texttt{\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 option maparabic in \babelprovide, \arabic is redefined to produce the native digits (this is done globally, to avoid inconsistencies in, for example, page numbering).

For example:

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

### 1.18 Getting the current language name

\texttt{\textbackslash language}

The control sequence \texttt{\textbackslash language} contains the name of the current language.

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

\texttt{\textbackslash iflanguage}

\begin{verbatim}
{\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 \texttt{\textbackslash 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 \texttt{\textbackslash language} also applies here – use \texttt{iflang} instead of \texttt{\textbackslash iflanguage} if possible.

### 1.19 Hyphenation tools

\texttt{\textbackslash babelhyphen}\{\langle type\rangle\}

\texttt{\textbackslash babelhyphen}\{\langle text\rangle\}

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

In \TeX{}, - and \texttt{\textbackslash -} 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 \texttt{\textbackslash -}, 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.
• \texttt{\textbackslash babelhyphen\{soft\}} and \texttt{\textbackslash babelhyphen\{hard\}} are self explanatory.
• \texttt{\textbackslash 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.
• \texttt{\textbackslash babelhyphen\{nobreak\}} inserts a hard hyphen without a break after it (even if a space follows).
• \texttt{\textbackslash babelhyphen\{empty\}} inserts a break opportunity without a hyphen at all.
• \texttt{\textbackslash babelhyphen\{⟨text⟩\}} is a hard “hyphen” using ⟨text⟩ instead. A typical case is \texttt{\textbackslash 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: \texttt{\textbackslash babelhyphen\*\{soft\}} (which in most cases is equivalent to the original \texttt{-}), \texttt{\textbackslash babelhyphen\*\{hard\}}, etc.

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

There are also some differences with \texttt{\LaTeX}: (1) the character used is that set for the current font, while in \texttt{\LaTeX} it is hardwired to - (a typical value); (2) the hyphen to be used in fonts with a negative \texttt{\hyphenchar} is -, like in \texttt{\LaTeX}, but it can be changed to another value by redefining \texttt{\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).

\texttt{\textbackslash 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 \texttt{\lccodes}’s done in \texttt{\extras⟨lang⟩} as well as the language specific encoding (not set in the preamble by default). Multiple \texttt{\babelhyphenation’s} are allowed. For example:

\texttt{\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.

\texttt{\textbackslash babelpatterns\{⟨language⟩, ⟨language⟩, ...\}⟨patterns⟩}

New 3.9m In \texttt{\LaTeX} only\textsuperscript{14} 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 \texttt{\lccodes}’s done in \texttt{\extras⟨lang⟩} as well as the language specific encoding (not set in the preamble by default). Multiple \texttt{\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.

\textsuperscript{14}With \texttt{\LaTeX} exceptions and patterns can be modified almost freely. However, this is very likely a task for a separate package and \texttt{babel} only provides the most basic tools.
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 encoding (ie, sets of glyphs) depending on the language, and therefore such a switch would be in a sense incomplete. 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.

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

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

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%).

\textbf{WARNING} Setting bidi text has many subtleties (see for example <https://www.w3.org/TR/html-bidi/>). \textit{This means the babel bidi code may take some time before it is truly stable}[^15] 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.

\texttt{bidi=}  
default | basic-r | basic

\textbf{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,

[^15]: 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.

[^16]: But still defined for backwards compatibility.

[^17]: A basic stable version for luatex is planned before Summer 2018. Other engines must wait very likely until Winter.
which handles numbers and unmarked \Ltext within an \Rtext context. \newpage

Finally, basic supports both \Ltext and \Rtext (see \ref{sec:bidi:basic}). (They are named basic mainly because they only consider the intrinsic direction of scripts and weak directionality.)

\textbf{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, which will be improved in the future. Remember basic-r is available in \texttt{luatex} only.\footnote{At the time of this writing some Arabic fonts are not rendered correctly by the default \texttt{luatex} font loader, with misplaced kerns inside some words, so double check the resulting text. It seems a fix is on the way, but in the meanwhile you could have a look at the workaround available on GitHub, under \texttt{/required/babel/samples}}

\begin{verbatim}
\documentclass{article}
\usepackage{babel}
\begin{document}
\textend{document}
\end{verbatim}

\texttt{layout=} sectioning | counters | lists | contents | footnotes | captions | columns | extras

\texttt{New 3.16} \textit{To be expanded}. Selects which layout elements are adapted in bidi documents, including some text elements. You may use several options with a comma-separated list (eg, \texttt{layout=counters.contents.sectioning}). This list will be expanded in future releases (tables, captions, etc.). Note not all options are required by all engines.

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

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

With counters, \texttt{\arabic} is not only considered \Ltext always (with \texttt{\babelsublr}, see below), but also an “isolated” block which does not interact with the surrounding chars. So, while \ref{sec:arabic:1} in \Rtext is rendered in that order with \texttt{bidi=basic} (as a decimal number), in \texttt{\arabic{c1}, \arabic{c2}} the visual order is \texttt{c2 c1}. Of course, you may always adjust the order by changing the language, if necessary.\footnote{Next on the roadmap are counters and numeral systems in general. Expect some minor readjustments.}

\texttt{lists} required in \texttt{xetex} and \texttt{pdftex}, but only in multilingual documents in \texttt{luatex}.

\texttt{contents} required in \texttt{xetex} and \texttt{pdftex}; in \texttt{luatex} \texttt{toc} entries are \Rtext by default if the main language is \Rtext.

\texttt{columns} required in \texttt{xetex} and \texttt{pdftex} to reverse the column order (currently only the standard two column mode); in \texttt{luatex} they are \Rtext by default if the main language is \Rtext (including \texttt{multicol}).
footnotes  not required in monolingual documents, but it may be useful in multilingual
documents in all engines; you may use alternatively \BabelFootnote described below
(what this options does exactly is also explained there).
captions is similar to sectioning, but for \caption; not required in monolingual
documents with luatex, but may be required in xetex and pdftex in some styles (support
for the latter two engines is still experimental) New 3.18
.tabular required in luatex for R tabular (it has been tested only with simple tables, so
expect some readjustments in the future); ignored in pdftex or xetex (which will not
support a similar option in the short term) New 3.18
extras is used for miscellaneous readjustments which do not fit into the previous groups.
Currently redefines in luatex \underline{and La}TeX2ε New 3.19 .

\babelsublr {⟨lr-text⟩}
Digits in pdftex must be marked up explicitly (unlike luatex with 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 \b babel sublr in explicit L mode is ignored. However, with bidi=basic and implicit L,
it first returns to R and then switches to explicit L. This is by design to provide the proper
behaviour in the most usual cases — if you need to use \ref in an L text inside R, it
must be marked up explicitly.

\BabelPatchSection {⟨section-name⟩}
Mainly for bidi text, but it could be useful in other cases. \BabelPatchSection and the
corresponding option layout=sectioning takes a more logical approach (at least in many
cases) because it applies the global language to the section format (including the
chapt er name 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, but with
This command you can set them individually if necessary (but note then tocs and marks
are not touched).

\BabelFootnote {⟨cmd⟩}{⟨local-language⟩}{⟨before⟩}{⟨after⟩}
New 3.17 Something like:

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

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

\footnote{⟨foreignlanguage{\languagename} {}{note}}}  

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

\BabelFootnote{\footnote}{\languagename}{()}{%
\BabelFootnote{\localfootnote}{\languagename}{()}{%  
\BabelFootnote{\mainfootnote}{()}{%  

(which also redefine \footnotetext and define \localfootnotetext and
\mainfootnotetext). If the language argument is empty, then no language is selected.

31
inside the argument of the footnote. Note this command is available always in bidi documents, even without layout=footnotes.

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

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

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

### 1.22 Language attributes

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

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

### 1.23 Hooks

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

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

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

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

- **addialect** (language name, dialect name) Used by luababel.def to load the patterns if not preloaded.
- **patterns** (language name, language with encoding) Executed just after the \language has been set. The second argument has the patterns name actually selected (in the form of either lang:ENC or lang).
- **hyphenation** (language name, language with encoding) Executed locally just before exceptions given in \babelhyphenation are actually set.
- **defaultcommands** Used (locally) in \StartBabelCommands.
- **encodedcommands** (input, font encodings) Used (locally) in \StartBabelCommands. Both xetex and luatex make sure the encoded text is read correctly.
- **stopcommands** Used to reset the 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\langle language\rangle. For example, the following deactivates shorthands in all languages:

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

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

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

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

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

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

In the following table most of the languages supported by babel with and .1df 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
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, frenchle, ethiopi 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:

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

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

### 1.25 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\extrasfrench{\inputencoding{latin1}}
\addto\extrasrussian{\inputencoding{koi8-r}}

(A recent version of inputenc is required.)

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

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

• Babel does not take into account \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.
mkpatternt Generates hyphenation patterns.
tracklang Tracks which languages have been requested.
1.26 Current and future work

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

It is possible now to typeset Arabic or Hebrew with numbers and L text. Next on the roadmap are line breaking in Thai and the like, as well as “non-European” digits. Also on the roadmap are R layouts (lists, footnotes, tables, column order), page and section numbering, and maybe kashida justification.

As to Thai line breaking, here is the basic idea of what luatex can do for us, with the Thai patterns and a little script (the final version will not be so little, of course). It replaces each discretionary by the equivalent to ZWJ.

\documentclass{article}
\usepackage{babel}
\babelprovide{import=th, main}{thai}
\babelfont{rm}{FreeSerif}
\directlua{
  local GLYF = node.id'glyph'
  function insertsp (head)
    local size = 0
    for item in node.traverse(head) do
      local i = item.id
      if i == GLYF then
        f = font.getfont(item.font)
        size = f.size
      elseif i == 7 then
        local n = node.new(12, 0)
        node.setglue(n, 0, size * 1) % 1 is a factor
        node.insert_before(head, item, n)
        node.remove(head, item)
      end
    end
  end
  luatexbase.add_to_callback('hyphenate',
    function (head, tail)
      lang.hyphenate(head)
      insertsp(head)
    end, 'insertsp')
}
\begin{document}
(Thai text.)
\end{document}

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 \LaTeX{} internals.

\footnote{See for example POSIX, ISO 14652 and the Unicode Common Locale Data Repository (CLDR). Those system, however, have limited application to \LaTeX{} because their aim is just to display information and not fine typesetting.}
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.o” may be referred to as either “item 3.o” or “3.er item”, and so on.

1.27 Tentative and experimental code

**Option bidi=basic**

New 3.19 With this package option 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:

```latex
\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 fuṣḥā l-ʻaṣr (MSA) and fuṣḥā t-turāth (CA).
\end{document}
```

What mapfont=direction means is, ‘when a character has the same direction as the script for the “provided” language (arabic in this case), then change its font to that set for this language’ (here defined via *arabic, because Crimson does not provide Arabic letters). 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):

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

In a future a more complete method, reading recursively boxed text, may be added. There are samples on GitHub, under /required/babel/samples: lua-bidibasic.tex and lua-seccenum.tex.

Old stuff

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

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

So, for example:

\begin{verbatim}
\setmainfont[Language=Turkish]{Minion Pro}
\babelFSstore{turkish}
\setmainfont{Minion Pro}
\babelFSfeatures{turkish}{Language=Turkish}
\end{verbatim}

\textbf{Bidi writing} in luatex is under development, but a basic implementation is almost finished. On the other hand, in xetex it is taking its first steps. The latter engine poses quite different challenges. An option to manage document layout in luatex (lists, footnotes, etc.) is almost finished, but xetex required more work.

See the code section for \texttt{\textbackslash foreignlanguage*} (a new starred version of \texttt{\textbackslash foreignlanguage}). xetex relies on the font to properly handle these unmarked changes, so it is not under the control of \TeX{}.

\section{Loading languages with language.dat}

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

\textbf{New 3.9q} With luatex, however, patterns are loaded on the fly when requested by the language (except the “0th” language, typically \texttt{english}, which is preloaded always)\textsuperscript{\texttt{\textbackslash e\textbackslash TeX}}. Until 3.9n, this task was delegated to the package \texttt{luatex-hyphen}, by Khaled Hosny, Élie Roux, and Manuel Pégourié-Gonnard, and required an extra file named language.dat.\texttt{.lua}, but now a new mechanism has been devised based solely on language.dat. \textbf{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)\textsuperscript{\texttt{\textbackslash e\textbackslash TeX}}.

\subsection{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\textsuperscript{\texttt{\textbackslash e\textbackslash TeX}}. 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:

\begin{verbatim}
% File    : language.dat
% Purpose : tell \texttt{\textbackslash iniT\textbackslash eX} what files with patterns to load.
english    english.hyphenations
\end{verbatim}

\textsuperscript{\texttt{\textbackslash e\textbackslash TeX}}This feature was added to 3.9o, but it was buggy. Both 3.9o and 3.9p are deprecated.

\textsuperscript{\texttt{\textbackslash e\textbackslash TeX}}The loader for lua(\textbackslash e)\TeX{} is slightly different as it’s not based on babel but on etex.\texttt{.5rc}. 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.

\textsuperscript{\texttt{\textbackslash e\textbackslash TeX}}This is because different operating systems sometimes use very different file-naming conventions.
You may also set the font encoding the patterns are intended for by following the language name by a colon and the encoding code. For example:

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

With the previous settings, if the encoding when the language is selected is T1 then the patterns in hyphenT1.ger are used, but otherwise use those in hyphen.ger (note the encoding could be set in \extras{lang}).

A typical error when using babel is the following:

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

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

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

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

The following assumptions are made:

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

- The common part of the babel system redefines a number of macros and environments (defined previously in the document style) to put in the names of macros that replace the previously hard-wired texts. These macros have to be defined in the language definition files.

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

---

This in not a new feature, but in former versions it didn't work correctly.
• When a language definition file is loaded, it can define `\language{lang}` to be a dialect of `\language{language0}` when `\language{lang}` is undefined.

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

• The semantics of modifiers is not defined (on purpose). In most cases, they will just be simple separated options (eg, `spanish`), but a language might require, say, a set of options organized as a tree with suboptions (in such a case, the recommended separator is `/`).

Some recommendations:

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

• Captions should not contain shorthands or encoding dependent commands (the latter is not always possible, but should be clearly documented). They should be defined using the LICR. You may also use the new tools for encoded strings, described below.

• Avoid adding things to `\noextras{lang}` except for umlauthigh and friends, `\bbi@deactivate`, `\bbi@frenchspacing`, and language specific macros. Use always, if possible, `\bbi@save` and `\bbi@savevariable` (except if you still want to have access to the previous value). Do not reset a macro or a setting to a hardcoded value. Never. Instead save its value in `\extras{lang}`.

• Do not switch scripts. If you want to make sure a set of glyphs is used, switch either the font encoding (low level) or the language (high level, which in turn may switch the font encoding). Usage of things like `\latinintext` is deprecated.\footnote{But not removed, for backward compatibility.}

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

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

3.1 Guidelines for contributed languages

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

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

• 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 ldffilesmay 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-outerversion 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

\substitutefontfamily (Deprecated.) This command takes three arguments, a font encoding and two font family names. It creates a font description file for the first font in the given encoding. This .fd file will instruct \LaTeX to use a font from the second family when a font from the first family in the given encoding seems to be needed.

3.3 Skeleton

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

\ProvidesLanguage{<language>}{[2016/04/23 v0.0 <Language> support from the babel system]}
\LdfInit{<language>}{captions<language>}
\ifx\undefined\l@<language>
\@nopatterns{<Language>}
\adddialect\l@<language>0
\fi
\adddialect\l@<dialect>\l@<language>
\bbl@declare@attribute{<language>}{attrib}{%
\expandafter\addto\expandafter\addto\extras<language>\%
\expandafter\addto\expandafter\extras<language> 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\monthiname{<name of first month>} % More strings
\StartBabelCommands*{<dialect>}{captions} \SetString\chaptername{<chapter name>} % More strings
\StartBabelCommands*{<dialect>}{date} \SetString\monthiname{<name of first month>}
3.4 Support for active characters

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

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

\bbl@activate  The command \bbl@activate is used to change the way an active character expands.

\bbl@deactivate  \bbl@activate ‘switches on’ the active behavior of the character. \bbl@deactivate lets the active character expand to its former (mostly) non-active self.

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

\bbl@add@special\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.” \[2] p. 380\]

It is used to set text ‘verbatim’. To make this work if more characters get a special category code, you have to add this character to the macro \dospecial. \TeX adds another macro called \@sanitize representing the same character set, but without the curly braces. The macros \bbl@add@special\langle char\rangle and \bbl@remove@special\langle char\rangle add and remove the character \langle char\rangle 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\[27\] To save the current meaning of any control sequence, the macro \bbl@save is provided. It takes one argument, \langle csname\rangle, 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 \langle variable\rangle.

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

3.6 Support for extending macros

\addto  The macro \addto\langle control sequence\rangle\{\langle \TeX code\rangle\} can be used to extend the definition of

\[27\]This mechanism was introduced by Bernd Raichle.
a macro. The macro need not be defined (ie, it can be undefined or \relax). This macro can, for instance, be used in adding instructions to a macro like \extrasenglish. Be careful when using this macro, because depending on the case the assignment could be either global (usually) or local (sometimes). That does not seem very consistent, but this behavior is preserved for backward compatibility. If you are using etoolbox, by Philipp Lehman, consider using the tools provided by this package instead of \addto.

3.7 Macros common to a number of languages

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

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

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

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

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

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

3.8 Encoding-dependent strings

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

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

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

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

\StartBabelCommands
\{\langle language-list\rangle\}\{\langle category\rangle\}\{\langle selector\rangle\}
\EndBabelCommands

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

A “selector” is a name to be used as value in package option strings, optionally followed by extra info about the encodings to be used. The name unicode must be used for xetex and luatex (the key strings has also other two special values: generic and encoded).
If a string is set several times (because several blocks are read), the first one take precedence (i.e., it works much like `\providecommand`). Encoding info is `charset=` followed by a charset, which if given sets how the strings should be translated to the internal representation used by the engine, typically `utf8`, which is the only value supported currently (default is no translations). Note `charset` is applied by `luatex` and `xetex` when reading the file, not when the macro or string is used in the document. A list of font encodings which the strings are expected to work with can be given after `fontenc=` (separated with spaces, if two or more) – recommended, but not mandatory, although blocks without this key are not taken into account if you have requested `strings=encoded`. Blocks without a selector are read always if the key `strings` has been used. They provide fallback values, and therefore must be the last blocks; they should be provided always if possible and all strings should be defined somehow inside it; they can be the only blocks (mainly LGC scripts using the LICR). Blocks without a selector can be activated explicitly with `strings=generic` (no block is taken into account except those). With `strings=encoded`, strings in those blocks are set as default (internally, `?`). With `strings=encoded` strings are protected, but they are correctly expanded in `\MakeUppercase` and the like. If there is no key `strings`, string definitions are ignored, but `\SetCases` are still honoured (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`).

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

A real example is:

```latex
\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{\monthiiiname}{M"{a}rz}
\SetString{\monthivname}{April}
\SetString{\monthvname}{Mai}
```

\[28\text{In future releases further categories may be added.}\]
When used in .ldf 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, ie, if \date\language exists).

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

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

\EndBabelCommands Marks the end of the series of blocks.

\AfterBabelCommands ⟨code⟩

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

\SetString ⟨macro-name⟩{⟨string⟩}

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

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

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

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

\SetStringLoop{abmon#1name}{en,fb, mr, ab, my, jn, j1, ag, sp, oc, nv, dc}
\SetStringLoop{abday#1name}{lu, ma, mi, vi, sa, do}

#1 is replaced by the roman numeral.

\footnote{This replaces in 3.9g a short-lived \UseStrings which has been removed because it did not work.}
\SetCase\{(map-list)\}\{<toupper-code>\}\{<tolower-code>\}

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

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

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

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

(Note the mapping for OT1 is not complete.)

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

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

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

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

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

An example is (which is redundant, because these assignments are done by both \luatex and \xetex):
This macro is not intended to fix wrong mappings done by Unicode (which are the default in both \texttt{xetex} and \texttt{luatex}) – if an assignment is wrong, fix it directly.

4 Changes

4.1 Changes in \texttt{babel} version 3.9

Most of changes in version 3.9 are related to bugs, either to fix them (there were lots), or to provide some alternatives. Even new features like \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 \texttt{german}, a \texttt{\select@language{spanish}} had no effect.

- \texttt{\foreignlanguage} and other language* messed up \texttt{\extra<language>}. Scripts, encodings and many other things were not switched correctly.

- The :ENC mechanism for hyphenation patterns used the encoding of the \textit{previous} language, not that of the language being selected.

- ' (with active acute) 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.

- \texttt{\textormath} raised and error with a conditional.

- \texttt{\aliasshorthand} didn’t work (or only in a few and very specific cases).

- \texttt{\l@english} was defined incorrectly (using \texttt{\let} instead of \texttt{\chardef}).

- \texttt{ldf} files not bundled with \texttt{babel} were not recognized when called as global options.

4.2 Changes in \texttt{babel} version 3.7

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

- Shorthands are expandable again. The disadvantage is that one has to type ‘{}a when the acute accent is used as a shorthand character. The advantage is that a number of other problems (such as the breaking of ligatures, etc.) have vanished.

- Two new commands, \texttt{\shorthand} and \texttt{\shorthandoff} have been introduced to enable to temporarily switch off one or more shorthands.

- Support for typesetting Hebrew (and potential support for typesetting other right-to-left written languages) is now available thanks to Rama Porrat and Boris Lavva.

- A language attribute has been added to the \texttt{\mark...} commands in order to make sure that a Greek header line comes out right on the last page before a language switch.
• Hyphenation pattern files are now read inside a group; therefore any changes a pattern file needs to make to lowercase codes, uppercase codes, and category codes are kept local to that group. If they are needed for the language, these changes will need to be repeated and stored in \extras...

• The concept of language attributes is introduced. It is intended to give the user some control over the features a language-definition file provides. Its first use is for the Greek language, where the user can choose the πολυτονικό (“polytonikό” or multi-accented) Greek way of typesetting texts.

• The environment hyphenrules is introduced.

• The syntax of the file language.dat has been extended to allow (optionally) specifying the font encoding to be used while processing the patterns file.

• The command \ providexyphemins should now be used in language definition files in order to be able to keep any settings provided by the pattern file.

## Part II

### The code

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

### 5 Identification and loading of required files

*Code documentation is still under revision.*

The babel package after unpacking consists of the following files:

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

1 <<(version=3.20)>
2 <<(date=2018/05/02)>

### 6 Tools

*Do not use the following macros in \l df 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 \LaTeX{} is executed twice, but we need them when defining options and babel.def cannot be load until options have been defined. This does not hurt, but should be fixed somehow.

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

\begin{verbatim}
\bbl@add@list
\end{verbatim}

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

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

\begin{verbatim}
\bbl@tempa
\bbl@trim@def
This code is based on code presented in TUGboat vol. 12, no2, June 1991 in “An expansion Power Lemma” by Sonja Maus.
\end{verbatim}

\footnote{This code is based on code presented in TUGboat vol. 12, no2, June 1991 in “An expansion Power Lemma” by Sonja Maus.}
To check if a macro is defined, we create a new macro, which does the same as `\@ifundefined`. However, in an `\epsilon-tex` engine, it is based on `\ifcsname`, which is more efficient, and do not waste memory.

```
\bbl@ifunset{ifcsname}{% \bbl@ifunset{ifcsname}{% \expandafter\ifx\csname#1\endcsname\relax \bbl@afterelse\expandafter\@firstoftwo \else \bbl@afterfi\expandafter\@secondoftwo \fi \else \expandafter\@firstoftwo \fi}
```

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

```
\bbl@ifblank{ifcsname}{% \bbl@ifblank{ifcsname}{% \expandafter\ifx\csname#1\endcsname\relax \else \bbl@ifblank{#1}{}{\bbl@forkv@eq#1=\@empty=\@nil{#1}} \expandafter\bbl@kvnext \fi \else \expandafter\@firstoftwo \fi}
```

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).

```
\bbl@foreach{#1}{\expandafter\bbl@vforeach\expandafter{#1}}
```

A `for` loop. Each item (trimmed), is #1. It cannot be nested (it's doable, but we don't need it).
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 expands its arguments and then compare their expansion (sanitized, so that the catcodes do not matter). `\bbl@engine` takes the following values: 0 is `pdfLaTeX`, 1 is `luatex`, and 2 is `xetex`. You may use the latter in your language style if you want.

Some files identify themselves with a \TeX macro. The following code is placed before them to define (and then undefine) if not in \TeX.
The following code is used in babel.sty and babel.def, and loads (only once) the data in language.dat.

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

6.1 Multiple languages

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

\addlanguage To add languages to \TeX’s memory plain \TeX version 3.0 supplies \newlanguage, in a pre-3.0 environment a similar macro has to be provided. For both cases a new macro is defined here, because the original \newlanguage was defined to be \outer. For a format based on plain version 2.x, the definition of \newlanguage can not be copied because \count 19 is used for other purposes in these formats. Therefore \addlanguage is defined using a definition based on the macros used to define \newlanguage in plain \TeX version 3.0.

For formats based on plain version 3.0 the definition of \newlanguage can be simply copied, removing \outer. Plain \TeX version 3.0 uses \count 19 for this purpose.
Now we make sure all required files are loaded. When the command \AtBeginDocument does’t exist we assume that we are dealing with a plain-based format or \LaTeX{}2.09. In that case the file plain.def is needed (which also defines \AtBeginDocument, and therefore it is not loaded twice). We need the first part when the format is created, and \orig@dump is used as a flag. Otherwise, we need to use the second part, so \orig@dump is not defined (plain.def undeneffes it).
Check if the current version of switch.def has been previously loaded (mainly, hyphen.cfg). If not, load it now. We cannot load babel.def here because we first need to declare and process the package options.

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

7.1 base

The first option to be processed is base, which set the hyphenation patterns then resets ver@babel.sty so that \LaTeX{}2\epsilon forgets about the first loading. After switch.def has been loaded (above) and \AfterBabelLanguage defined, exits.

If the format created a list of loaded languages (in \bbl@languages), get the name of the 0-th to show the actual language used.
Now the base option. With it we can define (and load, with \luatex) hyphenation patterns, even if we are not interested in the rest of babel. Useful for old versions of polyglossia, too.
The following macrosextract languagemodifiers, 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.

\special{cmd:\\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\@nnil{\ifx\@empty#2%\edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%\else\in@{=}{#1}\ifin@\edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%\else\edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%\bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%\fi\fi}%\let\bbl@tempc\@empty\bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty\@nnil}%\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.

\DeclareOption{KeepShorthandsActive}{\bbl@tempa shorthands=\bbl@tempa}\langle\langle More package options\rangle\rangle

Handling of package options is done in three passes. (I [JBL] am not very happy with the idea, anyway.) The first one processes options which has been declared above or follow the syntax <key>=<value>, the second one loads the requested languages, except the main one if set with the key main, and the third one loads the latter. First, we “flag” valid keys with a nil value.

\let\bbl@opt@shorthands\@nnil\let\bbl@opt@config\@nnil\let\bbl@opt@main\@nnil\let\bbl@opt@headfoot\@nnil\let\bbl@opt@layout\@nnil

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

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

### 7.3 Conditional loading of shorthands

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

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

```latex
\def\bbl@ifshorthand#1{\bbl@xin@{\string#1}{\bbl@opt@shorthands}\ifin@\expandafter\@firstoftwo\else\expandafter\@secondoftwo\fi}
```

```latex
\def\bbl@ifshorthand#1#2#3{#2}
```

```latex
\def\bbl@trace{Conditional loading of shorthands}
```

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

\edef\bbl@opt@shorthands{\expanded@after\string\bbl@opt@shorthands@empty}

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

\bbl@ifshorthand{'}\{
\PassOptionsToPackage{activeacute}{babel}\}
\bbl@ifshorthand{`}\{
\PassOptionsToPackage{activegrave}{babel}\}
\fi
\fi

With headfoot=lang we can set the language used in heads/foots. For example, in babel/3796 just adds headfoot=english. It misuses \resetactivechars but seems to work.

\ifx\bbl@opt@headfoot\@nnil\else
\g@addto@macro\resetactivechars{\protect\@resetactivechars\@empty}
\fi
\let\protect\noexpand

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

\edef\bbl@opt@safe{BR}
\fi
\ifx\bbl@opt@main\@nnil\else
\edef\bbl@language@opts{\ifx\bbl@language@opts\@empty\else\bbl@opt@main\fi}
\fi

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

\bbl@trace{Defining IfBabelLayout}
\ifx\bbl@opt@layout\@nnil
\newcommand\IfBabelLayout[3]{#3}\else
\newcommand\IfBabelLayout[1]{\IfExpandTwopartsin@{#1}{\bbl@opt@layout}{#1}}\fi

7.4 Language options

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

\bbl@trace{Language options}
\let\bbl@afterlang\relax
\let\BabelModifiers\relax
Now, we set language options whose names are different from .ldf files.

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.
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 \texttt{bbl@language@opts} are assumed to be languages (note this list also contains the language given with \texttt{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 \texttt{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 \texttt{BibTeX} processes before):

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

Last declared language option is \texttt{\bbl@tempc}, but the last processed one was \texttt{\bbl@tempa}.\%
The main language cannot be set as both a global and a package option. Use `main=\bbl@tempc' as an 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@error{You haven't specified a language option}{You need to specify a language, either as a global option or as an optional argument to the \string\usepackage\space command; You shouldn’t try to proceed from here, type x to quit.}
\fi
⟨⟨/package⟩⟩
⟨⟨/core⟩⟩

8 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 \TeX-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, etexet, 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.

8.1 Tools

\ifx\ldf@quit\@undefined
\else
\expandafter\endinput
\fi
⟨⟨Make sure ProvidesFile is defined⟩⟩
\ProvidesFile{babel.def}{⟨⟨date⟩⟩ ⟨⟨version⟩⟩ Babel common definitions}
\ProvidesFile{babel.def}{⟨⟨date⟩⟩ ⟨⟨version⟩⟩ Babel common definitions}
\ProvidesFile{babel.def}{⟨⟨date⟩⟩ ⟨⟨version⟩⟩ Babel common definitions}
The file babel.def expects some definitions made in the \LaTeX2ε style file. So, in \LaTeX2.09 and Plain we must provide at least some predefined values as well some tools to set them (even if not all options are available). There in no package options, and therefore and alternative mechanism is provided. For the moment, only \texttt{\b babeloptionstrings} and \texttt{\b babeloptionmath} are provided, which can be defined before loading babel.

\texttt{\B babelModifiers} can be set too (but not sure it works).

\begin{verbatim}
\ifx\bbl@ifshorthand\@undefined
  \let\bbl@opt@shorthands\@nnil
\def\bbl@ifshorthand#1#2#3{%\1#2#3\2}%
\let\bbl@language@opts\@empty
\ifx\b babeloptionstrings\@undefined
  \let\bbl@opt@strings\@nnil
\else
  \let\bbl@opt@strings\b babeloptionstrings
\fi
\def\B babelStringsDefault{generic}
\def\bbl@tempa{normal}
\ifx\babeloptionmath\bbl@tempa
  \def\bbl@mathnormal{\noexpand\textormath}
\fi
\def\AfterB babelLanguage#1#2{}
\ifx\B babelModifiers\@undefined
  \let\B babelModifiers\relax\fi
\let\bbl@afterlang\relax
\def\bbl@opt@safe{BR}
\ifx\@uclclist\@undefined
  \let\@uclclist\@empty\fi
\ifx\bbl@trace\@undefined
  \def\bbl@trace#1{}
\fi
\fi
\par
\and continue.
\fi
\def\bbl@switchflag\@undefined % Prevent double input
\let\bbl@switchflag\relax
\input switch.def\relax
\fi
\bbl@trace{Compatibility with language.def}
\ifx\directlua\@undefined
  \openin1 = language.def
\iff ill
    \message{I couldn't find the file language.def}
  \else
    \closein1
  \fi
\fi
\chardef\l@english\z@
\fi
\langle\langle Load patterns in luatex\rangle\rangle
\langle\langle Basic macros\rangle\rangle
\end{verbatim}
\addto \ For each language four control sequences have to be defined that control the language-specific definitions. To be able to add something to these macro once they have been defined the macro \addto is introduced. It takes two arguments, a \langle control sequence \rangle and \TeX-code to be added to the \langle control sequence \rangle.

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

\[
\begin{align*}
\text{512} & \ \text{\addto}\#1\#2\{\% \\
\text{513} & \ \text{\ifx}\#1\text@undefined \% \\
\text{514} & \ \text{\def}\#1\{\#2\}\% \\
\text{515} & \ \text{\else} \\
\text{516} & \ \text{\ifx}\#1\text@relax \% \\
\text{517} & \ \text{\def}\#1\{\#2\}\% \\
\text{518} & \ \text{\else} \\
\text{519} & \ \{\text{\toks}\expandafter{\#1\#2}\}\% \\
\text{520} & \ \text{\def}\#1\{\text{\the}\toks\}\% \\
\text{521} & \ \text{\fi} \\
\text{522} & \ \text{\fi}\}
\end{align*}
\]

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

\[
\begin{align*}
\text{523} & \ \text{\bbl@withactive}\#1\#2\{\% \\
\text{524} & \ \text{\begingroup} \\
\text{525} & \ \text{\lcname=`}\#2\text@relax \\
\text{526} & \ \text{\lowercase}\{\text{\endgroup}\#1\-}\}
\end{align*}
\]

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

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

\[
\begin{align*}
\text{527} & \ \text{\bbl@redefine}\#1\{\% \\
\text{528} & \ \text{\edef}\bbl@tempa\{\bbl@stripslash\#1\}\% \\
\text{529} & \ \text{\expandafter}\let\csname org@\bbl@tempa\endcsname\#1\% \\
\text{530} & \ \text{\expandafter}\expandafter\expandafter\def\csname\bbl@tempa\endcsname\}
\end{align*}
\]

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

\[
\begin{align*}
\text{531} & \ \@onlypreamble\bbl@redefine
\end{align*}
\]

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

\[
\begin{align*}
\text{532} & \ \text{\bbl@redefine@long}\#1\{\% \\
\text{533} & \ \text{\edef}\bbl@tempa\{\bbl@stripslash\#1\}\% \\
\text{534} & \ \text{\expandafter}\let\csname org@\bbl@tempa\endcsname\#1\% \\
\text{535} & \ \text{\expandafter}\expandafter\expandafter\def\csname\bbl@tempa\endcsname\}
\end{align*}
\]

\@onlypreamble\bbl@redefine@long

\bbl@redefinerobust
\ For commands that are redefined, but which might be robust we need a slightly more intelligent macro. A robust command foo is defined to expand to \protect\foo. So it is necessary to check whether \foo exists. The result is that the command that is being redefined is always robust afterwards. Therefore all we need to do now is define \foo.

\[
\begin{align*}
\text{537} & \ \text{\bbl@redefinerobust}\#1\{\% \\
\text{538} & \ \text{\edef}\bbl@tempa\{\bbl@stripslash\#1\}\%
\end{align*}
\]

63
This command should only be used in the preamble of the document.

8.2 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).

\babelensure

The user command just parses the optional argument and creates a new macro named \bbl@e@\langle language \rangle. 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@\langle language \rangle 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.
8.3 Setting up language files

\texttt{LdfInit} The second version of \texttt{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 \texttt{\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 \texttt{LdfInit} is a control sequence. We do that by looking at the first token after passing \texttt{#2} through \texttt{string}. When it is equal to \texttt{\@backslashchar} we are dealing with a control sequence which we can compare with \texttt{\@undefined}.

If so, we call \texttt{LdfQuit} to set the main language, restore the category code of the @-sign and call \texttt{endinput}.

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

Finally we check \texttt{\originalTeX}.

\begin{verbatim}
627 \bbl@trace{Macros for setting language files up}
628 \def\bbl@ldfinit{%
629 \let\bbl@screset@empty
630 \let\BabelStrings\bbl@opt@string
631 \let\BabelOptions@empty
632 \let\BabelLanguages\relax
633 \ifx\originalTeX@undefined
634 \let\originalTeX@empty
635 \else
636 \originalTeX
637 \fi}
638 \def\LdfInit#1#2{%
639 \chardef\atcatcode=@code`@\n
640 \catcode`@=11\relax
641 \chardef\eqcatcode=@code`=\n
642 \catcode`=12\relax

643 \expandafter\if\expandafter\@backslashchar
644 \expandafter\@car\string#2\@nil
645 \ifx#2\@undefined\else
646 \LdfQuit{#1}\
647 \fi
648 \else
649 \expandafter\ifx\csname#2\endcsname\relax\else
650 \LdfQuit{#1}\
651 \fi
652 \fi
653 \bbl@ldfinit}
\end{verbatim}

\texttt{LdfQuit} This macro interrupts the processing of a language definition file.

\begin{verbatim}
654 \def\LdfQuit#1{%
655 \expandafter\main@language\expandafter{#1}\
656 \catcode`@=\atcatcode \let\atcatcode\relax
657 \catcode`=\eqcatcode \let\eqcatcode\relax
658 \endinput}
\end{verbatim}
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.

\def\bbl@afterldf#1{%  
\let\bbl@afterlang\relax  
\let\BabelModifiers\relax  
\let\bbl@screset\relax}

\def\ldf@finish#1{%  
\loadlocalcfg{#1}  
\bbl@afterldf{#1}  
\expandafter\main@language\expandafter{#1}  
\catcode`@=\atcatcode \let\atcatcode\relax  
\catcode`==\eqcatcode \let\eqcatcode\relax}

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

\@onlypreamble\LdfInit
\@onlypreamble\ldf@quit
\@onlypreamble\ldf@finish

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.

\def\main@language#1{%  
\def\bbl@main@language{#1}  
\let\languagename\bbl@main@language  
\bbl@patterns{\languagename}}

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.

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

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

\def\selectlanguage@@#1{%  
\ifcase\bbl@select@type  
\bbl@ifsamestring{\languagename}{#1}{}{\selectlanguage{#1}}%  
\else  
  \selectlanguage{#1}%  
\fi}

8.4 Shorthands

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

Items are added to the lists without checking its existence or the original catcode. It does not hurt, but should be fixed. It's already done with \nfss@catcodes, added in 3.10.
\bbl@active@def#1#2#3#4{\@namedef{#3#1}{\expandafter\ifx\csname#2@sh@#1@\endcsname\relax\bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}\else\bbl@makeother{\{#1\}}\bbl@makeother{\{#2\}}\fi}}

\initiate@active@char

A language definition file can call this macro to make a character active. This macro takes one argument, the character that is to be made active. When the character was already active this macro does nothing. Otherwise, this macro defines the control sequence \normal@char{char} to expand to the character in its ‘normal state’ and it defines the active character to expand to \normal@char{char} by default ((char) being the character to be made active). Later its definition can be changed to expand to \normal@char{char} by calling \bbl@activate{char}.

For example, to make the double quote character active one could have \initiate@active@char{"} in a language definition file. This defines " as \active@prefix "\active@char" (where the first " is the character with its original catcode, when the shorthand is created, and \active@char" is a single token). In protected contexts, it expands to \protect " or \noexpand " (ie, with the original "); otherwise \active@char" is executed. This macro in turn expands to \normal@char" in “safe” contexts (eg, \label), but \user@active" in normal “unsafe” ones. The latter search a definition in the user, language and system levels, in this order, but if none is found, \normal@char" is used. However, a deactivated shorthand (with \bbl@deactivate is defined as \active@prefix "\normal@char".

The following macro is used to define shorthands in the three levels. It takes 4 arguments: the (string'ed) character, \level@group, \level@active and \next@level@active (except in system).

\bbl@active@def#1#2#3#4{\@namedef{#3#1}{\expandafter\ifx\csname#2@sh@#1@\endcsname\relax\bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}\else\bbl@makeother{\{#1\}}\bbl@makeother{\{#2\}}\fi}}
When there is also no current-level shorthand with an argument we will check whether there is a next-level defined shorthand for this active character.

\long@namedef{#3@arg#1}{##1}{% \expandafter\ifx\csname#2@sh@#1@string##1@endcsname\relax \bbl@afterelse\csname#4#1\endcsname##1\% \else \bbl@afterfi\csname#2@sh@#1@string##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 ('ed) and the original one. This trick simplifies the code a lot.

\def\initiate@active@char#1{% \bbl@ifunset{active@char\string#1}{\bbl@withactive \{\expandafter\@initiate@active@char\expandafter}#1\string#1#1\}{}\}

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).

\def\@initiate@active@char#1#2#3{% \bbl@csarg\edef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}% \ifx#1\@undefined \bbl@csarg\edef{oridef@#2}{\let
oexpand#1
oexpand\@undefined}\else \bbl@csarg\let{oridef@@#2}#1\bbl@csarg\edef{oridef@#2}{\let
oexpand#1\expandafter
oexpand\csname bbl@oridef@@#2\endcsname}\fi

If the character is already active we provide the default expansion under this shorthand mechanism. Otherwise we write a message in the transcript file, and define \normal@char(char) to expand to the character in its default state. If the character is mathematically active when babel is loaded (for example ‘) the normal expansion is somewhat different to avoid an infinite loop (but it does not prevent the loop if the mathcode is set to ’8000 a posteriori).

\ifx#1#3\relax \expandafter\let\csname normal@char#2\endcsname#3\else \bbl@info{Making #2 an active character}\fi

\ifnum\mathcode`#2=’8000 \namedef{normal@char#2}{% \textormath{#3}{\csname bbl@oridef@@#2@endcsname}}\else \namedef{normal@char#2}{#3}\fi

To prevent problems with the loading of other packages after babel we reset the catcode of the character to the original one at the end of the package and of each language file (except with KeepShorthandsActive). It is re-activate again at \begin{document}. We also need to make sure that the shorthands are active during the processing of the .aux file. Otherwise some citations may give unexpected results in the printout when a shorthand was used in
the optional argument of `\bibitem` for example. Then we make it active (not strictly necessary, but done for backward compatibility).

```latex
\bbl@restoreactive{#2}\
\AtBeginDocument{\catcode`#2=active\if@filesw\immediate\write\@mainaux{\catcode`string#2=active}\fi}%
\expandafter\bbl@add@special\csname#2\endcsname\catcode`#2=active\fi
```

Now we have set `\normal@char ⟨char⟩`, we must define `\active@char ⟨char⟩`, to be executed when the character is activated. We define the first level expansion of `\active@char ⟨char⟩` to check the status of the `@safe@actives` flag. If it is set to true we expand to the ‘normal’ version of this character, otherwise we call `\user@active ⟨char⟩` to start the search of a definition in the user, language and system levels (or eventually `\normal@char ⟨char⟩`).

```latex
\let\bbl@tempa\@firstoftwo\if\string^#2%\def\bbl@tempa{\noexpand\textormath}%\else\ifx\bbl@mathnormal\@undefined\else\let\bbl@tempa\bbl@mathnormal\fi\fi\expandafter\edef\csname active@char#2\endcsname{\bbl@tempa{\noexpand\if@safe@actives\noexpand\expandafter\expandafter\csname normal@char#2\endcsname\noexpand\else\noexpand\expandafter\expandafter\csname bbl@doactive#2\endcsname\noexpand\fi}\expandafter\noexpand\csname normal@char#2\endcsname}}\bbl@csarg\edef{doactive#2}{\noexpand\csname user@active#2\endcsname}\
```

We now define the default values which the shorthand is set to when activated or deactivated. It is set to the deactivated form (globally), so that the character expands to `\active@prefix ⟨char⟩ \normal@char ⟨char⟩` (where `\active@char ⟨char⟩` is one control sequence!).

```latex
\bbl@csarg\edef{active@#2}{\noexpand\active@prefix\noexpand#1\expandafter\noexpand\csname active@char#2\endcsname}\bbl@csarg\edef{normal@#2}{\noexpand\active@prefix\noexpand#1\expandafter\noexpand\csname normal@char#2\endcsname}\expandafter\let\expandafter#1\csname bbl@normal@#2\endcsname\
```

The next level of the code checks whether a user has defined a shorthand for himself with this character. First we check for a single character shorthand. If that doesn’t exist we check for a shorthand with an argument.

```latex
\bbl@active@def#2\user@group{user@active}{language@active}\bbl@active@def#2\language@group{language@active}{system@active}\bbl@active@def#2\system@group{system@active}{normal@char}\
```
In order to do the right thing when a shorthand with an argument is used by itself at the end of the line we provide a definition for the case of an empty argument. For that case we let the shorthand character expand to its non-active self. Also, When a shorthand combination such as ' ends up in a heading \TeX would see \protect\protect'. To prevent this from happening a couple of shorthand needs to be defined at user level.

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

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

This command helps the shorthand supporting macros to select how to proceed. Note that this macro needs to be expandable as do all the shorthand macros in order for them to work in expansion-only environments such as the argument of \hyphenation. This macro expects the name of a group of shorthands in its first argument and a shorthand character in its second argument. It will expand to either \bbl@firstcs or \bbl@scndcs. Hence two more arguments need to follow it.

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 \langle char \rangle.

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 \langle char \rangle in the case of \bbl@activate, or \normal@char \langle char \rangle 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 math mode. To achieve this the helper macro `\textormath` is provided.

```latex
\def\textormath{\ifmmode\expandafter\@secondoftwo\else\expandafter\@firstoftwo\fi}
```

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'.

```latex
\def\usergroup{user}
\def\languagegroup{english}
\def\systemgroup{system}
```

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.

```latex
\def\useshorthands{%
  \@ifstar{\bbl@usesh@s{}}{\bbl@usesh@x{}}%\bbl@usesh@s{x} {% 
  \bbl@ifshorthand{x}{%\AddBabelHook{babel-sh-\string{x}}{afterextras}{\bbl@activate(x)}% #1
  \bbl@ifshorthand{x}{%\bbl@usesh@s{x} {% 
  \bbl@ifshorthand{x}{%\bbl@activate(x)}% #1
  \bbl@error{Cannot declare a shorthand turned off (\string{x})}%
  {Sorry, but you cannot use shorthands which have been turned off in the package options}}}%
```
\defineshorthand

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

\def\user@language@group{user@\language@group}
\def\bbl@set@user@generic#1#2{\
  \bbl@ifunset{user@generic@active#1}\
  {\bbl@active@def#1\user@language@group{user@active}{user@generic@active}\
   \expandafter\edef\csname#2@sh@#1@@\endcsname{\
     \expandafter\noexpand\csname normal@char#1\endcsname}\
   \expandafter\edef\csname#2@sh@#1@\string\protect\endcsname{%\n   \expandafter\noexpand\csname user@active#1\endcsname}\
  }\
  \@empty}

\newcommand\defineshorthand[3][user]{\
  \edef\bbl@tempa{\zap@space#1 \@empty}\
  \bbl@for\bbl@tempb\bbl@tempa{\
    \if*\expandafter\@car\bbl@tempb\@nil\
      \edef\bbl@tempb{user@\expandafter\@gobble\bbl@tempb}\
      \@expandtwoargs\
      \bbl@set@user@generic{\expandafter\string\@car#2\@nil}\bbl@tempb\
    \fi\
    \declaredeshorthand{\bbl@tempb}{#2}{#3}}}|

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

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

\aliasshorthand
First the new shorthand needs to be initialized,

\def\aliasshorthand#1#2{\
  \bbl@ifshorthand{#2}{\expandafter\@ifundefined\@csname active@char\string#2\endcsname{\
    \document@notprerr\
    \notshorthand{#2}{\expandafter\@csname active@char\string#2\endcsname}}{\
    \else\
    \initiate@active@char{#2}}}\

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

\expandafter\let\csname active@char\string#2\endcsname\relax\
\if\document@notprerr\
\else\
\initiate@active@char{#2}}\
\fi\
\fi\
\{\bbl@error\
\{\Cannot declare a shorthand turned off \string#2{\}}\
\{\Sorry, but you cannot use shorthands which have been\%\
\turned off in the package options}}

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

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.
\newcommand*{\shorthandon}[1]{\bbl@switch@sh\@ne#1\@nnil}
\DeclareRobustCommand*{\shorthandoff}\
{\ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
\def\bbl@shorthandoff#1#2{\bbl@switch@sh#1#2\@nnil}

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.
\def\bbl@switch@sh#1#2{\
\ifx#2\@nnil\else
\bbl@ifunset{bbl@active@\string#2}\
{\bbl@error
{I cannot switch `\string#2' on or off--not a shorthand}\
{This character is not a shorthand. Maybe you made\%
a typing mistake? I will ignore your instruction}}\
{\ifcase#1\%
\catcode`#212\relax
\or
\catcode`#2\active
\or
\csname bbl@oricat@\string#2\endcsname
\csname bbl@oridef@\string#2\endcsname
\fi}\
\bbl@afterfi\bbl@switch@sh#1\%
\fi}

Note the value is that at the expansion time, eg, in the preamble shorthands are usually
deactivated.
\def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
\def\bbl@putsh#1{%\languagename
\bbl@ifunset{bbl@active@\string#1}%
{\bbl@error
{I cannot switch `\string#1' on or off--not a shorthand}%
{This character is not a shorthand. Maybe you made\%
a typing mistake? I will ignore your instruction}%
{\ifcase#1%
\catcode`#212\relax
\or
\catcode`#2\active
\or
\csname bbl@oricat@\string#2\endcsname
\csname bbl@oridef@\string#2\endcsname
\fi}\
\bbl@afterfi\bbl@switch@sh#1%
One of the internal macros that are involved in substituting \prime for each right quote in mathmode is \prim@s. This checks if the next character is a right quote. When the right quote is active, the definition of this macro needs to be adapted to look also for an active right quote; the hat could be active, too.

\def\prim@s{\prime\futurelet\@let@token\prim@s}
\def\prim@eif@primes#1#2{\ifx#1\@let@token\expandafter\@firstoftwo\else\ifx#2\@let@token\bbl@afterelse\expandafter\@firstoftwo\else\bbl@afterfi\expandafter\@secondoftwo\fi\fi}
\begin{group}
\catcode`\^=7 \catcode`*=12 \lccode`*=\^ \lccode`\^=`
\let\prim@eif@primes\pr@@s
\gdef\prim@s{\pr@@t\egroup}
\end{group}

Usually the ~ is active and expands to \penalty\@M\␣. When it is written to the .aux file it is written expanded. To prevent that and to be able to use the character ~ as a start character for a shorthand, it is redefined here as a one character shorthand on system level. The system declaration is in most cases redundant (when ~ is still a non-break space), and in some cases is inconvenient (if ~ has been redefined); however, for backward compatibility it is maintained (some existing documents may rely on the babel value).

\begin{group}
\catcode`\^=7 \catcode`*=12 \lccode`*=\^ \lccode`\^=`
\lowercase{\gdef\prim@s\pr@@s{\pr@@t\egroup}}
\end{group}

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.

\def\OT1dqpos{127}
\def\T1dqpos{4}

When the macro \f@encoding is undefined (as it is in plain \TeX) we define it here to expand to OT1
\def\f@encoding{OT1}
\fi
8.5 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.

```latex
\languageattribute[2]{% \def\bbl@tempc{#1} \bbl@fixname\bbl@tempc \bbl@iflanguage\bbl@tempc{% \bbl@vforeach{#2}{% Wewanttomakesurethateachattributeisselectedonlyonce;thereforewestorethe alreadyselectedattributesin `\bbl@known@attribs`. Whenthatcontrolsequenceisnotyet definedthisattributeiscertainlynotselectedbefore.
\ifs\bbl@known@attribs\@undefined \in@false \else \bbl@xin{,\bbl@tempc-##1,}{,\bbl@known@attribs,} \fi Whentheattributewasinthelistweissueawarning;thismightnotbetheusersintention.
\ifin@ \bbl@warning{You have more than once selected the attribute '##1' for language #1} \else Whenweendupheretheattributeisnotselectedbefore. So,wearrddittothelistof selectedattributesandexecutetheassociated \TeX-code.
\bbl@exp{\bbl@add@list\bbl@known@attribs{\bbl@tempc-##1}} \edef\bbl@tempa{\bbl@tempc-##1} \expandafter\bbl@ifknown@trib\expandafter{\bbl@tempa}\bbl@attributes\{\csname\bbl@tempc @attr@##1\endcsname\} \@attrerr{\bbl@tempc}{##1} \fi}}}
```

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

```latex
\@onlypreamble\languageattribute
```

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

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

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}`.
\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.

\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@attribs\@undefined
\in@false
\else
The we need to check the list of known attributes.
\bbl@xin@{,#1-#2,}{,\bbl@known@attribs,}%
\fi
When we're this far \in@ 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'.
\ifin@
\bbl@afterelse#3%
\else
\bbl@afterfi#4%
\fi%
}

\bbl@ifknown@trib An internal macro to check whether a given language/attribute is known. The macro takes 4 arguments, the language/attribute, the attribute list, the \TeX-code to be executed when the attribute is known and the \TeX-code to be executed otherwise.
\bbl@ifknown@trib#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.
\bbl@loopx\bbl@tempb(#2){%
\expandafter\in@\expandafter{\expandafter,\bbl@tempb,%}
When a match is found the definition of \bbl@tempa is changed.
\let\bbl@tempa\@firstoftwo
Finally we execute \bbl@tempa.
\bbl@tempa
}
\bbl@clear@ttribs  This macro removes all the attribute code from \LaTeX's memory at \begin{document} time (if any is present).
\def\bbl@clear@ttribs{%  \ifx\bbl@attributes\@undefined%\else%\bbl@loopx\bbl@tempa{\bbl@attributes}{%\expandafter\bbl@clear@ttrib\bbl@tempa.%}%\let\bbl@attributes\@undefined%\fi%\def\bbl@clear@ttrib#1-#2.{%\expandafter\let\csname#1@attr@#2\endcsname\@undefined%}
\AtBeginDocument{\bbl@clear@ttribs}

8.6 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.

\bbl@savecnt  The initialization of a new save cycle: reset the counter to zero.
\def\bbl@beginsave{\bbl@savecnt\z@}
\def\bbl@save#1{%\expandafter\let\csname babel@\number\bbl@savecnt\endcsname#1\relax%\toks@\expandafter{\originalTeX\let#1=}%\bbl@exp{%\def\originalTeX{\the\toks@\<babel@\number\bbl@savecnt>\relax}%\advance\bbl@savecnt\@one%}
\bbl@savevariable  The macro \bbl@savevariable{\csname} saves the current meaning of the control sequence \csname, to \originalTeX. 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.
\bbl@savevariable#1{%\toks@\expandafter{\originalTeX\let#1=}%\bbl@exp{%\def\originalTeX{\the\toks@\<bbl@\number\bbl@savecnt>\relax}%\advance\bbl@savecnt\@ne%}
\bbl@savevariable  The macro \bbl@savevariable{\variable} saves the value of the variable. \variable can be anything allowed after the \the primitive.
\def\bbl@savevariable#1%{\toks@\expandafter{\originalTeX \#1=}%\bbl@exp{%\def\originalTeX{\the\toks@\the\#1\relax}%}}
\bbl@frenchspacing  Some languages need to have \frenchspacing in effect. Others don't want that. The command \bbl@frenchspacing switches it on when it isn't already in effect and \bbl@nonfrenchspacing switches it off if necessary.
\def\bbl@frenchspacing%{\ifnum\the\sfcode`\m\relax%\let\bbl@nonfrenchspacing\@undefined%\let\bbl@frenchspacing\relax%}%
8.7 Short tags
\babeltags
This macro is straightforward. After zapping spaces, we loop over the list and define the macros \text{⟨tag⟩} and \⟨tag⟩. Definitions are first expanded so that they don't contain \csname but the actual macro.

8.8 Hyphens
\babelhyphenation
This macro saves hyphenation exceptions. Two macros are used to store them: \babelhyphenation@ for the global ones and \babelhyphenation<lang> for language ones. See \babelpatterns above for further details. We make sure there is a space between words when multiple commands are used.
This macro makes hyphenation possible. Basically its definition is nothing more than \nobreak \hskip Opt plus Opt.

\bbl@allowhyphens

Macros to insert common hyphens. Note the space before @ in \bbl@hyphen. 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@nullhyphen\bbl@hyphenchar}
\def\bbl@hyphen@i#1#2{%\bbl@ifunset{bbl@hy@#1#2\@empty}%{\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%\bbl@hy@#1#2\@empty\endcsname}

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

\nobreak is always preceded by \leavevmode, in case the shorthand starts a paragraph.

\def\bbl@usehyphen{\leavevmode\ifdim\lastskip>\z@\mbox{#1}\else
nobreak#1
\nobreak\hskip\z@skip}
\def\bbl@@usehyphen{\leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

The following macro inserts the hyphen char.

\def\bbl@hyphenchar{\ifnum\hyphenchar\font=\m@ne
\bbl@nullhyphen\char\hyphenchar\font\else
\char\hyphenchar\font\fi}

Finally, we define the hyphen “types”. Their names will not change, so you may use them in ldf’s. After a space, the \mbox in \bbl@hy@nobreak is redundant.

\def\bbl@hy@soft{\bbl@usehyphen\discretionary{\bbl@hyphenchar}{}}
\def\bbl@hy@soft{\bbl@usehyphen\discretionary{\bbl@hyphenchar}{}}
\def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
\def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
\def\bbl@hy@nobreak{\bbl@usehyphen\mbox{\bbl@hyphenchar}}
\def\bbl@hy@nobreak{\bbl@usehyphen\mbox{\bbl@hyphenchar}}
\def\bbl@hy@repeat{%\bbl@usehyphen\discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
\def\bbl@hy@repeat{%\bbl@usehyphen\discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
\def\bbl@hy@repeat{%\bbl@usehyphen{\bbl@hyphenchar{}}
\def\bbl@hy@repeat{%\bbl@usehyphen{\bbl@hyphenchar{}}

\textsuperscript{32}\TeX begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.
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

8.9 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 \texttt{luatex} and \texttt{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.

\bbl@trace{Multiencoding strings}
\def\bbl@toglobal#1{\global\let#1#1}
\def\bbl@recatcode#1{\@tempcnta="7F
\def\bbl@tempa{\ifnum\@tempcnta>"FF\else\catcode\@tempcnta=#1\relax\advance\@tempcnta\@ne\expandafter\bbl@tempa\fi}%
\bbl@tempa}

The second one. We need to patch \@uclclist, but it is done once and only if \texttt{SetCase} is used or if strings are encoded. The code is far from satisfactory for several reasons, including the fact \@uclclist is not a list any more. Therefore a package option is added to ignore it. Instead of gobbling the macro getting the next two elements (usually \reserved@a), we pass it as argument to \bbl@uclc. The parser is restarted inside \langle\texttt{lang}\rangle@bbl@uclc because we do not know how many expansions are necessary (depends on whether strings are encoded). The last part is tricky – when uppercasing, we have:

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

and starts over (and similarly when lowercasing).
\ifpackagewith{babel}{nocase}%
{\let\bbl@patchuclc\relax}%
{\def\bbl@patchuclc%\global\let\bbl@patchuclc\relax
\g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}%\gdef\bbl@uclc##1{\let\bbl@encoded\bbl@encoded@uclc
\bbl@ifunset{\languagename @bbl@uclc}% and resumes it
{##1}%\let\bbl@tempa##1\relax % Used by LANG@bbl@uclc\csname\languagename @bbl@uc\endcsname}%
{\bbl@tolower\@empty}\bbl@toupper\@empty}%
{\gdef\bbl@tolower{\csname\languagename @bbl@uc\endcsname}%\gdef\bbl@toupper{\csname\languagename @bbl@uc\endcsname}}%
\fi%
\def\bbl@patchuclc{\global\let\bbl@patchuclc\relax}
\g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}%\gdef\bbl@uclc##1{\let\bbl@encoded\bbl@encoded@uclc
\bbl@ifunset{\languagename @bbl@uclc}% and resumes it
{##1}%\let\bbl@tempa##1\relax % Used by LANG@bbl@uclc\csname\languagename @bbl@uc\endcsname}%
{\bbl@tolower\@empty}\bbl@toupper\@empty}%
{\gdef\bbl@tolower{\csname\languagename @bbl@uc\endcsname}%\gdef\bbl@toupper{\csname\languagename @bbl@uc\endcsname}}%

\DeclareOption{nocase}{}
\ DeclareOption{nocase}{}
\ DeclareOption{nocase}{}

82
The following package options control the behavior of \SetString.

```
1203 \{\langle More package options\rangle\} ≡
1204 \let\bbl@opt@strings\@nnil % accept strings=value
1205 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1206 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1207 \def\BabelStringsDefault{generic}
1208 \{\langle More package options\rangle\}
```

**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.

```
1209 \onlyin\preamble\StartBabelCommands
1210 \def\StartBabelCommands{%
1211 \begingroup
1212 \bbl@recatcode{11}%
1213 \{\langle Macros local to BabelCommands\rangle\}
1214 \def\bbl@provstring##1##2{%
1215 \providecommand##1{##2}%
1216 \bbl@toglobal##1}
1217 \global\let\bbl@scafter\@empty
1218 \let\StartBabelCommands\bbl@startcmds
1219 \ifx\BabelLanguages\relax
1220 \begingroup
1221 \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
1222 \StartBabelCommands%
1223 \ifx\bbl@screset\@nnil
1224 \bbl@usehooks{stopcommands}{}%
1225 \fi
1226 \fi
1227 \bbl@startcmds@i
1228 \bbl@startcmds@i
```

Parse the encoding info to get the label, input, and font parts. Select the behavior of \SetString. There are two main cases, depending if there is an optional argument: without it and strings=encoded, strings are defined always; otherwise, they are set only if they are still undefined (ie, fallback values). With labelled blocks and strings=encoded, define the strings, but with another value, define strings only if the current label or font encoding is the value of strings; otherwise (ie, no strings or a block whose label is not in strings=) do nothing. We presume the current block is not loaded, and therefore set (above) a couple of default values to gobble the arguments. Then, these macros are redefined if necessary according to several parameters.

```
1241 \newcommand\bbl@startcmds@i[1][\empty]{%
1242 \let\SetString\@gobbletwo
1243 \let\bbl@stringdef\@gobbletwo
```

83
There are two versions of \bbl@scswitch. The first version is used when ldfs are read, and it makes sure \langle group \rangle \langle language \rangle is reset, but only once (\bbl@sctest is used to keep
track of this). The second version is used in the preamble and packages loaded after babel and does nothing. The macro \bbl@forlang \bbl@L but its body is executed only if the value is in \BabelLanguages (inside babel) or \date\langle language\rangle is defined (after babel has been loaded). There are also two versions of \bbl@forlang. The first one skips the current iteration if the language is not in \BabelLanguages (used in ldfs), and the second one skips undefined languages (after babel has been loaded).

1300 \def\bbl@forlang#1#2{%
1301 \bbl@for#1\bbl@L{%
1302 \bbl@xin@{,#1,}{,\BabelLanguages,}%
1303 \ifin@#2\relax\fi}
1304 \def\bbl@scswitch{%
1305 \bbl@forlang\bbl@tempa{%
1306 \ifx\bbl@G\@empty\else
1307 \ifx\SetString\@gobbletwo\else
1308 \edef\bbl@GL{\bbl@G\bbl@tempa}%
1309 \bbl@xin@{,\bbl@GL,}{,\bbl@screset,}%
1310 \ifin@%
1311 \global\expandafter % TODO - con \bbl@exp ?
1312 \edef\bbl@screset{}%
1313 \fi
1314 \fi%
1315 \fi}
1316 \AtEndOfFile{%
1317 \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{}{#2}}}%
1318 \let\bbl@scswitch\relax}
1319 \onlypreamble\EndBabelCommands
1320 \def\EndBabelCommands{%
1321 \bbl@usehooks{stopcommands}{}%
1322 \endgroup
1323 \endgroup
1324 \bbl@scafter}

Now we define commands to be used inside \StartBabelCommands.

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

1325 \def\bbl@setstring#1#2{%
1326 \bbl@forlang\bbl@tempa{%
1327 \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
1328 \bbl@ifunset{\bbl@LC}{\ifgerman\bbl@setlanguage{\germanchaptername}{\bbl@tempa}{}}%
1329 \global\expandafter % TODO - con \bbl@exp ?
1330 \bbl@add\csname\bbl@LC\bbl@tempa\csname\endcsname\endcsname\expandafter
1331 \expandafter\bbl@setstring\bbl@LC\csname\endcsname\expandafter{\bbl@tempa}{}}%
1332 \}%
1333 \def\BabelString{#2}%
1334 \bbl@usehooks{stringprocess}{}%
1335 \expandafter\bbl@stringdef
1336 \csname\bbl@LC\bbl@tempa\expandafter\csname\endcsname\expandafter{\bbl@tempa}{}}%

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.

1337 \ifx\bbl@opt@strings\relax

85
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.

Delaying code

Now the definition of \AfterBabelCommands when it is activated.

Case mapping

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

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.
Macros local to BabelCommands
\newcommand\SetHyphenMap[1][]{\bbl@forlang\bbl@tempa% \expandafter\bbl@stringdef\csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}

There are 3 helper macros which do most of the work for you.
\newcommand\BabelLower[2]{% one to one.
  \ifnum\lccode#1=#2\else
    \babel@savevariable{\lccode#1}%
    \lccode#1=#2\relax
  \fi}
\newcommand\BabelLowerMM[4]{% many-to-many
  \@tempcnta=#1\relax
  \@tempcntb=#4\relax
  \def\bbl@tempa{%\expandafter\bbl@tempa
    \ifnum\@tempcnta>#2\else
      \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
      \advance\@tempcnta#3\relax
      \advance\@tempcntb#3\relax
      \expandafter\bbl@tempa
    \fi}%
  \bbl@tempa}
\newcommand\BabelLowerMO[4]{% many-to-one
  \@tempcnta=#1\relax
  \def\bbl@tempa{%\expandafter\bbl@tempa
    \ifnum\@tempcnta>#2\else
      \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
      \advance\@tempcnta#3
      \expandafter\bbl@tempa
    \fi}%
  \bbl@tempa}

The following package options control the behavior of hyphenation mapping.
\DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
\DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\@ne}
\DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
\DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@@}
\DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}

Initial setup to provide a default behavior if hyphenmap is not set.
\AtEndOfPackage{%
  \ifx\bbl@opt@hyphenmap\undefined
    \bbl@xin@{},%\bbl@language@opts\%
    \chardef\bbl@opt@hyphenmap\ifin@4\else\@ne\fi
  \fi}

8.10 Macros common to a number of languages
\set@low@box
The following macro is used to lower quotes to the same level as the comma. It prepares its argument in box register 0.
\save@sf@q The macro \save@sf@q is used to save and reset the current space factor.

\begin{verbatim}
def\save@sf@q#1{\leavevmode
\begingroup
\edef\@SF{\spacefactor\the\spacefactor}#1\@SF
\endgroup}
\end{verbatim}

8.11 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.

8.11.1 Quotation marks

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

\begin{verbatim}
\ProvideTextCommand{\quotedblbase}{OT1}{%
\save@sf@q{\set@low@box{\textquotedblright}\%\box\z@\kern-.04em\bbl@allowhyphens}}
\ProvideTextCommandDefault{\quotedblbase}{%\UseTextSymbol{OT1}{\quotedblbase}}
\end{verbatim}

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

\texttt{\textbackslash quotesinglbase} We also need the single quote character at the baseline.

\begin{verbatim}
\ProvideTextCommand{\quotesinglbase}{OT1}{%
\save@sf@q{\set@low@box{\textquoteright}\%\box\z@\kern-.04em\bbl@allowhyphens}}
\ProvideTextCommandDefault{\quotesinglbase}{%\UseTextSymbol{OT1}{\quotesinglbase}}
\end{verbatim}

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

\texttt{\textbackslash guillemotleft} The guillemet characters are not available in OT1 encoding. They are faked.

\begin{verbatim}
\ProvideTextCommand{\guillemotleft}{OT1}{%
\ifmmode\ll\else\save@sf@q{\nobreak\raise.2ex\hbox{$\scriptscriptstyle\ll$}}\bbl@allowhyphens}\fi}
\ProvideTextCommand{\guillemotright}{OT1}{%
\ifmmode\gg\else\save@sf@q{\nobreak\raise.2ex\hbox{$\scriptscriptstyle\gg$}}\bbl@allowhyphens}\fi}
\ProvideTextCommandDefault{\guillemotleft}{%\UseTextSymbol{OT1}{\guillemotleft}}
\ProvideTextCommandDefault{\guillemotright}{%\UseTextSymbol{OT1}{\guillemotright}}
\end{verbatim}

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

88
The single guillemets are not available in OT1 encoding. They are faked.

\provideTextCommand{\guilsinglleft}{OT1}{%}
\ifmmode < \else \save@sf@q{\nobreak \raise.2ex\hbox{$\scriptscriptstyle <$}\bbl@allowhyphens}\fi}
\provideTextCommand{\guilsinglright}{OT1}{%}
\ifmmode > \else \save@sf@q{\nobreak \raise.2ex\hbox{$\scriptscriptstyle >$}\bbl@allowhyphens}\fi}

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

\provideTextCommandDefault{\guilsinglleft}{OT1}{\UseTextSymbol{OT1}{\guilsinglleft}}
\provideTextCommandDefault{\guilsinglright}{OT1}{\UseTextSymbol{OT1}{\guilsinglright}}

\section*{8.11.2 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 \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.

\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.irb.hr).

\def\crrtic@{\hrule height0.1ex width0.3em}
\def\crttic@{\hrule height0.1ex width0.33em}
\def\ddj@{\setbox0\hbox{d}\dimen@=.55\ht0
\advance\dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font
\leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}}
\def\DDJ@{\setbox0\hbox{D}\dimen@=.55\ht0
\advance\dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font
\leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}}
\advance\dimen@ii.15ex % correction for the dash position
\advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
\dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
\leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\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.
\ProvideTextCommandDefault{\dj}{\UseTextSymbol{OT1}{\dj}}
\ProvideTextCommandDefault{\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.
\DeclareTextCommand{\SS}{OT1}{SS}
\ProvideTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}

8.11.3 Shorthands for quotation marks

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

\glq The 'german' single quotes.
\grq
\ProvideTextCommandDefault{\glq}{%}
\textormath{\quotingsinglbase}{\mbox{\quotingsinglbase}}

The definition of \grq depends on the font encoding. With T1 encoding no extra kerning is
needed.
\ProvideTextCommand{\grq}{T1}{%}
\textormath{\textquoteleft}{\mbox{\textquoteleft}}
\ProvideTextCommand{\grq}{TU}{%}
\ProvideTextCommand{\grq}{OT1}{%}
\save@sf@q{\kern-.0125em
\textormath{\textquoteleft}{\mbox{\textquoteleft}}%\kern.07em\relax}}
\ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}

\glqq The 'german' double quotes.
\grqq
\ProvideTextCommandDefault{\glqq}{%}
\textormath{\quotedblbase}{\mbox{\quotedblbase}}

The definition of \grqq depends on the font encoding. With T1 encoding no extra kerning is
needed.
\ProvideTextCommand{\grqq}{T1}{%}
\textormath{\textquotedblleft}{\mbox{\textquotedblleft}}
\ProvideTextCommand{\grqq}{TU}{%}
\ProvideTextCommand{\grqq}{OT1}{%}
\save@sf@q{\kern-.07em
\textormath{\textquotedblleft}{\mbox{\textquotedblleft}}%\kern.07em\relax}}
\ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}
The ‘french’ single guillemets.

\flq \ProvideTextCommandDefault{\flq}{{}\textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
\frq \ProvideTextCommandDefault{\frq}{{}\textormath{\guilsinglright}{\mbox{\guilsinglright}}}

The ‘french’ double guillemets.

\flqq \ProvideTextCommandDefault{\flqq}{{}\textormath{\guillemotleft}{\mbox{\guillemotleft}}}
\frqq \ProvideTextCommandDefault{\frqq}{{}\textormath{\guillemotright}{\mbox{\guillemotright}}}

8.11.4 Umlauts and tremas

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

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

\def\umlauthigh{% 
\def\@umlauta##1{\leavevmode\bgroup% 
\expandafter\accent\csname f@encoding dqpos\endcsname##1\bbl@allowhyphens\egroup}%
\let\@umlaute@\@umlauta}
\def\umlautlow{% 
\def\@umlauta{\protect\lower@umlaut}
\def\@umlaute@\@umlauta}
\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 \texttt{(dimen)} register.
\lower@umlaut

\expandafter\ifx\csname U@D\endcsname\relax 
\csname newdimen\endcsname\U@D
\fi

The following code fools T\TeX's make\_accent procedure about the current x-height of the font to force another placement of the umlaut character. First we have to save the current x-height of the font, because we'll change this font dimension and this is always done globally. Then we compute the new x-height in such a way that the umlaut character is lowered to the base character. The value of .45ex depends on the METAFONT parameters with which the fonts were built. (Just try out, which value will look best.) If the new x-height is too low, it is not changed. Finally we call the \texttt{\texttt{\accent}} primitive, reset the old x-height and insert the base character in the argument.

\def\lower@umlaut#1{% 
\leavevmode\bgroup% 
{\setbox@z@hbox{#1}}
\expandafter\char\csname f@encoding dqpos\endcsname%
\dimen@-.45ex\advance\dimen@ht\z@ 
\ifdim1ex<\dimen@\fontdimen5\font\dimen@i% 
\expandafter\accent\csname f@encoding dqpos\endcsname
}
For all vowels we declare \" to be a composite command which uses \umlauta or \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 \umlauta and/or \umlaute for a language in the corresponding ldf (using the babel switching mechanism, of course).

\AtBeginDocument{%
\DeclareTextCompositeCommand{"}{OT1}{a}{\umlauta{a}}%
\DeclareTextCompositeCommand{"}{OT1}{e}{\umlaute{e}}%
\DeclareTextCompositeCommand{"}{OT1}{i}{\umlaute{i}}%
\DeclareTextCompositeCommand{"}{OT1}{\i}{\umlaute{i}}%
\DeclareTextCompositeCommand{"}{OT1}{o}{\umlauta{o}}%
\DeclareTextCompositeCommand{"}{OT1}{u}{\umlauta{u}}%
\DeclareTextCompositeCommand{"}{OT1}{A}{\umlauta{A}}%
\DeclareTextCompositeCommand{"}{OT1}{E}{\umlaute{E}}%
\DeclareTextCompositeCommand{"}{OT1}{I}{\umlaute{I}}%
\DeclareTextCompositeCommand{"}{OT1}{O}{\umlauta{O}}%
\DeclareTextCompositeCommand{"}{OT1}{U}{\umlauta{U}}%
}

Finally, the default is to use English as the main language.
\ifx\l@english\undefined
  \chardef\l@english\z@
\fi
\main@language{english}

8.12 Layout

Work in progress.
Layout is mainly intended to set bidi documents, but there is at least a tool useful in general.
\@ifundefined{#1}{}{%
  \bbl@exp{\let\bbl@ss@#1<#1>}%
  \namedef{#1}{%\ifstar{\bbl@presec@s{#1}}{%\@dblarg{\bbl@presec@x{#1}}}}%
\def\bbl@presec@x#1[#2]#3{%\select@language@x{\main@language}#1}%
\def\bbl@presec@s#1#2{%\select@language@x{\main@language}#1}%
\def\bbl@exp{%\select@language@x{\main@language}}
Now we load definition files for engines.
\input txtbabel.def \or \input luababel.def \or \input xebabel.def
\section{Creating languages}
\verb!\babelprovide! is a general purpose tool for creating languages. Currently it just creates the language infrastructure, but in the future it will be able to read data from ini files, as well as to create variants. Unlike the nil pseudo-language, captions are defined, but with a warning to invite the user to provide the real string.
Depending on whether or not the language exists, we define two macros.

\def\bbl@provide@new#1{%
  \@namedef{date#1}{}% marks lang exists - required by \StartBabelCommands
  \@namedef{extras#1}{}%
  \@namedef{noextras#1}{}%
  \StartBabelCommands*{#1}{captions}%
  \ifx\bbl@KVP@captions\@nil % and also if import, implicit
  \def\bbl@tempb##1{% elt for \bbl@captionslist
    \if\empty##1\else
      \SetString\##1{\bbl@exp{\bbl@nocaption{\bbl@stripslash##1}{\bbl@stripslash##1}}}%
      \expandafter\bbl@tempb
    \fi}
    \expandafter\bbl@tempb\bbl@captionslist\@empty
  \else
    \bbl@read@ini{\bbl@KVP@captions}% Here all letters cat = 11
  \bbl@after@ini
  \bbl@savestrings
  \StartBabelCommands*{#1}{date}%
  \if\empty\bbl@KVP@import\else
    \bbl@savetoday
    \bbl@savedate
  \fi
  \EndBabelCommands
  \expandafter\bbl@exp{\def\bbl@hyphenmins{\bbl@exp{\bbl@ifunset{bbl@lfthm@#1}{2}{\bbl@lfthm@#1}\bbl@ifunset{bbl@rgthm@#1}{3}{\bbl@rgthm@#1}}}}%
The `hyphenrules` option is handled with an auxiliary macro.

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

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

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

 Key-value pairs are treated differently depending on the section in the ini file. The following macros are the readers for identification and typography.
The `\languagename` date format may not be suitable for proper typesetting, and therefore it very likely will change in a future release. Reported

Now captions and captions.licr, depending on the engine. And also for dates. They rely on a few auxiliary macros.

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

But dates are more complex. The full date format is stores in date.gregorian, so we must read it in non-Unicode engines, too.
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 the date or in the month name.

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

\def\bbl@provide@lsys#1{%"}
The following ini reader ignores everything but the identification section. It is called when a font is defined (i.e., 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.

\def\bbl@ini@ids#1{%
\def\BabelBeforeIni##1##2{%
\begingroup
\bbl@add\bbl@secpost@identification{\closein1 }%
\catcode`\[=12 \catcode`\]=12 \catcode`\==12
\bbl@read@ini{##1}%
\endgroup}% boxed, to avoid extra spaces:
{\setbox\z@hbox{\InputIfFileExists{babel-#1.tex}{}{}}}

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

9.1 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.

\def\format{lplain}
\ifx\fmtname\format
\else
\def\format{LaTeX2e}
\ifx\fmtname\format
\else
\aftergroup\endinput
\fi
\fi

9.2 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 \TeXbook (Appendix D, page 382). The primitive \meaning applied to a token expands to the current meaning of this token. For example, ‘\meaning\A’ with \A defined as ‘\def\A{\B}’ expands to the characters ‘macro:\#1->\B’ with all category codes set to ‘other’ or ‘space’.
\newlabel\label

The macro \label writes a line with a \newlabel command into the .aux file to define labels.
1994 \bbl@redefine\newlabel\#1\#2{% 
1995 \@safe@activestrue\org@newlabel\#1\#2\@safe@activesfalse
\@newl@bel

We need to change the definition of the \LaTeX-internal macro \@newl@bel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

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

1996 ⟨⟨More package options⟩⟩ ≡
1997 \DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
1998 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
1999 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
2000 ⟨⟨More package options⟩⟩

First we open a new group to keep the changed setting of \protect local and then we set the \@safe@actives switch to true to make sure that any shorthand that appears in any of the arguments immediately expands to its non-active self.

2001 \bbl@trace{Cross referencing macros}
2002 \if\bbl@opt@safe\@emptyelse
2003 \def\@newl@bel\#1\#2\#3{%
2004 \@safe@activestrue
2005 \bbl@ifunset\#1\#2%
2006 \relax
2007 \{\gdef\@multiplelabels{%
2008 \@latex@warning@no@line{There were multiply-defined labels}}%
2009 \@latex@warning@no@line{Label `\#2' multiply defined}}%
2010 \global\@namedef{\#1\#2}\#3}
\@testdef

An internal \LaTeX macro used to test if the labels that have been written on the .aux file have changed. It is called by the \enddocument macro. This macro needs to be completely rewritten, using \meaning. The reason for this is that in some cases the expansion of \#1\#2 contains the same characters as the #3; but the character codes differ. Therefore \LaTeX keeps reporting that the labels may have changed.

2011 \CheckCommand*\@testdef[3]{%
2012 \def\reserved@a[#3]{%
2013 \expandafter\ifx\csname #1\#2\endcsname\reserved@a
2014 \else
2015 \@tempswatrue
2016 \fi}

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

2017 \def\@testdef\#1\#2\#3{% 
2018 \@safe@activestrue

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

2019 \expandafter\let\expandafter\bbl@tempa\csname #1\#2\endcsname

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

2020 \def\bbl@tempb[#3]{%
2021 \@safe@activesfalse

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

2022 \ifx\bbl@tempa\relax
2023 \else
2024 \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
2025 \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.

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

\ref
\pageref
The same holds for the macro \ref that references a label and \pageref to reference a page. So we redefine \ref and \pageref. While we change these macros, we make them robust as well (if they weren't already) to prevent problems if they should become expanded at the wrong moment.

\bbl@xin@{R}\bbl@opt@safe
\ifin@
\bbl@redefinerobust\ref#1{\
\@safe@activestrue\org@ref{#1}\@safe@activesfalse}
\bbl@redefinerobust\pageref#1{\
\@safe@activestrue\org@pageref{#1}\@safe@activesfalse}
\else
\let\org@ref\ref
\let\org@pageref\pageref
\fi

\@citex
The macro used to cite from a bibliography, \cite, uses an internal macro, \@citex. It is this internal macro that picks up the argument(s), so we redefine this internal macro and leave \cite alone. The first argument is used for typesetting, so the shorthands need only be deactivated in the second argument.

\bbl@xin@{B}\bbl@opt@safe
\ifin@
\bbl@redefine\@citex[#1]#2{\
\@safe@activestrue\edef\@tempa{#2}\@safe@activesfalse
\org@@citex[#1]{\@tempa}}%
\else
\AtBeginDocument{%
@ifpackageloaded(natbib){%
Notice that we use \def here instead of \bbl@redefine because \org@citex is already defined and we don't want to overwrite that definition (it would result in parameter stack overflow because of a circular definition). (Recent versions of natbib change dynamically \@citex, so PR4087 doesn't seem fixable in a simple way. Just load natbib before.)
\def\citex[#1][#2][#3]{\
\@safe@activestrue\edef\@tempa{\@tempa}\@safe@activesfalse
\org@citex[#1][#2]{\@tempa})%
}}
\AtBeginDocument{%
@ifpackageloaded(cite){%
The package cite has a definition of \@citex where the shorthands need to be turned off in both arguments.
\def\citex[#1][#2]{%
\nocite The macro \nocite which is used to instruct BiB\TeX to extract uncited references from the database.
\bbl@redefine
\nocite#1{%\@safe@activestrue\org@@citex[#1]{#2}\@safe@activesfalse}\
\bibcite The macro that is used in the .aux file to define citation labels. When packages such as natbib or cite are not loaded its second argument is used to typeset the citation label. In that case, this second argument can contain active characters but is used in an environment where \@safe@activestrue is in effect. This switch needs to be reset inside the \hbox which contains the citation label. In order to determine during .aux file processing which definition of \bibcite is needed we define \bibcite in such a way that it redefines itself with the proper definition.
\bbl@redefine\bibcite{%\@safe@activestrue\org@nocite{#1}\@safe@activesfalse}
\bibcite We call \bbl@cite@choice to select the proper definition for \bibcite. This new definition is then activated.
\bbl@cite@choice
\bibcite The macro \bbl@bibcite holds the definition of \bibcite needed when neither natbib nor cite is loaded.
\def\bbl@bibcite#1#2{%\org@bibcite{#1}\{\@safe@activestrue#2}}
\bbl@cite@choice The macro \bbl@cite@choice determines which definition of \bibcite is needed.
\def\bbl@cite@choice{% \def\bbl@cite@choice{% 
First we give \bibcite its default definition.
\global\let\bibcite\bbl@bibcite 
Then, when natbib is loaded we restore the original definition of \bibcite.
\@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{}
For cite we do the same.
\@ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{}
Make sure this only happens once.
\global\let\bbl@cite@choice\relax}
When a document is run for the first time, no .aux file is available, and \bibcite will not yet be properly defined. In this case, this has to happen before the document starts.
\AtBeginDocument{\bbl@cite@choice}
\bibitem One of the two internal BiB\TeX macros called by \bibitem that write the citation label on the .aux file.
\bbl@redefine\bibitem#1{%\@safe@activestrue\org@bibitem[#1]\{\@safe@activesfalse}\
\else
\let\org@nocite\nocite
\let\org@citex\@citex
\let\org@bibcite\bibcite
\let\org@bibitem\bibitem
\fi


9.3 Marks

\markright Because the output routine is asynchronous, we must pass the current language attribute to the headlines, 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@activestru is in effect.

\bbl@trace{Marks}
\IfBabelLayout{sectioning}
{\ifx\bbl@opt@headfoot@nil
  \g@addto@macro\resetactivechars{
    \set@typeset@protect
    \expandafter\select@language@x\expandafter{\bbl@main@language}
    \let\protect\noexpand
    \let\protect\noexpand}
  \fi}
{\bbl@redefine\markright#1{\
  \bbl@ifblank{#1}{\org@markright{}}\
  \toks@{#1}\
  \bbl@exp{\org@markright{\protect\foreignlanguage{\languagename}{\protect\bbl@restore@actives\the\toks}}}}

\markboth The definition of \markboth is equivalent to that of \markright, except that we need two token registers. The document classes 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\@mkboth\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{\languagename}{\protect\bbl@restore@actives#1}}
  \bbl@ifblank{#1}{\toks@{#1}}
  \bbl@expafter{\bbl@tempb{#1}}
  \bbl@ifblank{#2}{\@temptokena{}}
  \@temptokena\expandafter{\bbl@exp{\org@markboth{\the\toks}{\the\@temptokena}}}

and copy it to \@mkboth if necessary.
{\bbl@tempc} % end \IfBabelLayout

104
9.4 Preventing clashes with other packages

9.4.1 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
\ifin@
\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.

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

9.4.2 varioref

\@@vpageref \vrefpagenum \Ref

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

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

9.4.3 \texttt{hhline}

\begin{verbatim}
\hhline Delaying the activation of the shorthand characters has introduced a problem with the \texttt{hhline} package. The reason is that it uses the ‘:’ character which is made active by the french support in \texttt{babel}. Therefore we need to reload the package when the ‘:’ is an active character.
So at \texttt{\begin{document}} we check whether \texttt{hhline} is loaded.
\begin{verbatim}
\AtBeginDocument{%
\ifpackageloaded{hhline}%
  Then we check whether the expansion of \texttt{\normal@char:} is not equal to \texttt{\relax}.
\begin{verbatim}
\expandafter\ifx\csname\string:endcsname\endcsname\relax
  \else
  In that case we simply reload the package. Note that this happens after the category code of the @-sign has been changed to other, so we need to temporarily change it to letter again.
\begin{verbatim}
\makeatletter
\def\@currname{hhline}\input{hhline.sty}\makeatother
\fi%
\end{verbatim}
\fi}
\end{verbatim}
\end{verbatim}

9.4.4 \texttt{hyperref}

\pdfstringdefDisableCommands A number of interworking problems between \texttt{babel} and \texttt{hyperref} are tackled by \texttt{hyperref} itself. The following code was introduced to prevent some annoying warnings but it broke bookmarks. This was quickly fixed in \texttt{hyperref}, which essentially made it no-op. However, it will not removed for the moment because \texttt{hyperref} is expecting it.

\begin{verbatim}
\AtBeginDocument{%
\if\pdfstringdefDisableCommands\@undefined\else
\pdfstringdefDisableCommands{%\languageshorthands{system}}%
\fi}
\end{verbatim}
9.4.5 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{\textbackslash foreignlanguage} which babel adds to the marks can end up inside the argument of \texttt{\textbackslash MakeUppercase}. To prevent unexpected results we need to define \texttt{\textbackslash FOREIGNLANGUAGE} here.

\begin{verbatim}
\DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
  \lowercase{\texttt{\textbackslash foreignlanguage\{#1\}}}%
}
\end{verbatim}

\substitutefontfamily The command \texttt{\textbackslash substitutefontfamily} creates an .fd file on the fly. The first argument is an encoding mnemonic, the second and third arguments are font family names.

\begin{verbatim}
\def\substitutefontfamily#1#2#3{%
  \lowercase{\immediate\openout15=#1#2.fd\relax}%
  \immediate\write15{%
    \string\ProvidesFile{#1#2.fd}[
      \the\year/\two@digits{\the\month}/\two@digits{\the\day}]
    space generated font description file^^J
    \string\DeclareFontFamily{#1}{#2}{m}{n}{<->ssub * #3/m/n}{}%^J
    \string\DeclareFontFamily{#1}{#2}{m}{it}{<->ssub * #3/m/it}{}%^J
    \string\DeclareFontFamily{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{}%^J
    \string\DeclareFontFamily{#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}

9.5 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{\LaTeX} and \texttt{\LaTeX} for them using \texttt{\textbackslash ensureascii}. The default ASCII encoding is set, too (in reverse order): the “main” encoding (when the document begins), the last loaded, or OT1.

\begin{verbatim}
\ensureascii
\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.

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

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

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

\AtBeginDocument{%
@ifpackageloaded{fontspec}%
 {\xdef\latinencoding{%
 {\if\UTFencname@undefined
 EU@ifcase\bbl@engine\or2\or1\fi
 \else
 \UTFencname
 \fi}}%
 {\edef\latinencoding{OT1}}%
 {\if\cfencoding\bbl@t@one
 \edef\latinencoding{\bbl@t@one}{%}
 {\if\@ifl@aded{def}{t1enc}{\xdef\latinencoding{\bbl@t@one}}{}}}
{\fi}\
%\
%\
%\\\% Now we can define the command \latintext which is a declarative switch to a latin font-encoding. Usage of this macro is deprecated.

\DeclareRobustCommand{\latintext}{%
 {\fontencoding{\latinencoding}\selectfont
 {\def\encodingdefault{\latinencoding}}}
\textlatin This command takes an argument which is then typeset using the requested font encoding. In order to avoid many encoding switches it operates in a local scope.

\begin{verbatim}
2235 \ifx\@undefined\DeclareTextFontCommand
2236 \DeclareRobustCommand\textlatin{[1]{\leavevmode{\latintext #1}}}
2237 \else \DeclareTextFontCommand\textlatin{[1]{\latintext}} \fi
\end{verbatim}

9.6 Basic bidi support

Work in progress. This code is currently placed here for practical reasons. It is loosely based on rl babel 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 AR ABI (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 rl babel did), and by introducing a “middle layer” just below the user interface (sectioning, footnotes).

- `pdftex` provides a minimal support for bidi text, and it must be done by hand. Vertical typesetting is not possible.

- `xetex` is somewhat better, thanks to its font engine (even if not always reliable) and a few additional tools. However, very little is done at the paragraph level. Another challenging problem is text direction does not honour \TeX grouping.

- `luatex` can provide the most complete solution, as we can manipulate almost freely the node list, the generated lines, and so on, but bidi text does not work out of the box and some development is necessary. It also provides tools to properly set left-to-right and right-to-left page layouts. As Lua\TeX-ja shows, vertical typesetting is possible, too. Its main drawback is font handling is often considered to be less mature than xetex, mainly in Indic scripts (but there are steps to make HarfBuzz, the xetex font engine, available in luatex; see <https://github.com/tatzetwerk/luatex-harfbuzz>.

\begin{verbatim}
2240 \bbl@trace{Basic (internal) bidi support}
2241 \def\bbl@alscripts{\Arabic,\Syriac,\Thaana,}
2242 \def\bbl@rsscripts{\%
2243 ,\Imperial\Aramaic,\Avestan,\Cypriot,\Hatran,\Hebrew,\%
2244 \Old\Hungarian,\Old\Hungarian,\Lydian,\Mandaean,\Manichaean,\%
2245 \Manichaean,\Meroitic\Cursive,\Meroitic,\Old\North\Arabian,\%
2246 \Nabataean,\N’Ko,\Orkhon,\Palmyrene,\Inscriptional\Pahlavi,\%
2247 \Psalter\Pahlavi,\Phoenician,\Inscriptional\Parthian,\Samaritan,\%
2248 \Old\South\Arabian,\%}
2249 \def\bbl@provide@dirs#1{\%
2250 \bbl@xin\{\csname bbl@sname@#1\endcsname\}\{\bbl@alscripts\bbl@rsscripts\}%
2251 \ifin\%
2252 \global\bbl@csarg\chardef\wdir@#1\@ne\%
2253 \bbl@xin\{\csname bbl@sname@#1\endcsname\}\{\bbl@alscripts\%
2254 \ifin\%
2255 \global\bbl@csarg\chardef\wdir@#1\tw@ \% useless in xetex
2256 \fi
2257 \else \%
2258 \global\bbl@csarg\chardef\wdir@#1\tw@ \%
2259 \fi
2260 \def\bbl@switchdir{\%
\end{verbatim}
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).

A tool for weak L (mainly digits).

\loadlocalcfg

At some sites it may be necessary to add site-specific actions to a language definition file. This can be done by creating a file with the same name as the language definition file, but with the extension .cfg. For instance the file norsk.cfg will be loaded when the language definition file norsk.ldf is loaded.

For plain-based formats we don’t want to override the definition of \loadlocalcfg from plain.def.

\DeclareRobustCommand\babelsublr[1]{\leavevmode{\bbl@textdir\z@#1}}
Just to be compatible with \LaTeX{} 2.09 we add a few more lines of code:

\begin{verbatim}
\ifx\@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\ifvmode\nobreak\fi\fi}
\fi
\end{verbatim}

\section{Multiple languages (switch.def)}

Plain \TeX{} 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.

\begin{verbatim}
\ProvidesFile{switch.def}[\date{} \version{} Babel switching mechanism]
\ProvidesFile{switch.def}[\date{} \version{} Define core switching macros]
\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.}
\def\bbl@version{\version{}}
\def\bbl@date{\date{}}
\def\adddialect#1#2{%
  \global\ch@r@def\string#1\string#2\relax
  \bbl@usehooks{adddialect}{\string#1}{\string#2}%
  \wlog{\string#1 = a dialect from \string\language#2}}
\bbl@iflanguage{\texttt{\bbl@iflanguage} executes code only if the language \texttt{l@} exists. Otherwise raises and error. The argument of \texttt{\bbl@iflanguage} 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{\foreignlanguage} and the like appear in a \texttt{\MakeXXXcase{}}. However, a lowercase form is not imposed to improve backward compatibility (perhaps you defined a language named \texttt{MYLANG}, but unfortunately mixed case names cannot be trapped). Note \texttt{l@} is encapsulated, so that its case does not change.}
\end{verbatim}
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.

```
def\iflanguage{\protect\expandafter\csname selectlanguage \endcsname}
```

### 10.1 Selecting the language

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.

```
def\iflanguage{\protect\expandafter\csname selectlanguage \endcsname}
```

### 10.1 Selecting the language

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.

```
\let\bbl@select@type\z@
def\selectlanguage{\noexpand\protect\expandafter\csname selectlanguage \endcsname}
```

Because the command \selectlanguage could be used in a moving argument it expands to \protect\selectlanguage. Therefore, we have to make sure that a macro \protect exists. If it doesn't it is \let to \relax.

```
\ifx\undefined\protect\let\protect\relax\fi
```
As \LaTeX 2.09 writes to files expanded whereas \LaTeX 2e takes care not to expand the arguments of \texttt{\write} statements we need to be a bit clever about the way we add information to \texttt{.aux} files. Therefore we introduce the macro \texttt{\xstring} which should expand to the right amount of \texttt{\string}'s.

\begin{verbatim}
2424 \ifx\documentclass@undefined
2425 \def\xstring{\string\string\string}
2426 \else
2427 \let\xstring\string
2428 \fi
\end{verbatim}

Since version 3.5 babel writes entries to the auxiliary files in order to typeset table of contents etc. in the correct language environment.

\texttt{\bbl@pop@language} \begin{itemize}
\item \texttt{\bbl@pop@language} \begin{itemize}
\item When the language change happens inside a group the end of the group doesn’t write anything to the auxiliary files. Therefore we need \LaTeX's \texttt{aftergroup} mechanism to help us. The command \texttt{\aftergroup} stores the token immediately following it to be executed when the current group is closed. So we define a temporary control sequence \texttt{\bbl@pop@language} to be executed at the end of the group. It calls \texttt{\bbl@set@language} with the name of the current language as its argument.
\end{itemize}
\end{itemize}

\texttt{\bbl@language@stack} \begin{itemize}
\item 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 \texttt{\bbl@language@stack} and initially empty.
\end{itemize}

\begin{verbatim}
2429 \def\bbl@language@stack{}
\end{verbatim}

\begin{itemize}
\item When using a stack we need a mechanism to push an element on the stack and to retrieve the information afterwards.
\end{itemize}

\texttt{\bbl@push@language} \begin{itemize}
\item The stack is simply a list of languagenames, separated with a ‘+’ sign; the push function can be simple:
\end{itemize}

\begin{verbatim}
2430 \def\bbl@push@language{%
2431 \edef\bbl@language@stack{\languagename+\bbl@language@stack}
\end{verbatim}

\begin{itemize}
\item 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 \texttt{\languagename}. For this we first define a helper function.
\end{itemize}

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

\begin{verbatim}
2432 \def\bbl@pop@language\#1+\#2-\#3{%
2433 \edef\languagename{\#1}\def\#3{\#2}
\end{verbatim}

\begin{itemize}
\item 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 \texttt{\bbl@pop@language} is executed \LaTeX first expands the stack, stored in \texttt{\bbl@language@stack}. The result of that is that the argument string of \texttt{\bbl@pop@language} 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.
\end{itemize}

\begin{verbatim}
2434 \let\bbl@ifrestoring\@secondoftwo
2435 \edef\bbl@pop@language{%
2436 \expandafter\bbl@pop@language\bbl@language@stack-\bbl@language@stack
2437 \let\bbl@ifrestoring\@firstoftwo
2438 \expandafter\bbl@set@language\expandafter{\languagename}%
2439 \let\bbl@ifrestoring\@secondoftwo}
\end{verbatim}

\begin{itemize}
\item Once the name of the previous language is retrieved from the stack, it is fed to \texttt{\bbl@set@language} to do the actual work of switching everything that needs switching.
\end{itemize}
\bbl@set@language  The macro \bbl@set@language takes care of switching the language environment and of writing entries on the auxiliary files. For historical reasons, language names can be either language of \language. To catch either form a trick is used, but unfortunately as a side effect the catcodes of letters in \languagename are not well defined. The list of auxiliary files can be extended by redefining \BabelContentsFiles, but make sure they are loaded inside a group (as aux, toc, lof, and lot do) or the last language of the document will remain active afterwards.

We also write a command to change the current language in the auxiliary files.

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 \TeX{} in a certain pre-defined state.

The name of the language is stored in the control sequence \languagename.
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.
otherlanguage The otherlanguage environment can be used as an alternative to using the \selectlanguage declarative command. When you are typesetting a document which mixes left-to-right and right-to-left typesetting you have to use this environment in order to let things work as you expect them to.

The \ignorespaces command is necessary to hide the environment when it is entered in horizontal mode.

\long\def\otherlanguage#1{%\ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel=\thr@@\fi\csname selectlanguage \endcsname{#1}\ignorespaces}

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

\long\def\endotherlanguage{%\global\@ignoretrue\ignorespaces}

otherlanguage* The otherlanguage environment is meant to be used when a large part of text from a different language needs to be typeset, but without changing the translation of words such as ‘figure’. This environment makes use of \foreignlanguage.

\long\def\csname otherlanguage*\endcsname#1{%\ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel=4\relax\fi\foreignlanguage{#1}}

At the end of the environment we need to switch off the extra definitions. The grouping mechanism of the environment will take care of resetting the correct hyphenation rules and “extras”.

\expandafter\let\csname endotherlanguage*\endcsname\relax

\foreignlanguage The \foreignlanguage command is another substitute for the \selectlanguage command. This command takes two arguments, the first argument is the name of the language to use for typesetting the text specified in the second argument.

Unlike \selectlanguage this command doesn’t switch everything, it only switches the hyphenation rules and the extra definitions for the language specified. It does this within a group and assumes the \extras⟨lang⟩ command doesn’t make any \global changes. The coding is very similar to part of \selectlanguage.

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

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

(3.11) Also experimental are the hook foreign and foreign*. With them you can redefine \BabelText which by default does nothing. Its behavior is not well defined yet. So, use it in horizontal mode only if you do not want surprises.
In other words, at the beginning of a paragraph `\foreignlanguage` enters into mode with the surrounding lang, and with `\foreignlanguage*` with the new lang.

```latex
\providecommand\bbl@beforeforeign{}
\edef\foreignlanguage{
\noexpand\protect \expandafter\noexpand\csname foreignlanguage \endcsname}
\expandafter\def\csname foreignlanguage \endcsname{\@ifstar\bbl@foreign@s\bbl@foreign@x}
\def\bbl@foreign@x#1#2{\begingroup \let\BabelText\@firstofone \bbl@beforeforeign \foreign@language{#1}% Now in horizontal mode!
\bbl@usehooks{foreign}{} \BabelText{#2}%
\endgroup}
\def\bbl@foreign@s#1#2{% TODO - \shapemode, \@setpar, ?\@@par
\begingroup {\par} \let\BabelText\@firstofone \foreign@language{#1}%
\bbl@usehooks{foreign*}{} \bbl@dirparastext \BabelText{#2}% Still in vertical mode!
{\par}
\endgroup}
```

`\foreign@language` This macro does the work for `\foreignlanguage` and the other `language*` environment. First we need to store the name of the language and check that it is a known language. Then it just calls `\bbl@switch`.

```latex
\def\foreignlanguage@#1{% \edef\languagename{#1}% \bbl@fixname\languagename \bbl@iflanguage\languagename{% 
\expandafter\ifx\csname date\languagename\endcsname\relax \bbl@warning {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.\%
I'll proceed, but expect wrong results.\%
Reported}\fi
\fi \let\bbl@select@type\@ne \expandafter\bbl@switch\expandafter{\languagename}}
```

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

```latex
\let\bbl@hyphlist\@empty \let\bbl@hyphenation@\relax
```
The environment `hyphenrules` can be used to select \textit{just} the hyphenation rules. This environment does \textit{not} change \texttt{language} name and when the hyphenation rules specified were not loaded it has no effect. Note however, \texttt{lccode}'s and font encodings are not set at all, so in most cases you should use other `language*`.

\begin{verbatim}
\def\hyphenrules#1{\edef\bbl@tempf{#1}\bbl@fixname\bbl@tempf\bbl@iflanguage\bbl@tempf{\expandafter\patterns\expandafter{\bbl@tempf}\languageshorthands{none}\expandafter\ifx\csname\bbl@tempf hyphenmins\endcsname\relax\set@hyphenmins\tw@\thr@@\relax\else\expandafter\expandafter\expandafter\set@hyphenmins\csname\bbl@tempf hyphenmins\endcsname\relax\fi}}\endverbatim

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

\begin{verbatim}
\def\providehyphenmins#1#2{\expandafter\ifx\csname #1hyphenmins\endcsname\relax\@namedef{#1hyphenmins}{#2}\fi}
\end{verbatim}

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

\begin{verbatim}
\def\set@hyphenmins{\lefthyphenmin\#1\righthyphenmin\#2}\end{verbatim}
\ProvidesLanguage  The identification code for each file is something that was introduced in \LaTeX2e. When the command \ProvidesFile does not exist, a dummy definition is provided temporarily. For use in the language definition file the command \ProvidesLanguage is defined by babel. Depending on the format, ie, on if the former is defined, we use a similar definition or not.

\ifx\ProvidesFile\@undefined
\def\ProvidesLanguage#1[#2 #3 #4]{%
  \wlog{Language: #1 #4 #3 <#2>}%
}
\else
\def\ProvidesLanguage#1{%
  \begingroup
  \catcode\ 10 
  \@makeother\/
  \ifnextchar\[%

\def\@provideslanguage#1#2\]{{@provideslanguage{#1}{#2}}}
\def\@provideslanguage#1[ ]{% 
  \wlog{Language: #1 #2}%
  \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}
  \endgroup}
\fi
%

\LdfInit This macro is defined in two versions. The first version is to be part of the ‘kernel’ of babel, ie. the part that is loaded in the format; the second version is defined in babel.def. The version in the format just checks the category code of the ampersand and then loads babel.def. The category code of the ampersand is restored and the macro calls itself again with the new definition from babel.def.

\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, \babel@beginsave, is not considered to be undefined.

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

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

\providecommand\setlocale{%
  \bbl@error
  \{Not yet available}\%
  \{Find an armchair, sit down and wait\}
  \let\uselocale\setlocale
  \let\locale\setlocale
  \let\selectlocale\setlocale
  \let\selectlanguage\setlocale
  \let\textlocale\setlocale
  \let\textlanguage\setlocale
  \let\languagetext\setlocale
10.2 Errors

The babel package will signal an error when a document 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.

When the package was loaded without options not everything will work as expected. An error message is issued in that case. When the format knows about \PackageError it must be \TeX, so we can safely use its error handling interface. Otherwise we’ll have to ‘keep it simple’.

\def\nulllanguage{\string\language=0}
\if\PackageError\undefined
\def\error#1#2{\begingroup\newlinechar=`\^^J\def\{\MessageBreak\errhelp{#2}\errmessage{\#1}\endgroup}
\def\warning#1{\begingroup\newlinechar=`\^^J\def\{\MessageBreak\message{\#1}\endgroup}
\def\info#1{\begingroup\newlinechar=`\^^J\def\{\MessageBreak\wlog{\#1}\endgroup}
\else
\def\error#1#2{\begingroup\def\{\PackageError{ babel}{\#1}{\#2}\endgroup}
\def\warning#1{\begingroup\def\{\PackageWarning{ babel}{\#1}\endgroup}
\def\info#1{\begingroup\def\{\PackageInfo{ babel}{\#1}\endgroup}
\fi
\ifpackagewith{ babel}{silent}
\let\info\@gobble\let\warning\@gobble
\def\nocaption#1#2{% 1: text to be printed 2: caption macro \langXname
\gdef#2{\textbf{?#1?}}\bbl@warning{\string#2 not set. Please, define\%

121
it in the preamble with something like:
\%\string\renewcommand\string#2{..}\%\%
\ Reporte})
\@nolanerr#1{%
\bbl@error
{You haven’t defined the language #1 \space yet}%
{Your command will be ignored, type <return> to proceed})
\@nopatterns#1{%
\bbl@warning
{No hyphenation patterns were preloaded for the language `#1' into the format.\%
Please, configure your TeX system to add them and\%
rebuild the format. Now I will use the patterns\%
preloaded for \bbl@nulllanguage\space instead})
\let\bbl@usehooks\@gobbletwo
⟨/kernel⟩
⟨∗patterns⟩

11 Loading hyphenation patterns

The following code is meant to be read by init\TeX{} because it should instruct \TeX{} to read
hyphenation patterns. To this end the docstrip option patterns can be used to include
this code in the file hyphen.cfg. Code is written with lower level macros.
toks8 stores info to be shown when the program is run.
We want to add a message to the message \TeX{} 2.09 puts in the \everyjob register. This
could be done by the following code:

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

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

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

- Plain \TeX{} does not use the \everyjob register so the message would not be displayed.

To circumvent this a ‘dirty trick’ can be used. As this code is only processed when creating
a new format file there is one command that is sure to be used, \dump. Therefore the
original \dump is saved in \org@dump and a new definition is supplied.
To make sure that \TeX{} 2.09 executes the \@begindocumenthook we would want to alter
\begin{document}, but as this done too often already, we add the new code at the front of
\@preamble. But we can only do that after it has been defined, so we add this piece of
code to \dump.
This new definition starts by adding an instruction to write a message on the terminal and
in the transcript file to inform the user of the preloaded hyphenation patterns.
Then everything is restored to the old situation and the format is dumped.
2727 ⟨⟨Make sure ProvidesFile is defined⟩⟩
2728 \ProvidesFile{hyphen.cfg}[⟨⟨date⟩⟩ ⟨⟨version⟩⟩ Babel hyphens]
2729 \xdef\bbl@format{\jobname}
2730 \if\AtBeginDocument\@undefined
2731 \def\@empty{}
2732 \let\orig@dump\dump
2733 \def\dump{%
2734 \if\@ztryfc\@undefined
2735 \else
2736 \toks0=\expandafter{\@preamblecmds}%
2737 \edef\@preamblecmds{\noexpand@begindocumenthook\the\toks0}%
2738 \def\@begindocumenthook{}%
2739 \fi
2740 \let\dump\orig@dump\let\orig@dump\@undefined\dump%
2741 \fi
2742 ⟨⟨Define core switching macros⟩⟩
2743 \toks8{Babel <<@version@>> and hyphenation patterns for }%
2744 \process@line Each line in the file language.dat is processed by \process@line after it is read. The first thing this macro does is to check whether the line starts with =. When the first token of a line is an =, the macro \process@synonym is called; otherwise the macro \process@language will continue.
2745 \def\process@line#1#2 #3 #4 {%
2746 \if=#1%
2747 \process@synonym{#2}%
2748 \else
2749 \process@language{#1#2}{#3}{#4}%
2750 \fi
2751 \ignorespaces}
2752 \process@synonym This macro takes care of the lines which start with an =. It needs an empty token register to begin with. \bbl@languages is also set to empty.
2753 \toks@{}
2754 \def\bbl@languages{}
2755 \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.
2756 \def\process@synonym#1{%
2757 \ifnum\last@language=\m@ne
2758 \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
2759 \else
2760 \expandafter\chardef\csname l@#1\endcsname\last@language
2761 \wlog{\string\l@#1=\string\language\the\last@language}%
2762 \expandafter\let\csname #1hyphenmins\expandafter\endcsname
2763 \csname\languagename hyphenmins\endcsname
2764 \let\bbl@elt\relax
2765 \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}{}}%
2766 \fi
2767 \process@language The macro \process@language is used to process a non-empty line from the ‘configuration file’. It has three arguments, each delimited by white space. The first argument is the ‘name’ of a language; the second is the name of the file that contains the patterns. The optional third argument is the name of a file containing hyphenation exceptions.
123
The first thing to do is call \addlanguage to allocate a pattern register and to make that register ‘active’. Then the ‘name’ of the language that will be loaded now is added to the token register \toks8. and finally the pattern file is read.

For some hyphenation patterns it is needed to load them with a specific font encoding selected. This can be specified in the file language.dat by adding for instance ‘:T1’ to the name of the language. The macro \bbl@get@enc extracts the font encoding from the language name and stores it in \bbl@hyph@enc. The latter can be used in hyphenation files if you need to set a behavior depending on the given encoding (it is set to empty if no encoding is given).

Pattern files may contain assignments to \lefthyphenmin and \righthyphenmin. \TeX{} does not keep track of these assignments. Therefore we try to detect such assignments and store them in the \langle lang\rangle hyphenmins macro. When no assignments were made we provide a default setting.

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

Then we globally store the settings of \lefthyphenmin and \righthyphenmin and close the group.

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

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

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

\begin{verbatim}
def\process@language#1#2#3{% 
\expandafter\addlanguage\csname l@#1\endcsname 
\expandafter\language\csname l@#1\endcsname 
\edef\languagename{#1}\
\bbl@hook@everylanguage{#1}\
\bbl@get@enc#1::\@@@ 
\begingroup 
\lefthyphenmin\m@ne 
\bbl@hook@loadpatterns{#2}\
\ifnum\lefthyphenmin=\m@ne 
\else 
  \expandafter\ifdef\csname #1hyphenmins\endcsname{% 
    \the\lefthyphenmin\the\righthyphenmin}% 
\fi 
\endgroup 
\def\bbl@tempa{#3}% 
\ifx\bbl@tempa\@empty\else 
  \bbl@hook@loadexceptions{#3}\
\fi 
\let\bbl@elt\relax 
\edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}\
\ifnum\the\language=\z@ 
  \ifdef\csname #1hyphenmins\endcsname{% 
    \set@hyphenmins\tw@\thr@@\relax 
  \else 
    \expandafter\ifdef\csname #1hyphenmins\endcsname{relax} 
  \set@hyphenmins\tw@\thr@@\relax 
\else 
  \expandafter\ifdef\csname #1hyphenmins\endcsname 
\csname #1hyphenmins\endcsname 
\fi 
\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.

\begin{verbatim}
def \bbl@get@enc#1:#2:#3@@@{\def \bbl@hyph@enc{#2}}
\end{verbatim}

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

\begin{verbatim}
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
\else
\input luababel.def
\input babel-\bbl@format.cfg
\if\ifeof{ babel-\bbl@format.cfg}\message{I couldn’t find the file language.dat, I will try the file hyphen.tex}\input hyphen.tex\relax\chardef \l@english\z@\else\input babel-\bbl@format.cfg\relax\fi\closein1\endgroup
\end{verbatim}

The configuration file can now be opened for reading.

\begin{verbatim}
\openin1 = language.dat
\see if the file exists, if not, use the default hyphenation file hyphen.tex. The user will be informed about this.
def \languagename(english)\%\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\z@\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
\end{verbatim}
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.

\endlinechar\@ne
\read1 to \bbl@line
\endlinechar`\^^M

If the file has reached its end, exit from the loop here. If not, empty lines are skipped. Add 3 space characters to the end of \bbl@line. This is needed to be able to recognize the arguments of \process@line later on. The default language should be the very first one.

\if T\ifeof1F\fi T\relax
\ifx\bbl@line\@empty\else
\edef\bbl@line{\bbl@line\space\space\space}%
\expandafter\process@line\bbl@line\relax
\fi
\repeat

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

\begingroup
\def\bbl@elt#1#2#3#4{% 
\global\language=#2\relax 
\gdef\languagename{#1}%
\def\bbl@elt##1##2##3##4{}}%
\bbl@languages
\endgroup

and close the configuration file.

\closein1

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

\if/\the\toks@/\else
\errhelp{language.dat loads no language, only synonyms}
\errmessage{Orphan language synonym}
\fi
\advance\last@language\@ne
\edef\bbl@tempa{%
\everyjob{%
\ifx\typeout\@undefined
\immediate\write16{}
\else
\noexpand\typeout
\fi
\fi}
\edef\last@language\@ne
\edef\bbl@tempa{%
\everyjob{%
\ifx\typeout\@undefined
\immediate\write16{}
\else
\noexpand\typeout
\fi
\fi}
\edef\last@language\@ne
\edef\bbl@tempa{%
\everyjob{%
\ifx\typeout\@undefined
\immediate\write16{}
\else
\noexpand\typeout
\fi
\fi}
\edef\last@language\@ne

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.

\let\bbl@line\@undefined
\let\process@line\@undefined
\let\process@synonym\@undefined
Here the code for init\TeX{} ends.

## 12 Font handling with fontspec

Add the bidi handler just before luaofload, which is loaded by default by LaTeX. Just in case, consider the possibility it has not been loaded. First, a couple of definitions related to bidi [misplaced].

With explicit languages, we could define the font at once, but we don't. Just wait and see if the language is actually activated.
\def\babelfont[2][]{% 1=langs/scripts 2=fam
\edef\bbl@tempa{#1}%
\def\bbl@tempb{#2}%
\ifx\fontspec\@undefined
\usepackage{fontspec}%
\fi
\EnableBabelHook{babel-fontspec}%
\bbl@bblfont}
\newcommand\bbl@bblfont[2][]{% 1=features 2=fontname
\if\isundefined{\bbl@tempb family}{\bbl@providefam\bbl@tempb}{}}%
\if\isundefined{bbl@lsys\languagename}{\bbl@provide@lsys\languagename}{}}%
\expandafter\bbl@ifblank\expandafter{\bbl@tempa}{}{% ie bbl@rmdflt@lang / *scrt
\bbl@csarg\edef\bbl@tempb dflt@{<>{#1}{#2}}% save bbl@rmdflt@
\bbl@exp{% 1=T - language, already defined
\def\bbl@tempa{}}%
\bbl@foreach\bbl@font@fams{% don't gather with prev for
\bbl@if\isundefined{\bbl@##1dflt@\languagename}{\bbl@cs{famrst@\##1}}%
\global\let\bbl@##1dflt@\relax}%
\bbl@foreach\bbl@font@fams{% order is relevant
\bbl@add\originalTeX{}}% 128
}
\if\isundefined{\bbl@##1dflt@\languagename}{\bbl@csarg\edef\bbl@##1dflt@\languagename{<>{#1}{#2}}{}}%
\bbl@foreach\bbl@font@fams{% 3=T - from generic
\global\let\bbl@##1dflt@\relax}%
\bbl@foreach\bbl@font@fams{% 2=T - from script
\global\let\bbl@##1dflt@\bbl@##1dflt@{}}%
\bbl@foreach\bbl@font@fams{% 1=T - language, already defined
\def\bbl@tempa{}}%
\bbl@warning{The current font is not a standard family:\%
\fontname\font\%
Script and Language are not applied. Consider defining a\%
ew family with \string\babelfont. Reported}%
\bbl@foreach\bbl@font@fams{% don’t gather with prev for
\bbl@cs{famrst@\##1}%
\global\bbl@csarg\let\famrst@\##1\relax}%
\bbl@exp{order is relevant
\bbl@add\originalTeX{}}%
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.

The default font families. They are eurocentric, but the list can be expanded easily with \babelfont.

The old tentative way. Short and preverved for compatibility, but deprecated. Note there is no direct alternative for \babelfont. The reason in explained in the user guide, but essentially – that was not the way to go :-).
13 Hooks for XeTeX and LuaTeX

13.1 XeTeX

Unfortunately, the current encoding cannot be retrieved and therefore it is reset always to utf8, which seems a sensible default. XeLaTeX 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 xelatex.ini, so here we just “undo” some of the changes done by XeLaTeX. Anyway, for consistency LuaLaTeX also resets the catcodes.

⟨⟨∗ Restore Unicode catcodes before loading patterns ⟩⟩ ≡

\begingroup
  \catcode`@=11 \count@=128
  \loop\ifnum\count@<192
    \uccode\count@=0 \lccode\count@=0
    \catcode\count@=12 \sfcode\count@=1000
    \advance\count@ by 1 \repeat

% Letter:
  \def\L ##1 ##2 ##3 {\catcode`##1=11
    \uccode`##1=##2 \lccode`##1=##3
    \ifnum##1=##3 \else \sfcode`##1=999 \fi}

% Letter without case mappings:
  \def\l ##1 {\L ##1 ##1 ##1 }

\l 00AA \L 00B5 039C 00B5 \l 00BA \O 00D7 \l 00DF
Some more common code.

```
\documentclass{article}
\usepackage{amsmath}
\usepackage{amsfonts}
\usepackage{amssymb}
\usepackage{bbm}
\usepackage{courier}
\usepackage{times}
\usepackage{amsthm}
\usepackage{enumerate}
\usepackage{hyperref}
\usepackage{url}
\usepackage{color}
\usepackage{verbatim}
\usepackage{enumitem}
\usepackage{tikz}
\usepackage{pgfplots}
\usepackage{pgfkeys}
\usepackage{pgffor}
\usepackage{pgfmath}
\usepackage{pgfplotstable}
\usepackage{array}
\usepackage{multirow}
\usepackage{multicol}
\usepackage{multibib}
\usepackage{natbib}
\usepackage{hyperref}
\usepackage{url}
\usepackage{color}
\usepackage{verbatim}
\usepackage{enumerate}
\usepackage{amssymb}
\usepackage{amsmath}
\usepackage{amsfonts}
\usepackage{bbm}
\usepackage{courier}
\usepackage{times}
\usepackage{amsthm}
\usepackage{enumitem}
\usepackage{tikz}
\usepackage{pgfplots}
\usepackage{pgfkeys}
\usepackage{pgffor}
\usepackage{pgfmath}
\usepackage{pgfplotstable}
\usepackage{array}
\usepackage{multirow}
\usepackage{multicol}
\usepackage{multibib}
\usepackage{natbib}
\usepackage{hyperref}
\usepackage{url}
\usepackage{color}
\usepackage{verbatim}

\begin{document}

\section{Footnotes}

Footnotes are used to provide additional information or references that are not suitable for inclusion in the main text. They are commonly used for citations, clarifications, or supplementary details.

Footnote changes:

```
\def\bbl@footnote#1#2#3{% 
  \ifnextchar[% 
    {\bbl@footnote@o{#1}{#2}{#3}}% 
    {\bbl@footnote@x{#1}{#2}{#3}}% 
  \bgroup 
  \select@language@x{\bbl@main@language}% 
  \bbl@fn@footnote{#2#1{\ignorespaces#3}}% 
  \egroup% 
}\def\bbl@footnote@o#1#2#3[#4]{% 
  \bgroup 
  \select@language@x{\bbl@main@language}% 
  \bbl@fn@footnote[#4]{#2#1{\ignorespaces#3}}% 
  \egroup% 
}\def\bbl@footnotetext#1#2#3{% 
  \ifnextchar[% 
    {\bbl@footnotetext@o{#1}{#2}{#3}}% 
    {\bbl@footnotetext@x{#1}{#2}{#3}}% 
  \bgroup 
  \select@language@x{\bbl@main@language}% 
  \bbl@fn@footnotetext{#2#1{\ignorespaces#3}}% 
  \egroup% 
}\def\bbl@footnotetext@o#1#2#3[#4]{% 
  \bgroup 
  \select@language@x{\bbl@main@language}% 
  \bbl@fn@footnotetext[#4]{#2#1{\ignorespaces#3}}% 
  \egroup% 
}\def\BabelFootnote#1#2#3#4{ 
  \ifx\bbl@fn@footnote\@undefined \let\bbl@fn@footnote\footnote \fi 
  \ifx\bbl@fn@footnotetext\@undefined \let\bbl@fn@footnotetext\footnotetext \fi 
  \bbl@ifblank{#2}{ 
    \def#1{\bbl@footnote{\@firstofone}{#3}{#4}}% 
    \@namedef{\bbl@stripslash#1text}{\bbl@footnotetext{\@firstofone}{#3}{#4}}\fi 
  \fi 
  \ifx\bbl@fn@footnotetext\undefined \let\bbl@fn@footnotetext\footnotetext \fi 
  \ifx\bbl@fn@footnotetext\undefined \def#1{\bbl@exp{\\bbl@footnote{\\foreignlanguage{#2}}}\\#3\#4} \@namedef{\bbl@stripslash#1text}{\bbl@exp{\\bbl@footnotetext{\\foreignlanguage{#2}}}\\#3\#4} \fi

\end{document}
```

Now, the code.

```
xetex
\def\BabelStringsDefault{unicode}
\let\xebbl@stop\relax
\if\bbl@beforeforeign\leavevmode
  \input #1\relax
\fi

⟨⟨\textit{\textbackslash O} \textit{00F7} \textit{\textbackslash L} \textit{00FF} \textit{0178} \textit{00FF} \textit{\textbackslash endgroup} \textit{\textbackslash input \#1\relax} \textit{\langle\langle \textbackslash Restore \text{Unicode catcodes before loading patterns}\rangle\rangle}⟩⟩

\begin{verbatim}
Some more common code.

\section{Footnotes}

Footnotes are used to provide additional information or references that are not suitable for inclusion in the main text. They are commonly used for citations, clarifications, or supplementary details.

Footnote changes:

```
\def\bbl@footnote#1#2#3{% 
  \ifnextchar[% 
    {\bbl@footnote@o{#1}{#2}{#3}}% 
    {\bbl@footnote@x{#1}{#2}{#3}}% 
  \bgroup 
  \select@language@x{\bbl@main@language}% 
  \bbl@fn@footnote{#2#1{\ignorespaces#3}}% 
  \egroup% 
}\def\bbl@footnote@o#1#2#3[#4]{% 
  \bgroup 
  \select@language@x{\bbl@main@language}% 
  \bbl@fn@footnote[#4]{#2#1{\ignorespaces#3}}% 
  \egroup% 
}\def\bbl@footnotetext#1#2#3{% 
  \ifnextchar[% 
    {\bbl@footnotetext@o{#1}{#2}{#3}}% 
    {\bbl@footnotetext@x{#1}{#2}{#3}}% 
  \bgroup 
  \select@language@x{\bbl@main@language}% 
  \bbl@fn@footnotetext{#2#1{\ignorespaces#3}}% 
  \egroup% 
}\def\bbl@footnotetext@o#1#2#3[#4]{% 
  \bgroup 
  \select@language@x{\bbl@main@language}% 
  \bbl@fn@footnotetext[#4]{#2#1{\ignorespaces#3}}% 
  \egroup% 
}\def\BabelFootnote#1#2#3#4{ 
  \ifx\bbl@fn@footnote\@undefined \let\bbl@fn@footnote\footnote \fi 
  \ifx\bbl@fn@footnotetext\@undefined \let\bbl@fn@footnotetext\footnotetext \fi 
  \bbl@ifblank{#2}{ 
    \def#1{\bbl@footnote{\@firstofone}{#3}{#4}}% 
    \@namedef{\bbl@stripslash#1text}{\bbl@footnotetext{\@firstofone}{#3}{#4}}\fi 
  \fi 
  \ifx\bbl@fn@footnotetext\undefined \let\bbl@fn@footnotetext\footnotetext \fi 
  \ifx\bbl@fn@footnotetext\undefined \def#1{\bbl@exp{\\bbl@footnote{\\foreignlanguage{#2}}}\\#3\#4} \@namedef{\bbl@stripslash#1text}{\bbl@exp{\\bbl@footnotetext{\\foreignlanguage{#2}}}\\#3\#4} \fi

\end{verbatim}
```

Now, the code.

```
xetex
\def\BabelStringsDefault{unicode}
\let\xebbl@stop\relax
```
3115 \AddBabelHook{xetex}{encodedcommands}{% 
3116 \def\bbl@tempa{#1}% 
3117 \ifx\bbl@tempa\empty 
3118 \TeXInputencoding"bytes"% 
3119 \else 
3120 \TeXInputencoding"#1"% 
3121 \fi 
3122 \def\xebbl@stop{\TeXInputencoding"utf8"}
3123 \AddBabelHook{xetex}{stopcommands}{% 
3124 \xebbl@stop 
3125 \let\xebbl@stop\relax} 
3126 \AddBabelHook{xetex}{loadkernel}{% 
3127 ⟨⟨ Restore Unicode catcodes before loading patterns ⟩⟩ 
3128 \ifx\DisableBabelHook\@undefined\endinput\fi 
3129 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont} 
3130 \DisableBabelHook{babel-fontspec} 
3131 ⟨⟨ Font selection ⟩⟩ 
3132 \input txtbabel.def 
3133 ⟨/xetex⟩
3134 ⟨∗texxet⟩
3135 \bbl@trace{Redefinitions for bidi layout} 
3136 \def\bbl@sspre@caption{% 
3137 \bbl@exp\everyhbox{\\bbl@textdir\bbl@cs{wdir@\bbl@main@language}}}
3138 \ifx\bbl@opt@layout\@nnil\endinput\fi % No layout 
3139 \def\bbl@startskip{%\ifcase\bbl@thepardir\leftskip\else\rightskip\fi} 
3140 \def\bbl@endskip{%\ifcase\bbl@thepardir\rightskip\else\leftskip\fi} 
3141 \ifx\bbl@beforeforeign\leavevmode\% A poor test for bidi= 
3142 \def\@hangfrom#1{% 
3143 \setbox\@tempboxa\hbox{{#1}}% 
3144 \hangindent\ifcase\bbl@thepardir\wd@\tempboxa\else-\wd@\tempboxa\fi 
3145 \noindent\box\@tempboxa} 
3146 \def\raggedright{% 
3147 \let\@@centercr\@centercr 
3148 \bbl@startskip\z@skip 
3149 \@rightskip\@flushglue 
3150 \bbl@endskip\@rightskip 
3151 \parindent\z@ 
3152 \parfillskip\bbl@startskip} 
3153 \def\raggedleft{% 
3154 \let\@@centercr\@centercr 
3155 \bbl@startskip\@flushglue

13.2 Layout

In progress.

Unfortunately, proper support for xetex requires patching somehow lots of macros and packages (and some issues related to \specials remain, like color and hyperlinks). At least at this stage, babel will not do it and therefore a package like bidi (by Vafa Khalighi) would be necessary to overcome the limitations of xetex. Any help in making babel and bidi collaborate will be welcome, although the underlying concepts in both packages seem very different. Note also elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titleps, and geometry. \bbl@startskip and \bbl@endskip are available to package authors. Thanks to the \TeX expansion mechanism the following constructs are valid: \adim\bbl@startskip, \advance\bbl@startskip\adim, \bbl@startskip\adim.

Consider txtbabel as a shorthand for tex–xet babel, which is the bidi model in both pdftex and xetex.
 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.

13.3 LuaTeX

The new loader for luatex is based solely on language.dat, which is read on the fly. The code shouldn’t be executed when the format is build, so we check if \AddBabelHook is defined. Then comes a modified version of the loader in hyphen.cfg (without the hyphenmins stuff, which is under the direct control of babel).

The names \l@<language> are defined and take some value from the beginning because all ldf files assume this for the corresponding language to be considered valid, but patterns are not loaded (except the first one). This is done later, when the language is first selected (which usually means when the ldf finishes). If a language has been loaded, \bbl@hyphendata@<num> exists (with the names of the files read).

The default setup preloads the first language into the format. This is intended mainly for ‘english’, so that it’s available without further intervention from the user. To avoid duplicating it, the following rule applies: if the “0th” language and the first language in language.dat have the same name then just ignore the latter. If there are new synonymous, 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 have been preloaded into the format. This is not optimal, but it shouldn’t happen very often – with luatex patterns are best loaded when the document is typeset, and the “0th” language is preloaded just for backwards compatibility.

As of 1.1b, luatex 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 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 have been preloaded into the format. This is not optimal, but it shouldn’t happen very often – with luatex patterns are best loaded when the document is typeset, and the “0th” language is preloaded just for backwards compatibility.

As of 1.1b, luatex 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 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 have been preloaded into the format. This is not optimal, but it shouldn’t happen very often – with luatex patterns are best loaded when the document is typeset, and the “0th” language is preloaded just for backwards compatibility.

As of 1.1b, luatex 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 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 have been preloaded into the format. This is not optimal, but it shouldn’t happen very often – with luatex patterns are best loaded when the document is typeset, and the “0th” language is preloaded just for backwards compatibility.

As of 1.1b, luatex 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 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 have been preloaded into the format. This is not optimal, but it shouldn’t happen very often – with luatex patterns are best loaded when the document is typeset, and the “0th” language is preloaded just for backwards compatibility.
\begin{verbatim}
\ifx\AddBabelHook\undefined
\bbl@trace{Read language.dat}
\begingroup
\toks@{}
\count@\z@ % 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>1\one
    \bbl@info{Non-standard hyphenation setup}%
  \fi
  \let\bbl@manylang\relax}
\def\bbl@process@language#1#2#3{%
  \ifcase\count@
    \or
    \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}}%
  \else
    \the\toks@
  \fi
  \def\bbl@process@synonym@aux#1#2{%}
  \def\bbl@process@synonym#1{%
    \ifcase\count@
      \toks@\expandafter{\the\toks@\relax}\bbl@process@synonym{#1}{}
      \or
      \@ifundefined{zth@#1}{\bbl@process@synonym@aux{#1}{0}}{}
    \else
      \bbl@process@synonym@aux{#1}{\the\bbl@last}{}
    \fi
  \fi
t\def\bbl@languages@format\bbl@languages
  \def\bbl@elt#1#2#3#4{% Remove all except language 0
    \ifnum#2>1\z@\else
      \global\let\bbl@languages@format\bbl@languages
      \def\bbl@elt#1#2#3#4{% Remove all except language 0
    \fi
    \global\let\bbl@languages\bbl@elt
    \global\let\bbl@languages@format\bbl@languages
    \def\bbl@elt#1#2#3#4{% Remove all except language 0
      \ifnum#2>1\z@\else
    \fi
\end{verbatim}
\csname l@#1:f@encoding\endcsname
\edef\bbl@tempa{#1:f@encoding}\
\fi\relax
\@namedef{lu@texhyphen@loaded@\the\language}{}
\@ifundefined{bbl@hyphendata@\the\language}{}
\ifnum\csname l@\bbl@tempa\endcsname=spanish, dutch:OT1...
\def\bbl@tempb{\ifx\bbl@tempb@empty\else if not a synonymous
\if\csname l@\bbl@tempa\endcsname=\empty\else % if not a synonymous
\fi
\bbl@csarg\xdef{hyphendata@\the\language}{\bbl@tempc}\
\fi}
\@ifundefined{bbl@hyphendata@\the\language}{}
{\bbl@info{No hyphenation patterns were set for language \bbl@tempa. Reported}}%
{\expandafter\expandafter\expandafter\bbl@luapatterns
\csname bbl@hyphendata@\the\language\endcsname}}{}
\endinput\fi
\begingroup
\catcode`%=12
\catcode`\'=12
\catcode`\%=12
\catcode`\#:12
\directlua{
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%?%.', '%%.') .. '%d?'
    ss = ss:gsub('%.%d%?', '%%.')
    pats, n = pats:gsub('%s '.. ss .. '%s', '%s' .. p .. ' ')
  end
end

138
if n == 0 then
tex.sprint(
    \[\text{New pattern: } \]
    p]
    pats = pats \ . \ . \ . \ . \ . \ p
else
tex.sprint(
    \[\text{Renew pattern: } \]
    p]
end
end
lang.patterns(lg, pats)
end

\endgroup
\def\BabelStringsDefault{unicode}
\let\luabbl@stop\relax
\AddBabelHook{luatex}{encodedcommands}{%\def\bbl@tempa{utf8}\def\bbl@tempb{#1}%\ifx\bbl@tempa\bbl@tempb\else\directlua{Babel.begin_process_input()}%\def\luabbl@stop{\directlua{Babel.end_process_input()}}%\fi}%\AddBabelHook{luatex}{stopcommands}{\luabbl@stop\let\luabbl@stop\relax}\AddBabelHook{luatex}{patterns}{\@ifundefined{bbl@hyphendata@\the\language}{\bbl@info{No hyphenation patterns were set for\%
language '#2'. Reported}}{}{\expandafter\expandafter\expandafter\bbl@luapatterns\csname bbl@hyphendata@\the\language\endcsname}{}}{}{\bbl@languages\@ifundefined{bbl@hyphendata@\the\language}{%}{\{\def\bbl@eit\#1\#2\#3\#4\%\ifnum\#2=csname l@\#2\endcsname % =spanish, dutch:OT1...
\def\bbl@tempb{\#3}\%\if\bbl@tempb@empty\else% if not a synonymous\def\bbl@tempc{\#3\#4}\%\fi\bbl@csarg\xdef{hyphendata@\#2}{\bbl@tempc}\%\fi\bbl@languages\@ifundefined{bbl@hyphendata@\the\language}{%}{\{%\bbl@info{No hyphenation patterns were set for\%
language ‘#2’. Reported}\%\expandafter\expandafter\expandafter\bbl@luapatterns\csname bbl@hyphendata@\the\language\endcsname\}{}%\@ifundefined{bbl@patterns@}{%}{\begingroup\bbl@xin@{,\number\language,},\bbl@pttnlist}\%\@if\empty\bbl@patterns@@empty\else\directlua{ Babel.addpatterns(\bbl@patterns@\}, \number\language)}%\fi\@if\empty\bbl@patterns@#1\%\empty\directlua{ Babel.addpatterns(\bbl@patterns@#1\}, \number\language)}%\xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%\fi\endgroup})}
\AddBabelHook{luatex}{everylanguage}{%
  \def\process@language##1##2##3{%\def\process@line####1####2 ####3 ####4 {}}}
\AddBabelHook{luatex}{loadpatterns}{%
  \input #1\relax
  \expandafter\gdef\csname bbl@hyphendata@\the\language\endcsname
  {{#1}{}}}
\AddBabelHook{luatex}{loadexceptions}{%
  \input #1\relax
  \def\bbl@tempb##1##2{{##1}{#1}}%
  \expandafter\xdef\csname bbl@hyphendata@\the\language\endcsname
  {\expandafter\expandafter\expandafter\bbl@tempb
    \csname bbl@hyphendata@\the\language\endcsname}}

\babelpatterns This macro adds patterns. Two macros are used to store them: \bbl@patterns@ for the
  global ones and \bbl@patterns@<lang> for language ones. We make sure there is a space
  between words when multiple commands are used.
\Onlypreamble\babelpatterns
\AtEndOfPackage{%
  \newcommand\babelpatterns[2][\@empty]{%\@empty
    \ifx\bbl@patterns@\relax
      \let\bbl@patterns@\@empty
    \fi
    \ifx\@empty#1%
      \protect@edef\bbl@patterns@\space#2
    \else
      \edef\bbl@tempb{\zap@space#1 \@empty}%
      \bbl@for\bbl@tempa\bbl@tempb{%\@empty
        \bbl@fixname\bbl@tempa
        \bbl@iflanguage\bbl@tempa{%\@empty
          \bbl@csarg\protect@edef\patterns@\bbl@tempa{\@empty
            \@ifundefined{bbl@patterns@\bbl@tempa}{}{\csname bbl@patterns@\bbl@tempa\endcsname\space}%
          \@empty
          \{\csname bbl@patterns@\bbl@tempa\endcsname\space}\space}%
        }%}
    \fi}
}

Common stuff.
\AddBabelHook{luatex}{loadkernel}{%
  ⟨⟨Restore Unicode catcodes before loading patterns⟩⟩}
\ifx\DisableBabelHook\@undefined\endinput\fi
\AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
\DisableBabelHook{babel-fontspec}

⟨⟨Font selection⟩⟩

13.4 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 \textdirection), and when \parbox and \hangindent are involved. Fortunately, latest releases of luatex simplify a lot the solution with \shapemode.

\begin{verbatim}
3573 \bbl@trace{Redefinitions for bidi layout}
3574 \ifx\@eqnnum\@undefined\else
3575 \edef\@eqnnum{\unexpanded{%
3576 \ifcase\bbl@attr@dir\else\bbl@textdir\@ne\fi%}
3577 \unexpanded\expandafter{\@eqnnum}}\fi
3578 \fi
3579 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
3580 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
3581 \def\bbl@nextfake#1{%
3582 \mathdir\bodydir % non-local, use always inside a group!
3583 \bbl@exp{% #1% Once entered in math, set boxes to restore values
3584 \everyvbox{% \the\everyvbox
3585 \bodydir\the\bodydir
3586 \mathdir\the\mathdir
3587 \everyhbox{\the\everyhbox}%
3588 \everyvbox{\the\everyvbox}}%
3589 \everyhbox{% \the\everyhbox
3590 \bodydir\the\bodydir
3591 \mathdir\the\mathdir
3592 \everyhbox{\the\everyhbox}%
3593 \everyvbox{\the\everyvbox}}%}
3594 \def\@hangfrom#1{\setbox\@tempboxa\hbox{{#1}}%
3595 \hangindent\wd\@tempboxa
3596 \ifnum\bbl@getluadir{page}=%\bbl@getluadir{par}\else
3597 \shapemode\@ne
3598 \fi
3599 \noindent\box\@tempboxa}
3600 \fi
3601 \IfBabelLayout{tabular}{\def\tabular{%
3602 \leavevmode\hbox\bgroup\bbl@nextfake$$%$
3603 \let\@acol\@tabacol \let\@classz\@tabclassz
3604 \let\@classiv\@tabclassiv \let\\@tabularcr\@tabarray}
3605 {}}
3606 \IfBabelLayout{lists}{\def\list#1#2{%
3607 \ifnum\@listdepth >5\relax
3608 \@toodeep
3609 \else
3610 \global\advance\@listdepth\@ne
3611 \shapemode\@ne
3612 \relax
3613 \fi
3614 \rightmargin\z@ \listparindent\z@ \itemindent\z@
3615 \csname @list\romannumeral\the@listdepth\endcsname
3616 \def@itemlabel{#1}%
3617 \let\makelabel\@mklab \@nmbrlistfalse #2\relax
3618 \fi
3619 {}}
3620 \end{verbatim}

141
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.

Some \LaTeX macros use internally the math mode for text formatting. They have very little in common and are grouped here, as a single option.

13.5 Auto bidi with basic-r

The file babel-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.

TODO: math mode (as weak L?)
function Babel.bidi(head, ispar)
    local first_n, last_n -- first and last char with nums
    local last_es -- an auxiliary 'last' used with nums
    local first_d, last_d -- first and last char in L/R block
    local dir, dir_real

    Next also depends on script/lang (<al>/<r>). To be set by babel. tex.pardir is dangerous, could be (re)set but it should be changed only in vmode. There are two strong's – strong = l/al/r and strong_lr = l/r (there must be a better way):
    local strong = ('TRT' == tex.pardir) and 'r' or 'l'
    local strong_lr = (strong == 'l') and 'l' or 'r'
    local outer = strong

    local new_dir = false
    local first_dir = false

    local last_lr

    local type_n = ''

    for item in node.traverse(head) do
        -- three cases: glyph, dir, otherwise
        if item.id == node.id'glyph'
            or (item.id == 7 and item.subtype == 2) then
            local itemchar
            if item.id == 7 and item.subtype == 2 then
                itemchar = item.replace.char
            else
                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
            dir = dir or 'l'

            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.

            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'
            end
        end
    end
end
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.

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' -- W3
end

Once finished the basic setup for glyphs, consider the two other cases: dir node and the rest.

elseif item.id == node.id'dir' then
    new_dir = true
    dir = nil
else
    dir = nil -- Not a char
end

Numbers in R mode. A sequence of <en>, <et>, <an>, <es> and <cs> is typeset (with some rules) in L mode. We store the starting and ending points, and only when anything different is found (including nil, i.e., 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>.

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
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, whatstits, etc., are ignored:

```lua
    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

    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

    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'
    elseif new_dir then
      last_lr = nil
    end

    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)
```

**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>, respectively, 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.
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

And here the Lua code for bidi=basic:

Babel = Babel or {}
Babel.fontmap = Babel.fontmap or {}
Babel.fontmap[0] = {} -- l
Babel.fontmap[1] = {} -- r

function Babel.pre_otfload_v(head)
  -- head = Babel.numbers(head)
  head = Babel.bidi(head, true)
  return head
end

function Babel.pre_otfload_h(head, gc, sz, pt, dir)
  -- head = Babel.numbers(head)
  head = Babel.bidi(head, false, dir)
  return head
end

require('babel-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
    d = node.new(DIR)
    d.dir = '+' .. dir
    node.insert_before(head, state.san, d)
    local d = node.new(DIR)
    d.dir = '-' .. dir
    node.insert_after(head, state.eim, d)
  end
  new_state.san, new_state.ean = nil, nil
  return head, new_state
end

local function insert_implicit(head, state, outer)
  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)
  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

-- \hbox with an explicit dir can lead to wrong results
-- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>

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 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')
else
    save_outer = ('TRT' == tex.pardir) 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'
end

if new_d then
    table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
    attr_d = node.get_attribute(item, ATDIR)
    attr_d = attr_d % 3
    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

elseif item.id == DIR then
    d = nil
    new_d = true
else
    d = nil
end
if last == 'al' and d == 'en' then
    d = 'an'  -- W3
elseif last == 'al' and (d == 'et' or d == 'es') then
    d = 'on'  -- W6
end

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

if d == 'en' and #nodes >= 2 then
    if (nodes[#nodes][2] == 'cs')
        and nodes[#nodes-1][2] == 'an' then
        nodes[#nodes][2] = 'an'
    end
end

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
    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})
else
    -- Not sure about the following. Looks too 'ad hoc', but it's
    -- required for numbers, so that 89 19 becomes 19 89. It also
    -- affects n+cs/es+n.
if prev_d == 'an' or prev_d == 'en' then
    table.insert(nodes, {item, 'on', nil})
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
```plaintext
item.char = characters[item.char].m or item.char
end

first_on = nil
end

if d == 'r' or d == 'l' then last = d end end

---------- 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.san = state.san or item
    state.ean = item
  elseif state.san then
    head, state = insert_numeric(head, state)
  end
  if outer == 'l' then
    if d == 'an' or d == 'r' then -- im -> implicit
      if d == 'r' then state.has_r = true end
      state.sim = state.sim or item
      state.eim = item
    elseif d == 'l' and 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
      state.sim = state.sim or item
      state.eim = item
    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
  end
  if isdir then
    last = d -- Don't search back - best save now
end
```

14 The ‘nil’ language

This ‘language’ does nothing, except setting the hyphenation patterns to nohyphenation. For this language currently no special definitions are needed or available. The macro \LdfInit takes care of preventing that this file is loaded more than once, checking the category code of the @ sign, etc.

\ProvidesLanguage{nil}[(\date) (\version) Nil language]
\LdfInit{nil}{datenil}

When this file is read as an option, i.e. by the \usepackage command, nil could be an ‘unknown’ language in which case we have to make it known.

\ifx\l@nohyphenation\@undefined
\@nopatterns{nil}
\adddialect\l@nil0
\else
\let\l@nil\l@nohyphenation
\fi

This macro is used to store the values of the hyphenation parameters \lefthyphenmin and \righthyphenmin.
\providehyphenmins{\CurrentOption}{\m@ne\m@ne}

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

\caption{nil}
\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}

15 Support for Plain \TeX (plain.def)

15.1 Not renaming hyphen.tex

As Don Knuth has declared that the filename hyphen.tex may only be used to designate his version of the american English hyphenation patterns, a new solution has to be found in order to be able to load hyphenation patterns for other languages in a plain-based \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 locallyhyphen.tex or whatever they like, but they mustn’t diddle with hyphen.tex (or plain.tex except to preload additional fonts).

The files bplain.tex and 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 initTeX, 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 initTeX sees, we need to set some category codes just to be able to change the definition of \input

\catcode`\{=1 % left brace is begin-group character
\catcode`\}=2 % right brace is end-group character
\catcode`\#=6 % hash mark is macro parameter character

Now let’s see if a file called hyphen.cfg can be found somewhere on TeX’s input path by trying to open it for reading...

\openin 0 hyphen.cfg

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

\ifeof 0
\else
When hyphen.cfg could be opened we make sure that it will be read instead of the file hyphen.tex which should (according to Don Knuth’s ruling) contain the american English hyphenation patterns and nothing else.
We do this by first saving the original meaning of \input (and I use a one letter control sequence for that so as not to waste multi-letter control sequence on this in the format).

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

\def\input #1 {%
\let\input\a
\a hyphen.cfg

Once that’s done the original meaning of \input can be restored and the definition of \a can be forgotten.

\let\a\undefined
\fi
//bplain | blplain

Now that we have made sure that hyphen.cfg will be loaded at the right moment it is time to load plain.tex.

(bplain)\a plain.tex
(bplain)\a lplain.tex

Finally we change the contents of \fmtname to indicate that this is not the plain format, but a format based on plain with the babel package preloaded.

(bplain)\def\fmtname{babel-plain}
bplain)\def\fmtname{babel-lplain}

When you are using a different format, based on plain.tex you can make a copy of blplain.tex, rename it and replace plain.tex with the name of your format file.
15.2 Emulating some \LaTeX features

The following code duplicates or emulates parts of \LaTeX\ 2ε that are needed for babel.

\begin{verbatim}
\long\def\@firstofone#1{#1}
\long\def\@firstoftwo#1#2{#1}
\long\def\@secondoftwo#1#2{#2}
\def\@nnil\@nil
\def\@gobbletwo#1#2{
\def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
\def\@star@or@long#1{
\@ifstar{
\let\l@ngrel@x\relax#1}{\let\l@ngrel@x\long#1}}
\let\l@ngrel@x\relax
\long\def\@car#1#2\@nil{\@firstofone{#1}}
\long\def\@cdr#1#2\@nil{\@firstofone{#2}}
\let\@typeset@protect\relax
\long\def\@gobble#1{}
\edef\@backslashchar{\expandafter\@gobble\string\\}
\def\strip@prefix#1>{
\def\g@addto@macro#1#2{\expandafter\xdef#1\expandafter{\the#1#2}}}
\long\def\@namedef#1\csname \endcsname
\long\def\@nameuse#1\csname \endcsname
\if\@ifundefined{\expandafter\if\csname#1\endcsname\relax
\else\expandafter\@secondoftwo\fi}
\def\@expandtwoargs#1#2#3{\edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
\long\def\zap@space#1 #2{#1\ifx#2\@empty\else\expandafter\zap@space\fi #2}
\end{verbatim}

15.3 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}
\long\def\@secondoftwo#1#2{#2}
\def\@nnil\@nil
\def\@gobbletwo#1#2{
\def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
\def\@star@or@long#1{
\@ifstar{
\let\l@ngrel@x\relax#1}{\let\l@ngrel@x\long#1}}
\let\l@ngrel@x\relax
\long\def\@car#1#2\@nil{\@firstofone{#1}}
\long\def\@cdr#1#2\@nil{\@firstofone{#2}}
\let\@typeset@protect\relax
\long\def\@gobble#1{}
\edef\@backslashchar{\expandafter\@gobble\string\\}
\def\strip@prefix#1>{
\def\g@addto@macro#1#2{\expandafter\xdef#1\expandafter{\the#1#2}}}
\long\def\@namedef#1\csname \endcsname
\long\def\@nameuse#1\csname \endcsname
\if\@ifundefined{\expandafter\if\csname#1\endcsname\relax
\else\expandafter\@secondoftwo\fi}
\def\@expandtwoargs#1#2#3{\edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
\long\def\zap@space#1 #2{#1\ifx#2\@empty\else\expandafter\zap@space\fi #2}
\end{verbatim}

155
\TeX\ has the command \texttt{\@onlypreamble} which adds commands to a list of commands that are no longer needed after \texttt{\begin{document}}.

\begin{verbatim}
4279 \ifx\@preamblecmds\@undefined
4280 \def\@preamblecmds{}
4281 \fi
4282 \def\onlypreamble#1{%
4283 \expandafter\gdef\expandafter\@preamblecmds\expandafter{%\@preamblecmds\do#1}
4284 \@onlypreamble\@onlypreamble
\end{verbatim}

Mimic \TeX\'s \texttt{\AtBeginDocument}; for this to work the user needs to add \texttt{\begindocument} to his file.

\begin{verbatim}
4286 \def\begindocument{%
4287 \@begindocumenthook
4288 \global\let\@begindocumenthook\@undefined
4289 \def\do##1{%\global\let##1\@undefined}\
4290 \@preamblecmds
4291 \global\let\do\noexpand
4292 \ifx\@begindocumenthook\@undefined
4293 \def\@begindocumenthook{}
4294 \fi
4295 \onlypreamble\@begindocumenthook
4296 \def\AtBeginDocument{\g@addto@macro\@begindocumenthook}
\end{verbatim}

We also have to mimic \TeX's \texttt{\AtEndOfPackage}. Our replacement macro is much simpler; it stores its argument in \texttt{\@endofldf}.

\begin{verbatim}
4297 \def\AtEndOfPackage#1{\g@addto@macro\@endofldf{#1}}
4298 \onlypreamble\AtEndOfPackage
4299 \def\@endofldf{}
4300 \onlypreamble\@endofldf
4301 \let\bbl@afterlang\@empty
4302 \chardef\bbl@opt@hyphenmap\@empty
\end{verbatim}

\TeX\ needs to be able to switch off writing to its auxiliary files; plain doesn't have them by default.

\begin{verbatim}
4303 \ifx\if@filesw\@undefined
4304 \expandafter\let\csname\endcsname\if@filesw\expandafter\endcsname
4305 \csname iffalse\endcsname
4306 \fi
\end{verbatim}

Mimick \TeX\'s commands to define control sequences.

\begin{verbatim}
4307 \def\newcommand{%\star@or@long\new\command}
4308 \def\new\command#1{%
4309 \@testopt{\\new\command#1}{0}
4310 \def\new\command#1[#2]{%\ifnextchar [{{}@argdef#1[#2]}%\else
4311 \@argdef#1[#2]{\@xargdef#1[#2]}%\fi
4312 \argdef#1[#2]}%\fi
4313 \long\def\argdef#1[#2]{%\@yargdef#1[#2]{\@ne}}%\fi
4314 \long\def\@yargdef#1#2#3{\@tempcnta#3\relax
4315 \advance\@tempcnta\@ne
4316 \expandafter\def\expandafter#2\expandafter{\@tempcnta}{\@tempcnta}{
4317 \expandafter\def\expandafter#2\expandafter{\@tempcnta}{\@tempcnta}{
4318 \expandafter\def\expandafter#2\expandafter{\@tempcnta}{\@tempcnta}{
4319 \expandafter\def\expandafter#2\expandafter{\@tempcnta}{\@tempcnta}{
4320 \two@{\@ne}{#2}{#3}}%\fi
4321 \long\def\@yargdef#1#2#3{%\fi
4322 \@tempcnta\@ne\relax
4323 \advance\@tempcnta\@one
\end{verbatim}

156
The following little macro \in@ is taken from latex.ltx; it checks whether its first argument is part of its second argument. It uses the boolean \in@; allocating a new boolean inside conditionally executed code is not possible, hence the construct with the temporary definition of \bbl@tempa.

\def\bbl@tempa{\csname newif\endcsname\in@}
\iffalse\in@\@undefined
\def\in@#1#2{%\iffalse\in@##1\else\in@false\fi}
\fi\bbl@tempa
\fi\else
\let\bbl@tempa\@empty
\fi
\bbl@tempa

\iffalse\in@ is taken from latex.ltx; it checks 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}
\def\@ifpackagewith#1#2#3#4{}\end{verbatim}

The \TeX macro \@ifpackagewith 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.

\begin{verbatim}
\def\@ifpackagewith#1#2#3#4{#3}
\def\@ifpackagewith#1#2#3#4{}
\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 \TeX versions; just enough to make things work in plain \TeX environments.

\begin{verbatim}
\ifx\@tempcnta\@undefined\csname newcount\endcsname\@tempcnta\relax\fi
\ifx\@tempcntb\@undefined\csname newcount\endcsname\@tempcntb\relax\fi
\end{verbatim}

To prevent wasting two counters in \TeX 2.09 (because counters with the same name are allocated later by it) we reset the counter that holds the next free counter (\count10).

\begin{verbatim}
\def\@ifnextchar#1#2#3{\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\reserved@c}
\def\:{\let\@sptoken=} \:{\expandafter\def: {\futurelet\@let@token\@ifnch}}
\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax #2\relax}\fi}
\long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum\else\expandafter\@gobble\fi{#1}}
\end{verbatim}

\begin{verbatim}
\def\@testopt#1#2{\@ifnextchar\[{#1}{#1[#2]}}\def\@protected@testopt#1{\ifx\protect\@typeset@protect\expandafter\@testopt\else\@x@protect#1\fi}
\long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax #2\relax}\fi}
\end{verbatim}

\begin{verbatim}
\def\@x@protect#1{\ifx\@tempcnta\@undefined\csname newcount\endcsname\@tempcnta\relax\fi}
\def\@x@protect#1{\ifx\@tempcntb\@undefined\csname newcount\endcsname\@tempcntb\relax\fi}
\end{verbatim}

\begin{verbatim}
\ifx\@tempcnta\@undefined\csname newcount\endcsname\@tempcnta\relax\fi
\ifx\@tempcntb\@undefined\csname newcount\endcsname\@tempcntb\relax\fi
\end{verbatim}
15.4 Encoding related macros

Code from loutenc.dtx, adapted for use in the plain \TeX environment.

\begin{verbatim}
\def\DeclareTextCommand{%\@dec@text@cmd\providecommand
\def\ProvideTextCommand{%\@dec@text@cmd\providecommand
\def\DeclareTextSymbol#1#2#3{%\@dec@text@cmd\chardef#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{%\@dec@text@cmd\ifx\protect\@typeset@protect\else\noexpand#1\@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\def\csname ?\string#1\endcsname{%\@changed@x@err{#1}\fi\global\expandafter\let\csname\cf@encoding \string#1\expandafter\endcsname\csname ?\string#1\endcsname\else\noexpand#1\@gobble\fi}
\def\@changed@x@err#1{%\errhelp{Your command will be ignored, type <return> to proceed}\errmessage{Command \protect#1 undefined in encoding \cf@encoding}}\def\DeclareTextCommandDefault#1{%\DeclareTextCommand#1?\ProvideTextCommandDefault#1{%\expandafter\let\csname OT1-cmd\endcsname\@current@cmd\expandafter\let\csname?-cmd\endcsname\@changed@cmd\def\DeclareTextAccent#1#2#3{%\DeclareTextCommand#1{#2}[1]{\accent#3 ##1}\endcsname\else\noexpand#1\fi\@changed@x@err{#1}\{
\end{verbatim}

159
Currently we only use the \textipaZ 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 \textipaZ-control sequence \scriptsize to be available. Because plain \TeX doesn't have such a sophisticated font mechanism as \textipaZ has, we just \let it to \sevenrm.

\let\scriptsize\sevenrm
\let\scriptsize\sevenrm
\fi
\fi

\section*{Acknowledgements}

I would like to thank all who volunteered as \textipaZ-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


