Babel

Localization and internationalization

\TeX
\pdf\TeX
\Lua\TeX
\LuaHBT\TeX
\Xe\TeX
Contents

I User guide 4

1 The user interface 4
   1.1 Monolingual documents 4
   1.2 Multilingual documents 6
   1.3 Mostly monolingual documents 7
   1.4 Modifiers 8
   1.5 Troubleshooting 8
   1.6 Plain 8
   1.7 Basic language selectors 9
   1.8 Auxiliary language selectors 9
   1.9 More on selection 10
   1.10 Shorthands 12
   1.11 Package options 15
   1.12 The base option 17
   1.13 ini files 18
   1.14 Selecting fonts 25
   1.15 Modifying a language 27
   1.16 Creating a language 28
   1.17 Digits and counters 31
   1.18 Accessing language info 32
   1.19 Hyphenation and line breaking 33
   1.20 Selecting scripts 35
   1.21 Selecting directions 36
   1.22 Language attributes 40
   1.23 Hooks 41
   1.24 Languages supported by babel with ldf files 42
   1.25 Unicode character properties in luatex 43
   1.26 Tweaking some features 43
   1.27 Tips, workarounds, known issues and notes 44
   1.28 Current and future work 45
   1.29 Tentative and experimental code 45

2 Loading languages with language.dat 46
   2.1 Format 46

3 The interface between the core of babel and the language definition files 47
   3.1 Guidelines for contributed languages 48
   3.2 Basic macros 49
   3.3 Skeleton 50
   3.4 Support for active characters 51
   3.5 Support for saving macro definitions 51
   3.6 Support for extending macros 51
   3.7 Macros common to a number of languages 52
   3.8 Encoding-dependent strings 52

4 Changes 56
   4.1 Changes in babel version 3.9 56

II Source code 56

5 Identification and loading of required files 57
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>locale directory</td>
<td>57</td>
</tr>
<tr>
<td>7</td>
<td>Tools</td>
<td>57</td>
</tr>
<tr>
<td>7.1</td>
<td>Multiple languages</td>
<td>62</td>
</tr>
<tr>
<td>8</td>
<td>The Package File ($\LaTeX$, babel.sty)</td>
<td>63</td>
</tr>
<tr>
<td>8.1</td>
<td>base</td>
<td>63</td>
</tr>
<tr>
<td>8.2</td>
<td>key=value options and other general option</td>
<td>65</td>
</tr>
<tr>
<td>8.3</td>
<td>Conditional loading of shorthands</td>
<td>66</td>
</tr>
<tr>
<td>8.4</td>
<td>Language options</td>
<td>68</td>
</tr>
<tr>
<td>9</td>
<td>The kernel of Babel (babel.def, common)</td>
<td>70</td>
</tr>
<tr>
<td>9.1</td>
<td>Tools</td>
<td>70</td>
</tr>
<tr>
<td>9.2</td>
<td>Hooks</td>
<td>73</td>
</tr>
<tr>
<td>9.3</td>
<td>Setting up language files</td>
<td>75</td>
</tr>
<tr>
<td>9.4</td>
<td>Shorthands</td>
<td>77</td>
</tr>
<tr>
<td>9.5</td>
<td>Language attributes</td>
<td>86</td>
</tr>
<tr>
<td>9.6</td>
<td>Support for saving macro definitions</td>
<td>89</td>
</tr>
<tr>
<td>9.7</td>
<td>Short tags</td>
<td>89</td>
</tr>
<tr>
<td>9.8</td>
<td>Hyphens</td>
<td>90</td>
</tr>
<tr>
<td>9.9</td>
<td>Multiencoding strings</td>
<td>91</td>
</tr>
<tr>
<td>9.10</td>
<td>Macros common to a number of languages</td>
<td>97</td>
</tr>
<tr>
<td>9.11</td>
<td>Making glyphs available</td>
<td>97</td>
</tr>
<tr>
<td>9.11.1</td>
<td>Quotation marks</td>
<td>98</td>
</tr>
<tr>
<td>9.11.2</td>
<td>Letters</td>
<td>99</td>
</tr>
<tr>
<td>9.11.3</td>
<td>Shorthands for quotation marks</td>
<td>100</td>
</tr>
<tr>
<td>9.11.4</td>
<td>Umlauts and tremas</td>
<td>101</td>
</tr>
<tr>
<td>9.12</td>
<td>Layout</td>
<td>102</td>
</tr>
<tr>
<td>9.13</td>
<td>Load engine specific macros</td>
<td>103</td>
</tr>
<tr>
<td>9.14</td>
<td>Creating languages</td>
<td>103</td>
</tr>
<tr>
<td>10</td>
<td>Adjusting the Babel behavior</td>
<td>117</td>
</tr>
<tr>
<td>11</td>
<td>The kernel of Babel (babel.def for $\LaTeX$ only)</td>
<td>118</td>
</tr>
<tr>
<td>11.1</td>
<td>The redefinition of the style commands</td>
<td>118</td>
</tr>
<tr>
<td>11.2</td>
<td>Cross referencing macros</td>
<td>118</td>
</tr>
<tr>
<td>11.3</td>
<td>Marks</td>
<td>122</td>
</tr>
<tr>
<td>11.4</td>
<td>Preventing clashes with other packages</td>
<td>123</td>
</tr>
<tr>
<td>11.4.1</td>
<td>ifthen</td>
<td>123</td>
</tr>
<tr>
<td>11.4.2</td>
<td>varioref</td>
<td>123</td>
</tr>
<tr>
<td>11.4.3</td>
<td>hline</td>
<td>124</td>
</tr>
<tr>
<td>11.4.4</td>
<td>hyperref</td>
<td>124</td>
</tr>
<tr>
<td>11.4.5</td>
<td>fancyhdr</td>
<td>125</td>
</tr>
<tr>
<td>11.5</td>
<td>Encoding and fonts</td>
<td>125</td>
</tr>
<tr>
<td>11.6</td>
<td>Basic bidi support</td>
<td>127</td>
</tr>
<tr>
<td>11.7</td>
<td>Local Language Configuration</td>
<td>130</td>
</tr>
<tr>
<td>12</td>
<td>Multiple languages (switch.def)</td>
<td>131</td>
</tr>
<tr>
<td>12.1</td>
<td>Selecting the language</td>
<td>132</td>
</tr>
<tr>
<td>12.2</td>
<td>Errors</td>
<td>141</td>
</tr>
<tr>
<td>13</td>
<td>Loading hyphenation patterns</td>
<td>142</td>
</tr>
<tr>
<td>14</td>
<td>Font handling with fontspec</td>
<td>147</td>
</tr>
</tbody>
</table>
15 Hooks for Xe\TeX and Lua\TeX
  15.1 Xe\TeX ............................................. 152
  15.2 Layout ............................................. 154
  15.3 Lua\TeX ............................................. 156
  15.4 Southeast Asian scripts ................................ 162
  15.5 CJK line breaking .................................. 165
  15.6 Automatic fonts and ids switching ......... 166
  15.7 Layout ............................................. 173
  15.8 Auto bidi with basic and basic-r .................................. 175

16 Data for CJK

17 The ‘nil’ language

18 Support for Plain \TeX (plain.def)
  18.1 Not renaming hyphen.tex .................................. 187
  18.2 Emulating some \LaTeX features .......................... 188
  18.3 General tools ....................................... 189
  18.4 Encoding related macros ............................ 192

19 Acknowledgements

Troubleshooting

\begin{quote}
Paragraph ended before $\backslash$UTFviii@three@octets was complete
\end{quote}

\begin{quote}
No hyphenation patterns were preloaded for (babel) the language ‘LANG’ into the format
\end{quote}

\begin{quote}
You are loading directly a language style
\end{quote}

\begin{quote}
Unknown language ‘LANG’
\end{quote}

\begin{quote}
Argument of $\backslash$language@active@arg” has an extra)
\end{quote}

\begin{quote}
Package fontspec Warning: ‘Language ‘LANG’ not available for font ‘FONT’ with script ‘SCRIPT’ ‘Default’ language used instead
\end{quote}

\begin{quote}
Package babel Info: The following fonts are not babel standard families
\end{quote}
Part I

User guide

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

• Changes and new features with relation to version 3.8 are highlighted with New X.XX, and there are some notes for the latest versions in the babel wiki. The most recent features could be still unstable. Please, report any issues you find in GitHub, which is better than just complaining on an e-mail list or a web forum.

• If you are interested in the \TeX multilingual support, please join the kadingiramail list. You can follow the development of babel in GitHub (which provides many sample files too).

• See section 3.1 for contributing a language.

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

1 The user interface

1.1 Monolingual documents

In most cases, a single language is required, and then all you need in \LaTeX is to load the package using its standard mechanism for this purpose, namely, passing that language as an optional argument. In addition, you may want to set the font and input encodings. Many languages are compatible with xetex and luatex. With them you can use babel to localize the documents. When these engines are used, 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 lmrroman. Other scripts require loading fontspec. You may want to set the font attributes with fontspec, too.

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

```latex
\documentclass{article}
\usepackage[T1]{fontenc}
% \usepackage[utf8]{inputenc} % Uncomment if LaTeX < 2018-04-01
\usepackage[french]{babel}
\begin{document}

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

\end{document}
```

**EXAMPLE** And now a simple monolingual document in Russian (text from the Wikipedia) with xetex or luatex. 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).

```latex
\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}
```

**TROUBLESHOOTING**  A common source of trouble is a wrong setting of the input encoding. Depending on the \LaTeX{} version you could get the following somewhat cryptic error:

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

Or the more explanatory:

```
! Package inputenc Error: Invalid UTF-8 byte ...
```

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:

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

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

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

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

```
Package babel Warning: No hyphenation patterns were preloaded for \language=LANG into the format.
```

```
Package babel Warning: 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 a list of the required languages as package or class options. The last language is considered the main one, activated by default. Sometimes, the main language changes the document layout (e.g., Spanish and French).

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

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

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

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

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

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

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

\begin{verbatim}
\PassOptionsToPackage{main=english}{babel}
\end{verbatim}

**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`:

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

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

**EXAMPLE**  With xetex and luatex, 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}

### 1.3 Mostly monolingual documents

**New 3.39**  Very often, multilingual documents consist of a main language with small pieces of text in another languages (words, idioms, short sentences). Typically, all you need is to set the line breaking rules and, perhaps, the font. In such a case, babel now does not require declaring these secondary languages explicitly, because the basic settings are loaded on the fly when the language is selected (and also when provided in the optional argument of \babelfont, if used.)

This is particularly useful, too, when there are short texts of this kind coming from an external source whose contents are not known on beforehand (for example, titles in a bibliography). At this regard, it is worth remembering that \babelfont does not load any font until required, so that it can be used just in case.

**EXAMPLE**  A trivial document is:

\documentclass{article}
\usepackage[english]{babel}
1.4 Modifiers

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

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

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

1.5 Troubleshooting

• Loading directly sty files in \LaTeX (ie, \usepackage{⟨language⟩}) is deprecated and you will get the error: \[2\]

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

• Another typical error when using babel is the following: \[3\]

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

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):

1No predefined “axis” for modifiers are provided because languages and their scripts have quite different needs.
2In old versions the error read “You have used an old interface to call babel”, not very helpful.
3In old versions the error read “You haven't loaded the language LANG yet.”
WARNING Not all languages provide a sty file and some of them are not compatible with Plain.  

1.7 Basic language selectors

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

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

\selectlanguage{german}  
This command can be used as environment, too.

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

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

{\selectlanguage{⟨inner-language⟩} ...}\selectlanguage{⟨outer-language⟩}

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

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

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

\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.

The environment `\begin{otherlanguage*}{language}` is similar to `\begin{otherlanguage}{language}`, but as an 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 environments 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{language}`, except when the option `bidi` is set – in this case, `\foreignlanguage{language}` emits a `\leavevmode`, while `otherlanguage*` does not.

The environment `\begin{hyphenrules}{language}` 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, ‘donebysomelanguages’ (e.g., Italian, French, Ukrainian).

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

### 1.9 More on selection

\begin{verbatim}
\text{tag1}\text{tag2} = \text{language1}, \text{tag2} = \text{language2}, ...
\end{verbatim}

\textbf{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 \begin{verbatim}\text{tag1}\text{tag2}\end{verbatim} to be `\foreignlanguage{language1}\{\text{tag2}\}`, and `\begin{verbatim}\text{tag1}\end{verbatim}` to be `\begin{otherlanguage*}{language1}\{\text{tag1}\}`, and so on. Note `\text{tag1}` is also allowed, but remember to set it locally inside a group.

\textbf{EXAMPLE} With

\footnote{Even in the babel kernel there were some macros not compatible with plain. Hopefully these issues have been fixed.}
\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\langle tag\rangle, namely, it is not affected by \MakeUppercase (while \foreignlanguage is).

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

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

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

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

```
\babelensure{polish}
```

By default only the basic captions and \today are redefined, but you can add further macros with the key include in the optional argument (without commas). Macros not to be modified are listed in exclude. You can also enforce a font encoding with fontenc\footnote{With it, encoded strings may not work as expected}.

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

They are activated when the language is selected (at the afterextras event), and it makes some assumptions which could not be fulfilled in some languages. Note also you should include only macros defined by the language, not global macros (e.g., \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, for example: (1) in some languages shorthands such as "a are defined to be able to hyphenate the word if the encoding is OT1; (2) in some languages shorthands such as \ are used to insert the right amount of white space; (3) several kinds of discretionary and breaks can be inserted easily with ", "", etc. The package inputenc as well as xetex and luatex have alleviated entering non-ASCII characters, but minority languages and some kinds of text can still require characters not directly available on the keyboards (and sometimes not even as separated or precomposed Unicode characters). As to the point 2, now pdf\TeX 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).

TROUBLESHOOTING 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, "{}`).

```
\shortandon
\{\shorthands-list\}
\shorthandoff
```

\shortandon
\{\shorthands-list\}
\shorthandoff

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

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

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

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

If you do not need shorthands, or prefer an alternative approach of your own, you may want to switch them off with the package option shorthands=off, as described below.
The command \useshorthands \{\langle char \rangle\} initiates the definition of user-defined shorthand sequences. It has one argument, the character that starts these personal shorthands.

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

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

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

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

\begin{verbatim}
\useshorthands*{
\defineshorthand{*}{\texttt{\babelhyphen{soft}}} \\
\defineshorthand{-}{\texttt{\babelhyphen{hard}}}
\end{verbatim}

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

\begin{verbatim}
\defineshorthand[\texttt{*polish,*portuguese}]{\texttt{-}}{\texttt{\babelhyphen{repeat}}}
\end{verbatim}

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 \languageshorthands \{\langle language \rangle\} can be used to switch the shorthands on the language level. It takes one argument, the name of a language or none (the latter does what its name suggests).\footnote{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.} 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{\languageshorthandsn{german}}

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

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

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

\babelshorthand{(\textit{shorthand})}

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

**EXAMPLE** Since by default shorthands are not activated until \begin{document}, you may use this macro when defining the \texttt{title} in the preamble:

\texttt{\title{Documento científico\babelshorthand{"-}\técnico}}

For your records, here is a list of shorthands, but you must double check them, as they may change\footnote{Thanks to Enrico Gregorio}{\footnote{This declaration serves to nothing, but it is preserved for backward compatibility.}}.

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

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

Basque " ' ~
Breton : ; ? !
Catalan " ' \nCzech " -
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\footnote{This declaration serves to nothing, but it is preserved for backward compatibility.}.
\ifbeshorthand \{\langle\text{character}\rangle}\{\langle\text{true}\rangle}\{\langle\text{false}\rangle}\}

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

\aliashorthand \{\langle\text{original}\rangle}\{\langle\text{alias}\rangle}\}

The command \aliashorthand 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 \aliashorthand{"}\{/\}. For the reasons in the warning below, usage of this macro is not recommended.

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

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

\aliashorthand{-}\{^\}
\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.

### 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=** \langle\text{char}\rangle\langle\text{char}\rangle... | off

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

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

If ' is included, activeacute is set; if ` is included, activegrave is set. Active characters (like ~) should be preceded by \string (otherwise they will be expanded by \LaTeX before they are passed to the package and therefore they will not be recognized); however, t is provided for the common case of ~ (as well as c for not so common case of the comma). With shorthands=off no language shorthands are defined, As some languages use this mechanism for tools not available otherwise, a macro \beshorthand is defined, which allows using them; see above.
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. As of \texttt{New 3.34}, in \LaTeX based engines (ie, almost every engine except the oldest ones) shorthands can be used in these macros (formerly you could not).

Shorthands are mainly intended for text, not for math. By setting this option with the value \texttt{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 anymore.

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

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.

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.

Global and language default config files are not loaded, so you can make sure your document is not spoilt by an unexpected .cfg file. However, if the key \texttt{config} is set, this file is loaded.

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.

Language settings for uppercase and lowercase mapping (as set by \texttt{\SetCase}) are ignored. Use only if there are incompatibilities with other packages.

No warnings and no \texttt{infos} are written to the log file\footnote{You can use alternatively the package \texttt{silence}}.

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

You can use alternatively the package \texttt{silence}.\footnotemark
**New 3.9g** Sets the behavior of case mapping for hyphenation, provided the language defines it\(^{10}\). 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\(^1\);
- **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 foots (if the option headfoot is used) and in auxiliary files (ie, at `\select@language`), and it's the default if several language options have been stated. The option **first** can be regarded as an optimized version of **other*** for monolingual documents\(^2\).

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

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

**layout** =

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

### 1.12 The base option

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

`\AfterBabelLanguage{⟨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

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

does ... at the end of french.1df. It can be used in ldf files, too, but in such a case the code is executed only if ⟨option-name⟩ is the same as `\CurrentOption` (which could not be the same as the option name as set in `\usepackage`).

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

---

\(^{10}\)Turned off in plain.  
\(^{11}\)Duplicated options count as several ones.  
\(^{12}\)Providing foreign is pointless, because the case mapping applied is that at the end of the paragraph, but if either xetex or luatex change this behavior it might be added. On the other hand, other is provided even if I [[BLL]] think it isn't really useful, but who knows.
1.13 ini files

An alternative approach to define a language (or, more precisely, a locale) is by means of an ini file. Currently babel provides about 200 of these files containing the basic data required for a locale.

ini files are not meant only for babel, and they have been devised as a resource for other packages. To easy interoperability between \TeX and other systems, they are identified with the BCP 47 codes as preferred by the Unicode Common Language Data Repository, which was used as source for most of the data provided by these files, too (the main exception being the \ldots name strings).

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 study. In other words, \babelprovide is mainly meant for auxiliary tasks.

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

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

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

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

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

Devanagari In luatex many fonts work, but some others do not, the main issue being the 'ra'. It is advisable to set explicitly the script to either deva or dev2, eg:
Other Indic scripts are still under development in luatex. On the other hand, xetex is better. The upcoming lualatex will be based on luahbtex, so Indic scripts will be rendered correctly with the option Renderer=Harfbuzz in fontspec.

Southeast scripts

Thai works in both luatex and xetex, but line breaking differs (rules can be modified in luatex; they are hard-coded in xetex). Lao seems to work, too, but there are no patterns for the latter in luatex. Khmer clusters are rendered wrongly. The comment about Indic scripts and lualatex also applies here. Some quick patterns could help, with something similar to:

\babelprovide[import,hyphenrules=+]{lao}
\babelpatterns{lao}{1ديل 1مشار 1_HERE 1_HERE 1_HERE} % Random

East Asia scripts

Settings for either Simplified or Traditional should work out of the box. luatex does basic line breaking, but currently xetex does not (you may load zhspacing). Although for a few words and short texts the ini files should be fine, CJK texts are best set with a dedicated framework (CJK, luatexja, kotex, CTeX, etc.). This is what the class ltjbook does with luatex, which can be used in conjunction with the ldf for Japanese, because the following piece of code loads luatexja:

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

NOTE

Wikipedia defines a locale as follows: “In computing, a locale is a set of parameters that defines the user’s language, region and any special variant preferences that the user wants to see in their user interface. Usually a locale identifier consists of at least a language code and a country/region code.” Babel is moving gradually from the old and fuzzy concept of language to the more modern of locale. Note each locale is by itself a separate “language”, which explains why there are so many files. This is on purpose, so that possible variants can be created and/or redefined easily.

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

<table>
<thead>
<tr>
<th></th>
<th>Afrikaansul</th>
<th>bem</th>
<th>Bemba</th>
</tr>
</thead>
<tbody>
<tr>
<td>agq</td>
<td>Aghem</td>
<td>bez</td>
<td>Bena</td>
</tr>
<tr>
<td>ak</td>
<td>Akan</td>
<td>bg</td>
<td>Bulgarianul</td>
</tr>
<tr>
<td>am</td>
<td>Amharicul</td>
<td>bm</td>
<td>Bambara</td>
</tr>
<tr>
<td>ar</td>
<td>Arabicul</td>
<td>bn</td>
<td>Banglaul</td>
</tr>
<tr>
<td>ar-DZ</td>
<td>Arabicul</td>
<td>bo</td>
<td>Tibetanu</td>
</tr>
<tr>
<td>ar-MA</td>
<td>Arabicul</td>
<td>brx</td>
<td>Bodo</td>
</tr>
<tr>
<td>ar-SY</td>
<td>Arabicul</td>
<td>bs-Cyr</td>
<td>Bosnian</td>
</tr>
<tr>
<td>as</td>
<td>Assamese</td>
<td>bs-Latin</td>
<td>Bosnianul</td>
</tr>
<tr>
<td>asa</td>
<td>Asu</td>
<td>bs</td>
<td>Bosnianul</td>
</tr>
<tr>
<td>ast</td>
<td>Asturianul</td>
<td>ca</td>
<td>Catalanul</td>
</tr>
<tr>
<td>az-Cyrl</td>
<td>Azerbaijani</td>
<td>ce</td>
<td>Chechen</td>
</tr>
<tr>
<td>az-Latn</td>
<td>Azerbaijani</td>
<td>cgg</td>
<td>Chiga</td>
</tr>
<tr>
<td>az</td>
<td>Azerbaijaniul</td>
<td>chr</td>
<td>Cherokee</td>
</tr>
<tr>
<td>bas</td>
<td>Basaa</td>
<td>ckb</td>
<td>Central Kurdish</td>
</tr>
<tr>
<td>be</td>
<td>Belarusianul</td>
<td>cop</td>
<td>Coptic</td>
</tr>
<tr>
<td>Code</td>
<td>Language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cs</td>
<td>Czech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cu</td>
<td>Church Slavonic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cu-Cyrs</td>
<td>Church Slavonic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cu-Glag</td>
<td>Church Slavonic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cy</td>
<td>Welsh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>da</td>
<td>Danish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dav</td>
<td>Taita</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de-AT</td>
<td>German</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de-CH</td>
<td>German</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>German</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dje</td>
<td>Zarma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dsb</td>
<td>Lower Sorbian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dua</td>
<td>Duala</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dyo</td>
<td>Jola-Fonyi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dz</td>
<td>Dzongkha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ebu</td>
<td>Embu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ee</td>
<td>Ewe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>el</td>
<td>Greek</td>
<td></td>
<td></td>
</tr>
<tr>
<td>en-AU</td>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>en-CA</td>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>en-GB</td>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>en-NZ</td>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>en-US</td>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>en</td>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eo</td>
<td>Esperanto</td>
<td></td>
<td></td>
</tr>
<tr>
<td>es-MX</td>
<td>Spanish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>es</td>
<td>Spanish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>et</td>
<td>Estonian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eu</td>
<td>Basque</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ewo</td>
<td>Ewondo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fa</td>
<td>Persian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ff</td>
<td>Fulah</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fi</td>
<td>Finnish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fil</td>
<td>Filipino</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fo</td>
<td>Faroese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fr</td>
<td>French</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fr-BE</td>
<td>French</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fr-CA</td>
<td>French</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fr-CH</td>
<td>French</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fr-LU</td>
<td>French</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fur</td>
<td>Friulian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fy</td>
<td>Western Frisian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ga</td>
<td>Irish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gd</td>
<td>Scottish Gaelic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gl</td>
<td>Galician</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gsw</td>
<td>Swiss German</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gu</td>
<td>Gujarati</td>
<td></td>
<td></td>
</tr>
<tr>
<td>guz</td>
<td>Gusii</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gv</td>
<td>Manx</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ha-GH</td>
<td>Hausa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ha-NE</td>
<td>Hausa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ha</td>
<td>Hausa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>haw</td>
<td>Hawaiian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>he</td>
<td>Hebrew</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hi</td>
<td>Hindi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hr</td>
<td>Croatian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsb</td>
<td>Upper Sorbian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hu</td>
<td>Hungarian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hy</td>
<td>Armenian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ia</td>
<td>Interlingua</td>
<td></td>
<td></td>
</tr>
<tr>
<td>id</td>
<td>Indonesian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ig</td>
<td>Igbo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii</td>
<td>Sichuan Yi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>is</td>
<td>Icelandic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>it</td>
<td>Italian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ja</td>
<td>Japanese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>jgo</td>
<td>Ngomba</td>
<td></td>
<td></td>
</tr>
<tr>
<td>jmc</td>
<td>Machame</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka</td>
<td>Georgian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kab</td>
<td>Kabyle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kam</td>
<td>Kamba</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kde</td>
<td>Makonde</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kea</td>
<td>Kabuverdianu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>khq</td>
<td>Koyra Chini</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ki</td>
<td>Kikuyu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kk</td>
<td>Kazakh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kkj</td>
<td>Kako</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kl</td>
<td>Kalaallisut</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kln</td>
<td>Kalenjin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>km</td>
<td>Khmer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kn</td>
<td>Kannada</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ko</td>
<td>Korean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kok</td>
<td>Konkani</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ks</td>
<td>Kashmir</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ksb</td>
<td>Shambala</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ksf</td>
<td>Bafia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ksh</td>
<td>Colognian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kw</td>
<td>Cornish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ky</td>
<td>Kyrgyz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lag</td>
<td>Langi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lb</td>
<td>Luxembourgish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lg</td>
<td>Ganda</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lkt</td>
<td>Lakota</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln</td>
<td>Lingala</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lo</td>
<td>Lao</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lrc</td>
<td>Northern Luri</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lt</td>
<td>Lithuanian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lu</td>
<td>Luba-Katanga</td>
<td></td>
<td></td>
</tr>
<tr>
<td>luo</td>
<td>Luo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>luy</td>
<td>Luyia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lv</td>
<td>Latvian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mas</td>
<td>Masai</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mer</td>
<td>Meru</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mfe</td>
<td>Morisyen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mg</td>
<td>Malagasy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mgh</td>
<td>Makhuwa-Meetto</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mgo</td>
<td>Meta’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mk</td>
<td>Macedonian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Language</td>
<td>Code</td>
<td>Language</td>
</tr>
<tr>
<td>------</td>
<td>-------------------</td>
<td>------</td>
<td>-------------------</td>
</tr>
<tr>
<td>ml</td>
<td>Malayalam</td>
<td>shi</td>
<td>Tachelhit</td>
</tr>
<tr>
<td>mn</td>
<td>Mongolian</td>
<td>shi-Tfng</td>
<td>Tachelhit</td>
</tr>
<tr>
<td>mr</td>
<td>Marathi</td>
<td>si</td>
<td>Sinhala</td>
</tr>
<tr>
<td>ms-BN</td>
<td>Malay</td>
<td>sk</td>
<td>Slovak</td>
</tr>
<tr>
<td>ms-SG</td>
<td>Malay</td>
<td>sl</td>
<td>Slovenian</td>
</tr>
<tr>
<td>ms</td>
<td>Malay</td>
<td>smn</td>
<td>Inari Sami</td>
</tr>
<tr>
<td>mt</td>
<td>Maltese</td>
<td>sn</td>
<td>Shona</td>
</tr>
<tr>
<td>mua</td>
<td>Mundang</td>
<td>so</td>
<td>Somali</td>
</tr>
<tr>
<td>my</td>
<td>Burmese</td>
<td>sq</td>
<td>Albanian</td>
</tr>
<tr>
<td>mzn</td>
<td>Mazanderani</td>
<td>sr-Cyrl-BA</td>
<td>Serbian</td>
</tr>
<tr>
<td>naq</td>
<td>Nama</td>
<td>sr-Cyrl-ME</td>
<td>Serbian</td>
</tr>
<tr>
<td>nb</td>
<td>Norwegian Bokmål</td>
<td>sr-Cyrl-XK</td>
<td>Serbian</td>
</tr>
<tr>
<td>nd</td>
<td>North Ndebele</td>
<td>sr-Cyrl</td>
<td>Serbian</td>
</tr>
<tr>
<td>ne</td>
<td>Nepali</td>
<td>sr-Latn-BA</td>
<td>Serbian</td>
</tr>
<tr>
<td>nl</td>
<td>Dutch</td>
<td>sr-Latn-ME</td>
<td>Serbian</td>
</tr>
<tr>
<td>nmg</td>
<td>Kwasio</td>
<td>sr-Latn-XK</td>
<td>Serbian</td>
</tr>
<tr>
<td>nn</td>
<td>Norwegian Nynorsk</td>
<td>sr-Cyrl</td>
<td>Serbian</td>
</tr>
<tr>
<td>nnh</td>
<td>Ngjemboon</td>
<td>sr-Latn</td>
<td>Serbian</td>
</tr>
<tr>
<td>nus</td>
<td>Nuer</td>
<td>sr</td>
<td>Serbian</td>
</tr>
<tr>
<td>nyn</td>
<td>Nyankole</td>
<td>sv</td>
<td>Swedish</td>
</tr>
<tr>
<td>om</td>
<td>Oromo</td>
<td>sw</td>
<td>Swahili</td>
</tr>
<tr>
<td>or</td>
<td>Odia</td>
<td>ta</td>
<td>Tamil</td>
</tr>
<tr>
<td>os</td>
<td>Ossetic</td>
<td>te</td>
<td>Telugu</td>
</tr>
<tr>
<td>pa-Arab</td>
<td>Punjabi</td>
<td>teo</td>
<td>Teso</td>
</tr>
<tr>
<td>pa-Guru</td>
<td>Punjabi</td>
<td>th</td>
<td>Thai</td>
</tr>
<tr>
<td>pa</td>
<td>Punjabi</td>
<td>ti</td>
<td>Tigrinya</td>
</tr>
<tr>
<td>pl</td>
<td>Polish</td>
<td>tk</td>
<td>Turkmen</td>
</tr>
<tr>
<td>pms</td>
<td>Piedmontese</td>
<td>to</td>
<td>Tongan</td>
</tr>
<tr>
<td>ps</td>
<td>Pashto</td>
<td>tr</td>
<td>Turkish</td>
</tr>
<tr>
<td>pt-BR</td>
<td>Portuguese</td>
<td>twq</td>
<td>Tasawaq</td>
</tr>
<tr>
<td>pt-PT</td>
<td>Portuguese</td>
<td>tzm</td>
<td>Central Atlas Tamazight</td>
</tr>
<tr>
<td>pt</td>
<td>Portuguese</td>
<td>ug</td>
<td>Uyghur</td>
</tr>
<tr>
<td>qu</td>
<td>Quechua</td>
<td>uk</td>
<td>Ukrainian</td>
</tr>
<tr>
<td>rm</td>
<td>Romansh</td>
<td>ur</td>
<td>Urdu</td>
</tr>
<tr>
<td>rn</td>
<td>Rundi</td>
<td>uz-Arab</td>
<td>Uzbek</td>
</tr>
<tr>
<td>ro</td>
<td>Romanian</td>
<td>uz-Cyrl</td>
<td>Uzbek</td>
</tr>
<tr>
<td>rof</td>
<td>Rombo</td>
<td>uz-Latn</td>
<td>Uzbek</td>
</tr>
<tr>
<td>ru</td>
<td>Russian</td>
<td>uz</td>
<td>Uzbek</td>
</tr>
<tr>
<td>rw</td>
<td>Kinyarwanda</td>
<td>vai-Latn</td>
<td>Vai</td>
</tr>
<tr>
<td>rwk</td>
<td>Rwa</td>
<td>vai-Vaii</td>
<td>Vai</td>
</tr>
<tr>
<td>sa-Beng</td>
<td>Sanskrit</td>
<td>vai</td>
<td>Vai</td>
</tr>
<tr>
<td>sa-Dev</td>
<td>Sanskrit</td>
<td>vi</td>
<td>Vietnamese</td>
</tr>
<tr>
<td>sa-Gujr</td>
<td>Sanskrit</td>
<td>vun</td>
<td>Unjo</td>
</tr>
<tr>
<td>sa-Knda</td>
<td>Sanskrit</td>
<td>wae</td>
<td>Walser</td>
</tr>
<tr>
<td>sa-Mlym</td>
<td>Sanskrit</td>
<td>xog</td>
<td>Soga</td>
</tr>
<tr>
<td>sa-Telu</td>
<td>Sanskrit</td>
<td>yav</td>
<td>Yangben</td>
</tr>
<tr>
<td>sa</td>
<td>Sanskrit</td>
<td>yi</td>
<td>Yiddish</td>
</tr>
<tr>
<td>sah</td>
<td>Sakha</td>
<td>yo</td>
<td>Yoruba</td>
</tr>
<tr>
<td>saq</td>
<td>Samburu</td>
<td>yue</td>
<td>Cantonese</td>
</tr>
<tr>
<td>sbp</td>
<td>Sangu</td>
<td>zgh</td>
<td>Standard Moroccan</td>
</tr>
<tr>
<td>se</td>
<td>Northern Sami</td>
<td>zh-Hans-HK</td>
<td>Chinese</td>
</tr>
<tr>
<td>seh</td>
<td>Sena</td>
<td>zh-Hans-MO</td>
<td>Chinese</td>
</tr>
<tr>
<td>ses</td>
<td>Koyraboro Senni</td>
<td>zh-Hans-SG</td>
<td>Chinese</td>
</tr>
<tr>
<td>sg</td>
<td>Sango</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In some contexts (currently \babel\font) an ini file may be loaded by its name. Here is the list of the names currently supported. With these languages, \babel\font loads (if not done before) the language and script names (even if the language is defined as a package option with an ldf file). These are also the names recognized by \babel\provide with a valueless import.

<table>
<thead>
<tr>
<th>Language</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>aghem</td>
<td>burmese</td>
</tr>
<tr>
<td>akan</td>
<td>canadian</td>
</tr>
<tr>
<td>albanian</td>
<td>cantonese</td>
</tr>
<tr>
<td>american</td>
<td>catalan</td>
</tr>
<tr>
<td>amharic</td>
<td>centralatlastamazight</td>
</tr>
<tr>
<td>arabic</td>
<td>centalkurdish</td>
</tr>
<tr>
<td>arabic-algeria</td>
<td>chechen</td>
</tr>
<tr>
<td>arabic-DZ</td>
<td>cherokee</td>
</tr>
<tr>
<td>arabic-morocco</td>
<td>chiga</td>
</tr>
<tr>
<td>arabic-MA</td>
<td>chinese-hans-hk</td>
</tr>
<tr>
<td>arabic-syria</td>
<td>chinese-hans-mo</td>
</tr>
<tr>
<td>arabic-SY</td>
<td>chinese-hans-sg</td>
</tr>
<tr>
<td>armenian</td>
<td>chinese-hans</td>
</tr>
<tr>
<td>assamese</td>
<td>chinese-hant-hk</td>
</tr>
<tr>
<td>asturian</td>
<td>chinese-hant-mo</td>
</tr>
<tr>
<td>asu</td>
<td>chinese-hant</td>
</tr>
<tr>
<td>australian</td>
<td>chinese-simplified-hongkongsarchina</td>
</tr>
<tr>
<td>austrian</td>
<td>chinese-simplified-macausarchina</td>
</tr>
<tr>
<td>azerbaijani-cyrillic</td>
<td>chinese-simplified-singapore</td>
</tr>
<tr>
<td>azerbaijani-cyrl</td>
<td>chinese-simplified</td>
</tr>
<tr>
<td>azerbaijani-latin</td>
<td>chinese-traditional-hongkongsarchina</td>
</tr>
<tr>
<td>azerbaijani-latn</td>
<td>chinese-traditional-macausarchina</td>
</tr>
<tr>
<td>azerbaijani</td>
<td>chinese-traditional</td>
</tr>
<tr>
<td>bafia</td>
<td>chinese</td>
</tr>
<tr>
<td>bambara</td>
<td>churchslavic</td>
</tr>
<tr>
<td>basaa</td>
<td>churchslavic-cyrs</td>
</tr>
<tr>
<td>basque</td>
<td>churchslavic-oldcyrillic</td>
</tr>
<tr>
<td>belarusian</td>
<td>churchslavic-glag</td>
</tr>
<tr>
<td>bemba</td>
<td>churchslavic-glotional</td>
</tr>
<tr>
<td>bena</td>
<td>colognian</td>
</tr>
<tr>
<td>bengali</td>
<td>cornish</td>
</tr>
<tr>
<td>bodo</td>
<td>croatian</td>
</tr>
<tr>
<td>bosnian-cyrillic</td>
<td>czech</td>
</tr>
<tr>
<td>bosnian-cyrl</td>
<td>danish</td>
</tr>
<tr>
<td>bosnian-latin</td>
<td>duala</td>
</tr>
<tr>
<td>bosnian-latn</td>
<td>dutch</td>
</tr>
<tr>
<td>bosnian</td>
<td>dzongkha</td>
</tr>
<tr>
<td>brazilian</td>
<td>embu</td>
</tr>
<tr>
<td>breton</td>
<td>english-au</td>
</tr>
<tr>
<td>british</td>
<td>english-australia</td>
</tr>
<tr>
<td>bulgarian</td>
<td>english-ca</td>
</tr>
</tbody>
</table>

---

13The name in the CLDR is Old Church Slavonic Cyrillic, but it has been shortened for practical reasons.
<table>
<thead>
<tr>
<th>Language</th>
<th>ISO Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>nepali</td>
<td>sanskrit-telu</td>
</tr>
<tr>
<td>newzealand</td>
<td>sanskrit-telugu</td>
</tr>
<tr>
<td>ngiemboon</td>
<td>sanskrit</td>
</tr>
<tr>
<td>ngomba</td>
<td>scottishgaelic</td>
</tr>
<tr>
<td>norsk</td>
<td>sena</td>
</tr>
<tr>
<td>northernluri</td>
<td>serbian-cyrillic-bosniaherzegovina</td>
</tr>
<tr>
<td>northernsami</td>
<td>serbian-cyrillic-kosovo</td>
</tr>
<tr>
<td>northndebele</td>
<td>serbian-cyrillic-montenegro</td>
</tr>
<tr>
<td>norwegianbokmal</td>
<td>serbian-cyrl</td>
</tr>
<tr>
<td>norwegiannynorsk</td>
<td>serbian-cyrl-ba</td>
</tr>
<tr>
<td>nsweissgerman</td>
<td>serbian-cyrl-me</td>
</tr>
<tr>
<td>nuer</td>
<td>serbian-cyrl-xk</td>
</tr>
<tr>
<td>nyankole</td>
<td>serbian-latin</td>
</tr>
<tr>
<td>nynorsk</td>
<td>serbian-latin-bosniaherzegovina</td>
</tr>
<tr>
<td>occitan</td>
<td>serbian-latin-montenegro</td>
</tr>
<tr>
<td>oriya</td>
<td>serbian-latin</td>
</tr>
<tr>
<td>oromo</td>
<td>serbian-latn</td>
</tr>
<tr>
<td>ossetic</td>
<td>serbian-latn-ba</td>
</tr>
<tr>
<td>pashto</td>
<td>serbian-latn-me</td>
</tr>
<tr>
<td>persian</td>
<td>serbian-latn-xk</td>
</tr>
<tr>
<td>piedmontese</td>
<td>serbian-latn</td>
</tr>
<tr>
<td>polish</td>
<td>shambala</td>
</tr>
<tr>
<td>portuguese-br</td>
<td>shona</td>
</tr>
<tr>
<td>portuguese-brazil</td>
<td>sichuanyi</td>
</tr>
<tr>
<td>portuguese-portugal</td>
<td>sinhala</td>
</tr>
<tr>
<td>portuguese-pt</td>
<td>slovak</td>
</tr>
<tr>
<td>portuguese</td>
<td>slovene</td>
</tr>
<tr>
<td>punjabi-arab</td>
<td>slovenian</td>
</tr>
<tr>
<td>punjabi-arabic</td>
<td>soga</td>
</tr>
<tr>
<td>punjabi-gurmukhi</td>
<td>somali</td>
</tr>
<tr>
<td>punjabi-guru</td>
<td>spanish-mexico</td>
</tr>
<tr>
<td>punjabi</td>
<td>spanish-mx</td>
</tr>
<tr>
<td>quechua</td>
<td>spanish</td>
</tr>
<tr>
<td>romanian</td>
<td>standardmoroccantamazight</td>
</tr>
<tr>
<td>romansh</td>
<td>swahili</td>
</tr>
<tr>
<td>rombo</td>
<td>swedish</td>
</tr>
<tr>
<td>rundi</td>
<td>swissgerman</td>
</tr>
<tr>
<td>russian</td>
<td>tachelhit-latin</td>
</tr>
<tr>
<td>rwa</td>
<td>tachelhit-latn</td>
</tr>
<tr>
<td>sakha</td>
<td>tachelhit-tfng</td>
</tr>
<tr>
<td>samburu</td>
<td>tachelhit-tifinagh</td>
</tr>
<tr>
<td>samin</td>
<td>tachelhit</td>
</tr>
<tr>
<td>sango</td>
<td>taita</td>
</tr>
<tr>
<td>sangu</td>
<td>tamil</td>
</tr>
<tr>
<td>sanskrit-beng</td>
<td>tasawaq</td>
</tr>
<tr>
<td>sanskrit-bengali</td>
<td>telugu</td>
</tr>
<tr>
<td>sanskrit-deva</td>
<td>teso</td>
</tr>
<tr>
<td>sanskrit-devanagari</td>
<td>thai</td>
</tr>
<tr>
<td>sanskrit-gujarati</td>
<td>tibetan</td>
</tr>
<tr>
<td>sanskrit-gujr</td>
<td>tigrinya</td>
</tr>
<tr>
<td>sanskrit-kannada</td>
<td>tongan</td>
</tr>
<tr>
<td>sanskrit-knda</td>
<td>turkish</td>
</tr>
<tr>
<td>sanskrit-malayalam</td>
<td>turkmen</td>
</tr>
<tr>
<td>sanskrit-mlym</td>
<td></td>
</tr>
</tbody>
</table>
Modifying and adding values to ini files

New 3.39 There is a way to modify the values of ini files when they get loaded with `\babelprovide` and `import`. To set, say, `digits.native` in the `numbers` section, use something like `numbers/digits.native=abcdefghij`. Keys may be added, too. Without `import` you may modify the identification keys.

This can be used to create private variants easily. All you need is to import the same ini file with a different locale name and different parameters.

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`.\[14\]

`\babelfont` \[⟨language-list⟩\]{⟨font-family⟩}{⟨font-options⟩}{⟨font-name⟩}

The main purpose of `\babelfont` is to define at once in a multilingual document the fonts required by the different languages, with their corresponding language systems (script and language). So, if you load, say, 4 languages, `\babelfont{rm}{FreeSerif}` defines 4 fonts (with their variants, of course), which are switched with the language by babel. It is a tool to make things easier and transparent to the user.

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

If no language is given, then it is considered the default font for the family, activated when a language is selected.

On the other hand, if there is one or more languages in the optional argument, the font will be assigned to them, overriding the default one. Alternatively, you may set a font for a script – just precede its name (lowercase) with a star (eg, *devanagari*). With this optional argument, the font is not yet defined, but just predeclared. This means you may define as many fonts as you want ‘just in case’, because if the language is never selected, the corresponding `\babelfont` declaration is just ignored.

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.

\[14\]See also the package `combofont` for a complementary approach.
\documentclass{article}
\usepackage[swedish, bidi=default]{babel}
\babelprovide[import]{hebrew}
\babelfont{rm}{FreeSerif}
\begin{document}
Svenska \foreignlanguage{hebrew}{עברית} svenska.
\end{document}

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

\babelfont{rm}{Iwona}
\babelfont{hebrew}{rm}{FreeSerif}

\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 families.

\textbf{EXAMPLE} Here is how to do it:

\babelfont{kai}{FandolKai}

Now, \texttt{kaiFamily} and \texttt{kaiDefault}, as well as \texttt{textkai} are at your disposal.

\textbf{NOTE} You may load fontspec explicitly. For example:

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

This makes sure the OpenType script for Devanagari is deva and not dev2, in case it is not detected correctly. You may also pass some options to fontspec: with silent, the warnings about unavailable scripts or languages are not shown (they are only really useful when the document format is being set up).

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

\textbf{NOTE} \texttt{fontspec} is not touched at all, only the preset font families (\texttt{rm}, \texttt{sf}, \texttt{tt}, and the like). If a language is switched when an \texttt{ad hoc} font is active, or you select the font with this command, neither the script nor the language is passed. You must add them by hand. This is by design, for several reasons—for example, each font has its own set of features and a generic setting for several of them could be problematic, and also a “lower-level” font selection is useful.
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 \babel\provide provides default values for \babel\font if omitted, but the opposite is not true. See the note above for the reasons of this behavior.

Using \setxxxxfont and \babel\font at the same time is discouraged, but very often works as expected. However, be aware with \setxxxxfont the language system will not be set by babel and should be set with fontspec if necessary.

This is not an error. This warning is shown by fontspec, not by babel. It could be irrelevant for English, but not for many other languages, including Urdu and Turkish. This is a useful and harmless warning, and if everything is fine with your document the best thing you can do is just to ignore it altogether.

This is not an error. babel assumes that if you are using \babel\font for a family, very likely you want to define the rest of them. If you don't, you can find some inconsistencies between families. This checking is done at the beginning of the document, at a point where we cannot know which families will be used.

Actually, there is no real need to use \babel\font in a monolingual document, if you set the language system in \setmainfont (or not, depending on what you want).

As the message explains, there is nothing intrinsically wrong with not defining all the families. In fact, there is nothing intrinsically wrong with not using \babel\font at all. But you must be aware that this may lead to some problems.

### 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:

  ```latex
  \addto\captionsenglish{%
  \renewcommand\contentsname{Foo}%
  }
  ```

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

- 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 \babel\provide and its key import, is:

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

- Macros to be run when a language is selected can be add to \extras⟨lang⟩:

  ```latex
  \addto\extrarussian{\mymacro}
  ```
There is a counterpart for code to be run when a language is unselected: ```\noextras\lang```

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

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

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

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

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

```\usepackage[danish]{babel} \babelprovide[\texttt{\captions=da,\hyphenrules=\texttt{nohyphenation}}]{danish}``` 

first loads danish.1df, and then redefines the captions for danish (as provided by the ini file) and prevents hyphenation. The rest of the language definitions are not touched.

### 1.16 Creating a language

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

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

If the language ```\langle\texttt{language-name}\rangle``` has not been loaded as class or package option and there are no ```\langle\texttt{options}\rangle```, it creates an “empty” one with some defaults in its internal structure: the hyphen rules, if not available, are set to the current ones, left and right hyphen mins are set to 2 and 3. In either case, caption, date and language system are not defined.

If no ini file is imported with ```\import{\langle\texttt{language-name}\rangle}``` it is still relevant because in such a case the hyphenation and like breaking rules (including those for South East Asian and CJK) are based on it as provided in the ini file corresponding to that name; the same applies to OpenType language and script.

Conveniently, some options allow to fill the language, and babel warns you about what to do if there is a missing string. Very likely you will find alerts like that in the log file:

```Package babel Warning: \mylangchaptername not set. Please, define \renewcommand\mylangchaptername{..} 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} ```
Locales with names based on BCP 47 codes can be created with something like:

\babelprovide[import=en-US]{enUS}

Note, however, mixing ways to identify locales can lead to problems. For example, is yi the name of the language spoken by the Yi people or is it the code for Yiddish?

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.

\begin{quote}
import = \langle language-tag \rangle
\end{quote}

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

\begin{quote}
\babelprovide[import=hu]{hungarian}
\end{quote}

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

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

\begin{quote}
\babelprovide[import]{hungarian}
\end{quote}

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 (french, breton, and occitan). Besides \today, this option defines an additional command for dates: \langle language\rangle date, which takes three arguments, namely, year, month and day numbers. In fact, \today calls \langle language\rangle today, which in turn calls \langle language\rangle date{\the\year}{\the\month}{\the\day}.

\begin{quote}
captions = \langle language-tag \rangle
\end{quote}

Loads only the strings. For example:

\begin{quote}
\babelprovide[captions=hu]{hungarian}
\end{quote}

\begin{quote}
hyphenrules = \langle language-list \rangle
\end{quote}

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{quote}
\babelprovide[hyphenrules=chavacano spanish italian]{chavacano}
\end{quote}
If none of the listed hyphen rules 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 \babelpatterns, as for example:

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

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

\texttt{main}

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

\texttt{script= \{script-name\}}

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

\texttt{language= \{language-name\}}

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

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

\texttt{onchar= ids | fonts}

\texttt{New 3.38} This option is much like an ‘event’ called when a character belonging to the script of this locale is found. There are currently two ‘actions’, which can be used at the same time (separated by a space): with ids the \language and the \localeid are set to the values of this locale; with fonts, the fonts are changed to those of this locale (as set with \babelfont). This option is not compatible with mapfont. Characters can be added with \babelchar property.

\texttt{mapfont= direction}

Assigns the font for the writing direction of this language (only with bidi=basic). Whenever possible, instead of this option use onchar, based on the script, which usually makes more sense. More precisely, what \texttt{mapfont=direction} means is, ‘when a character has the same direction as the script for the “provided” language, then change its font to that set for this language’. There are 3 directions, following the bidi Unicode algorithm, namely, Arabic-like, Hebrew-like and left to right. So, there should be at most 3 directives of this kind.

\texttt{intraspace= \{base\} \{shrink\} \{stretch\}}

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

\texttt{intrapenalty\{penalty\}}

Sets the interword penalty for the writing system of this language. Currently used only in Southeast Asian scripts, like Thai. Ignored if 0 (which is the default value).

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

1.17 Digits and counters

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

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

Languages providing native digits in all or some variants are:

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Persian</th>
<th>Lao</th>
<th>Odia</th>
<th>Urd</th>
<th>Lao</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assamese</td>
<td>Gujarati</td>
<td>Northern Luri</td>
<td>Punjabi</td>
<td>Uzbek</td>
<td>Lao</td>
</tr>
<tr>
<td>Bangla</td>
<td>Hindi</td>
<td>Malayalam</td>
<td>Pashto</td>
<td>Vai</td>
<td>Uzbek</td>
</tr>
<tr>
<td>Tibetan</td>
<td>Khmer</td>
<td>Marathi</td>
<td>Tamil</td>
<td>Cantonese</td>
<td>Vai</td>
</tr>
<tr>
<td>Bodo</td>
<td>Kannada</td>
<td>Burmese</td>
<td>Telugu</td>
<td>Chinese</td>
<td>Tibetan</td>
</tr>
<tr>
<td>Central Kurdish</td>
<td>Konkani</td>
<td>Mazanderani</td>
<td>Thai</td>
<td>Chinese</td>
<td>Burmese</td>
</tr>
<tr>
<td>Dzongkha</td>
<td>Kashmiri</td>
<td>Nepali</td>
<td>Uyghur</td>
<td>Chinese</td>
<td>Nepali</td>
</tr>
</tbody>
</table>

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

New 4.41 Many ‘ini’ locale files has been extended with information about non-positional numerical systems, based on those predefined in CSS. They only work with xetex and luatex and are fully expendable (even inside an \texttt{edef}). Currently, they are limited to numbers below 10000.

There are several ways to use them (for the available styles in each language, see the list below):

- \texttt{localenumeral\{\texttt{style}\}\{\texttt{number}\}}, like \texttt{localenumeral\{abjad\}\{15\}}
• \localecounter{\{style\}\{counter\}}, like \localecounter{lower}{section}

• In \babelprovide, as an argument to the keys alph and Alph, which redefine what \alph and \Alph print. For example:

\babelprovide[alph=alphabetic]{thai}

The styles are:

- **Ancient Greek** lower.ancient, upper.ancient
- **Arabic** abjad, maghrebi.abjad
- **Belarusan, Bulgarian, Macedonian, Serbian** lower, upper
- **Hebrew** letters (neither gershd nor gershyim yet)
- **Hindi** alphabetic
- **Armenian** lower, upper
- **Japanese** hiragana, hiragana.iroha, katakana, katakana.iroha, circled.katakana, informal, formal, cjk-earthly-branch, cjk-heavenly-stem, fullwidth.lower.alpha, fullwidth.upper.alpha
- **Georgian** letters
- **Greek** lower.modern, upper.modern, lower.ancient, upper.ancient (all with keraia)
- **Khmer** consonant
- **Korean** consonant, syllable, hanja.informal, hanja.formal, hangul.formal, cjk-earthly-branch, cjk-heavenly-stem, fullwidth.lower.alpha, fullwidth.upper.alpha
- **Persian** abjad, alphabetic
- **Russian** lower, lower.full, upper, upper.full
- **Tamil** ancient
- **Thai** alphabetic
- **Ukrainian** lower, lower.full, upper, upper.full
- **Chinese** cjk-earthly-branch, cjk-heavenly-stem, fullwidth.lower.alpha, fullwidth.upper.alpha

### 1.18 Accessing language info

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

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

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

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

\localeinfo {\{field\}}

**New 3.38** If an \ini file has been loaded for the current language, you may access the information stored in it. This macros is fully expandable and the available fields are:

name.english as provided by the Unicode CLDR.
tag.ini is the tag of the ini file (the way this file is identified in its name).
tag.bcp47 is the BCP 47 language tag.
tag.opentype is the tag used by OpenType (usually, but not always, the same as BCP 47).

script.name as provided by the Unicode CLDR.
script.tag.bcp47 is the BCP 47 language tag of the script used by this locale.
script.tag.opentype is the tag used by OpenType (usually, but not always, the same as BCP 47).

\getlocaleproperty ⟨macro⟩{⟨locale⟩}{⟨property⟩}

New 3.42 The value of any locale property as set by the ini files (or added/modified with \babelprovide) can be retrieved and stored in a macro with this command. For example, after:

\getlocaleproperty \hechap{hebrew}{captions/chapter}

the macro \hechap will contain the string קרפ.

Babel remembers which ini files have been loaded. There is a loop named \LocaleForEach to traverse the list, where #1 is the name of the current item, so that \LocaleForEach{\message{ **#1** }} just shows the loaded ini's.

NOTE ini files are loaded with \babelprovide and also when languages are selected if there is a \babelfont. To ensure the ini files are loaded (and therefore the corresponding data) even if these two conditions are not met, write \BabelEnsureInfo in the preamble.

1.19 Hyphenation and line breaking

Babel deals with three kinds of line breaking rules: Western, typically the LGC group, South East Asian, like Thai, and CJK, but support depends on the engine: pdftex only deals with the former, xetex also with the second one, while luatex provides basic rules for the latter, too.

\babelhyphen *{(type)}
\babelhyphen *{(text)}

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

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

- \babelhyphen{soft} and \babelhyphen{hard} are self explanatory.
- \babelhyphen{repeat} inserts a hard hyphen which is repeated at the beginning of the next line, as done in languages like Polish, Portuguese and Spanish.
\babelhyphenation{nobreak} inserts a hard hyphen without a break after it (even if a space follows).

\babelhyphenation{empty} inserts a break opportunity without a hyphen at all.

\babelhyphenation{⟨text⟩} is a hard “hyphen” using ⟨text⟩ instead. A typical case is \babelhyphenation{⟨⟩}.

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

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

There are also some differences with \TeX: (1) the character used is that set for the current font, while in \TeX it is hardwired to - (a typical value); (2) the hyphen to be used in fonts with a negative \hyphenchar is -, like in \TeX, but it can be changed to another value by redefining \babelnullhyphen; (3) a break after the hyphen is forbidden if preceded by a glue >0 pt (at the beginning of a word, provided it is not immediately preceded by, say, a parenthesis).

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

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

\small
\begin{verbatim}
\babelhyphenation{Wal-hal-la Dar-bhan-ga}
\end{verbatim}

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

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

\babelpatterns{⟨language⟩, ⟨language⟩, …}{⟨patterns⟩}

**New 3.9m** In luatex only adds or replaces patterns for the languages given or, without the optional argument, for all languages. If a pattern for a certain combination already exists, it gets replaced by the new one. It can be used only in the preamble, and patterns are added when the language is first selected, thus taking into account changes of \lccodes’s done in \extras⟨lang⟩ as well as the language-specific encoding (not set in the preamble by default). Multiple \babelpatterns’s are allowed.

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

---

15\ With luatex exceptions and patterns can be modified almost freely. However, this is very likely a task for a separate package and babel only provides the most basic tools.
New 3.31 (Only luatex.) With \babelprovide and imported CJK languages, a simple
generic line breaking algorithm (push-out-first) is applied, based on a selection of the
Unicode rules (New 3.32) it is disabled in verbatim mode, or more precisely when the
hyphenrules are set to nohyphenation). It can be activated alternatively by setting
explicitly the intraspace.

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

New 3.37-3.39 With luatex it is now possible to define non-standard hyphenation rules,
like \(f-f \rightarrow ff-f\), repeated hyphens, ranked ruled (or more precisely, ‘penalized’
hyphenation points), and so on. No rules are currently provided by default, but they can be
defined as shown in the following example, where \{1\} is the first captured char (between
\()\ in the pattern):

\begin{verbatim}
\b babelposthyphenation{german}{([fmtrp]) | {1}}
  { no = {1}, pre = {1}{1}- }, % Replace first char with disc
  remove, % Remove automatic disc (2nd node)
  {} % Keep last char, untouched
\}
\end{verbatim}

In the replacements, a captured char may be mapped to another, too. For example, if the
first capture reads \{[\text{ā}\text{ò}]\}, the replacement could be \{\text{	ext{ā}}\text{	ext{ò}}\}, which maps \text{	ext{ò}}\to\text{	ext{ò}}\, and \text{	ext{ò}}\ to \text{	ext{ò}}, so that the diaeresis is removed.
This feature is activated with the first \b babelposthyphenation.
See the [babel wiki] for a more detailed description and some examples. It also describes an
additional replacement type with the key string.

**EXAMPLE** Although the main purpose of this command is non-standard hyphenation, it
may actually be used for other transformations (after hyphenation is applied, so you
must take discretionaries into account). For example, you can use the string
replacement to replace a character (or series of them) by another character (or series of
them). Thus, to enter \text{	ext{ā}} as \text{zh} and \text{	ext{ò}} as \text{sh} in a newly created locale for transliterated
Russian:

\begin{verbatim}
\b babelprovide[hyphenrules=+]{russian-latin} % Create locale
\b babelposthyphenation{russian-latin}{([sz])h} % Create rule
  { string = {1|sz|sz} },
  remove
\}
\end{verbatim}

In other words, it is a quite general tool. (A counterpart \b babelprehyphenation is on
the way.)

1.20 Selecting scripts
Currently babel provides no standard interface to select scripts, because they are best
selected with either \fontencoding (low-level) or a language name (high-level). Even the
Latin script may require different encodings (ie, sets of glyphs) depending on the language, and therefore such a switch would be in a sense incomplete\textsuperscript{16}

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\textsuperscript{17}

\texttt{\textbackslash ensureasci i \{\textbackslash text\}}

**New 3.9i** This macro makes sure \texttt{(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 \texttt{\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, \texttt{\ensureascii} switches to the encoding at the beginning of the document if ASCII-savvy, or else the last ASCII-savvy encoding loaded. For example, if you load LY1, LGR, then it is set to LY1, but if you load LY1, T2A it is set to T2A. The symbol encodings TS1, T3, and TS3 are not taken into account, since they are not used for “ordinary” text (they are stored in \texttt{\BabelNonText}, used in some special cases when no Latin encoding is explicitly set).

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

### 1.21 Selecting directions

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

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

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

**WARNING** If characters to be mirrored are shown without changes with lualatex, try with the following line:

\texttt{\textbackslash ensureasci i \{\textbackslash text\}}

\textsuperscript{16}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.

\textsuperscript{17}But still defined for backwards compatibility.
There are some package options controlling bidi writing.

\begin{verbatim}
\texttt{\textbackslash babeladjust\{bidi.mirroring=off\}}
\end{verbatim}

There are some package options controlling bidi writing.

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

\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 \texttt{xetex} and \texttt{pdftex} this is the only option.

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

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

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

\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.

Remember \texttt{basic} is available in \texttt{luatex} only.

\begin{verbatim}
\documentclass{article}
\usepackage[bidi=basic]{babel}
\babelprovide[import, main]{arabic}
\babelfont{rm}{FreeSerif}
\begin{document}
\textquotedblleftـﺑ(ﻲﻘﻳﺮﻏﻻا)ﻲﻨﻴﻠﻴﻬﻟاﺮﺼﻌﻟاﺔﻠﻴﻃبﺮﻌﻟاةﺮﻳﺰﺟﻪﺒﺷﺖﻓﺮﻋﺪﻗو
Arabia
وأ
Aravia )
ﺔﻴﻘﻳﺮﻏﻻﺎﺑ
Αραβία (ثﻼﺛنﺎﻣوﺮﻟامﺪﺨﺘﺳا،
"ـﺑتﺎﺋدﺎﺑ
Arabia"
ﺎﻬﻧأﻻإ،ﺔﻴﺑﺮﻌﻟاةﺮﻳﺰﺠﻟاﻪﺒﺷﻦﻣﻖﻃﺎﻨﻣثﻼﺛﻰﻠﻋ
مﻮﻴﻟاﻪﻴﻠﻋفﺮﻌﺗﺎﻤﻣﺮﺒﻛأﺖﻧﺎﻛًﺔﻘﻴﻘﺣ
\end{document}
\end{verbatim}

\textbf{EXAMPLE} With \texttt{bidi=basic both} L and R text can be mixed without explicit markup (the latter will be only necessary in some special cases where the Unicode algorithm fails). It is used much like \texttt{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 \texttt{\textbackslash babelprovide}, as illustrated:

\begin{verbatim}
\documentclass{book}
\usepackage[english, bidi=basic]{babel}
\babelprovide[onchar=ids fonts]{arabic}
\end{document}
\end{verbatim}
Most Arabic speakers consider the two varieties to be two registers of one language, although the two registers can be referred to in Arabic as \textit{fuṣḥā l-ʻaṣr} (MSA) and \textit{fuṣḥā t-turāth} (CA).

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

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

\newcommand\refrange[2]{\babelsublr{\text{the}{\ref{#1}}-\text{the}{\ref{#2}}}}

In the future a more complete method, reading recursively boxed text, may be added.
**WARNING** As of April 2019 there is a bug with \parshape in \texttt{luatex} (a \TeX{} primitive) which makes lists to be horizontally misplaced if they are inside a \texttt{vbox} (like \texttt{minipage}) and the current direction is different from the main one. A workaround is to restore the main language before the box and then set the local one inside.

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

\textbf{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 R by default if the main language is R (including \texttt{multicol}).

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

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

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

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

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

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

\begin{verbatim}
\usepackage[bidi=basic,
            layout=counters.tabular]{babel}
\end{verbatim}

\begin{verbatim}
\babelsublr \{\langle lr-text\rangle\}
\end{verbatim}

Digits in \texttt{pdftex} must be marked up explicitly (unlike \texttt{luatex} with \texttt{bidi=basic} or \texttt{bidi=basic-r} and, usually, \texttt{xetex}). This command is provided to set \{\langle lr-text\rangle\} 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 \texttt{RL} counterpart.

Any \texttt{\babelsublr} in \texttt{explicit} L mode is ignored. However, with \texttt{bidi=basic} and \texttt{implicit} L, it first returns to R and then switches to explicit L. To clarify this point, consider, in an R context:

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

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

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

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 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

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 (e.g., \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 \left\langle \langle \text{lang} \rangle \right\rangle \left\langle \langle \text{name} \rangle \right\rangle \left\langle \langle \text{event} \rangle \right\rangle \left\langle \langle \text{code} \rangle \right\rangle

The same name can be applied to several events. Hooks may be enabled and disabled for all defined events with \EnableBabelHook \langle \text{name} \rangle, \DisableBabelHook \langle \text{name} \rangle. Names containing the string babel are reserved (they are used, for example, by \useshortands* to add a hook for the event afterextras). **New 3.33** They may also be applied to a specific language with the optional argument; language-specific settings are executed after global ones.

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

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

\AddBabelHook \langle \langle \text{no} \rangle \rangle \langle \langle \text{after} \rangle \rangle \langle \langle \text{extras} \rangle \rangle \langle \langle \text{language} \rangle \rangle \langle \langle \text{shorthands} \rangle \rangle \langle \langle \text{none} \rangle \rangle

- **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 \langle \langle \text{myhook} \rangle \rangle \langle \langle \text{stringprocess} \rangle \rangle \langle \% \rangle \langle \protected@edef \BabelString {\BabelString} \rangle

- **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 \text{language} \rangle and \date \langle \text{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 file or to load nothing.

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

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

\BabelContentsFiles

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

### 1.24 Languages supported by babel with ldf files

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

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

42
There are more languages not listed above, including hindi, thai, thaicjk, latvian, turkmen, magyar, mongolian, romansh, lithuanian, spanglish, vietnamese, japanese, pinyin, arabic, farsi, ibygreek, bgreek, serbianc, frenchie, ethiop, and friulan.

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

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

Then you preprocess it with devnag \textlangle file\textrangle, which creates \textlangle file\textrangle .tex; you can then typeset the latter with \textLaTeX.

### 1.25 Unicode character properties in \textLaTeX

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

\begin{verbatim}
\babelcharproperty \langle char-code \rangle \langle to-char-code \rangle \langle property \rangle \langle value \rangle
\end{verbatim}

Here, \langle char-code \rangle is a number (with TeX syntax). With the optional argument, you can set a range of values. There are three properties (with a short name, taken from Unicode): direction (bc), mirror (bmg), linebreak (lb). The settings are global, and this command is allowed only in vertical mode (the preamble or between paragraphs).

For example:

```latex
\babelcharproperty{`¿}{mirror}{`?}
\babelcharproperty{`-}{direction}{1} % or al, r, en, an, on, et, cs
\babelcharproperty{`‘}{linebreak}{cl} % or id, op, cl, ns, ex, in, hy
```

Another property is locale, which adds characters to the list used by onchar in \textbackslash babelprovide, or, if the last argument is empty, removes them. The last argument is the locale name:

```latex
\babelcharproperty{`‘}{locale}{english}
```

### 1.26 Tweaking some features
\babeladjust{(key-value-list)}

New 3.36  Sometimes you might need to disable some babel features. Currently this macro understands the following keys (and only for \LaTeX), with values on or off: bidi.text, bidi.mirroring, bidi.mapdigits, layout.lists, layout.tabular, linebreak.sea, linebreak.cjk. For example, you can set \texttt{\babeladjust{bidi.text=off}} if you are using an alternative algorithm or with large sections not requiring it. With \texttt{luatex} you may need \texttt{bidi.mirroring=off}. Use with care, because these options do not deactivate other related options (like paragraph direction with \texttt{bidi.text}).

1.27  **Tips, workarounds, known issues and notes**

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

- Both \texttt{ltxdoc} and babel use \texttt{\AtBeginDocument} to change some catcodes, and babel reloads \texttt{hhline} to make sure : has the right one, so if you want to change the catcode of \texttt{|} it has to be done using the same method at the proper place, with

\begin{verbatim}
\AtBeginDocument{\DeleteShortVerb{|}}
\end{verbatim}

\textit{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 \texttt{hhline} (babel, now with the correct catcodes for \texttt{|} and \texttt{:}).

- 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:

\begin{verbatim}
\addto\extrasfrench{\inputencoding{latin1}}
\addto\extrarussian{\inputencoding{koi8-r}}
\end{verbatim}

(A recent version of \texttt{inputenc} is required.)

- For the hyphenation to work correctly, lccodes cannot change, because \LaTeX{} only takes into account the values when the paragraph is hyphenated, i.e., when it has been finished\footnote{This explains why \LaTeX{} assumes the lowercase mapping of T1 and does not provide a tool for multiple mappings. Unfortunately, \texttt{\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 \texttt{\foreignlanguage}, the apostrophes might not be taken into account. This is a limitation of \LaTeX{}, not of babel. Alternatively, you may use \texttt{\useshortshands} to activate \texttt{'} and \texttt{\defineshorthand}, or redefine \texttt{\textquoteright} (the latter is called by the non-ASCII right quote).

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

- Babel does not take into account \texttt{\normalsfcodes} and (non-)French spacing is not always properly (un)set by languages. However, problems are unlikely to happen and therefore this part remains untouched in version 3.9 (but it is in the 'to do' list).
• Using a character mathematically active (ie, with math code "8000") as a shorthand can make \TeX enter in an infinite loop in some rare cases. (Another issue in the ‘to do’ list, although there is a partial solution.)

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

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

1.28 Current and future work

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

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

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

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

An option to manage bidirectional document layout in luatex requires patching somehow lots of macros and packages (and some issues related to \specials remain, like color and hyperlinks), so babel resorts to the bidi package (by Vafa Khalighi). See the babel repository for a small example (xe-bidi).

1.29 Tentative and experimental code

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

Old and deprecated stuff

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

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

\footnote{See for example POSIX, ISO 14652 and the Unicode Common Locale Data Repository (CLDR). Those systems, however, have limited application to \TeX because their aim is just to display information and not fine typesetting.}
• \babelFSdefault{\fontspecfeatures} patches \fontspec so that the given features are always passed as the optional argument or added to it (not an ideal solution).

So, for example:

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

2 Loading languages with language.dat

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

New 3.9q With luatex, however, patterns are loaded on the fly when requested by the language (except the “0th” language, typically english, which is preloaded always). Until 3.9n, this task was delegated to the package luatex-hyphen, by Khaled Hosny, Élie Roux, and Manuel Pégourié-Gonnard, and required an extra file named language.dat.lua, but now a new mechanism has been devised based solely on language.dat. You must rebuild the formats if upgrading from a previous version. You may want to have a local language.dat.dat for a particular project (for example, a book on Chemistry).

2.1 Format

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

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

\begin{verbatim}
% File : language.dat
% Purpose : tell ini\TeX what files with patterns to load.
english english.hyphenations
=british

dutch hyphen.dutch.exceptions.dutch % Nederlands
german hyphen.ger
\end{verbatim}

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:

\begin{verbatim}
% File : language.dat
% Purpose : tell ini\TeX what files with patterns to load.
english:encoding=latin1 english.hyphenations
=british

dutch:encoding=latin1 hyphen.dutch.exceptions.dutch % Nederlands
german:encoding=latin1 hyphen.ger
\end{verbatim}

21 This feature was added to 3.9a, but it was buggy. Both 3.9a and 3.9p are deprecated.
22 The loader for lua(e)tex is slightly different as it's not based on babel but on etex .src. Until 3.9p it just didn't work, but thanks to the new code it works by reloading the data in the babel way, i.e., with language.dat.
23 This is because different operating systems sometimes use very different file-naming conventions.
24 This is not a new feature, but in former versions it didn't work correctly.
With the previous settings, if the encoding when the language is selected is T1 then the patterns in hyphen\texttt{T1}.ger are used, but otherwise use those in hyphen\texttt{.ger} (note the encoding could be set in \texttt{\set{	extasciitilde}{lang}}). A typical error when using babel is the following:

\begin{quote}
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}
\end{quote}

It simply means you must reconfigure \texttt{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 \texttt{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\texttt{.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 \texttt{\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 \texttt{\langle\langle lang\rangle\rangle}, \texttt{\langle captions\langle lang\rangle\rangle}, \texttt{\langle date\langle lang\rangle\rangle}, \texttt{\langle extras\langle lang\rangle\rangle} and \texttt{\langle noextras\langle lang\rangle\rangle}(the last two may be left empty); where \texttt{\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, \texttt{\langle date\langle lang\rangle\rangle} but not \texttt{\langle captions\langle lang\rangle\rangle} does not raise an error but can lead to unexpected results.

- When a language definition file is loaded, it can define \texttt{\language{\langle lang\rangle}} to be a dialect of \texttt{\language{0}} when \texttt{\language{\langle lang\rangle}} is undefined.

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

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

• The preferred shorthand is ", which is not used in \TeX (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.).

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

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

• Do not switch scripts. If you want to make sure a set of glyphs is used, switch either the font encoding (low-level) or the language (high-level, which in turn may switch the font encoding). Usage of things like \latintext is deprecated.²⁵

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

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

3.1 Guidelines for contributed languages

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

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

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

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

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

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

²⁵But not removed, for backward compatibility.
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 \language. Here “language” is used in the \TeX sense of set of hyphenation patterns.

\langle lang\rangle hyphenmins The macro \langle lang\rangle 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\langle lang\rangle 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\langle lang\rangle The macro \captions\langle lang\rangle defines the macros that hold the texts to replace the original hard-wired texts.

\date\langle lang\rangle The macro \date\langle lang\rangle defines \today.

\extras\langle lang\rangle The macro \extras\langle lang\rangle 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\langle lang\rangle 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\langle lang\rangle, a macro that brings \TeX into a predefined state is needed. It will be no surprise that the name of this macro is \noextras\langle lang\rangle.

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

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

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

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

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

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

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

\substitutefontfamily

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

### 3.3 Skeleton

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

```
\ProvidesLanguage{<language>}
    [2016/04/23 v0.0 <Language> support from the babel system]
\LdfInit{<language>}{captions<language>}
\ife\undefined\l@<language>
    \@nopatterns{<Language>}
\adddialect\l@<language>0
\fi
\adddialect\l@<dialect>\l@<language>
\bbl@declare@attribute{<language>}{<attrib>}{% 
    \expandafter\addto\expandafter\extras<language>
    \expandafter\extras<attrib><language>\% 
    \let\captions<language>\captions<attrib><language>\% 
\providehyphenmins{<language>}{\tw@@\thr@@}
\StartBabelCommands*{<language>}{captions}
\SetString\chaptername{<chapter name>}
% More strings
\StartBabelCommands*{<language>}{date}
\SetString\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>}
% More strings
\EndBabelCommands
\addto\extras<language>{}
\addto\noextras<language>{}
\let\extras<dialect>\extras<language>{}
\let\noextras<dialect>\noextras<language>{}
\ldf@finish{<language>}
```

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

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

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.

- \texttt{\initiate@active@char}\ The internal macro \texttt{\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.
- \texttt{\bbl@activate}\ The command \texttt{\bbl@activate} is used to change the way an active character expands. \texttt{\bbl@activate} ‘switches on’ the active behavior of the character. \texttt{\bbl@deactivate} lets the active character expand to its former (mostly) non-active self.
- \texttt{\declare@shorthand}\ The macro \texttt{\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”.)

- \texttt{\bbl@add@special}\ \texttt{\bbl@remove@special}\ The \TeX{}book states: “Plain \TeX{} includes a macro called \texttt{\dospecials} that is essentially a set macro, representing the set of all characters that have a special category code.” [4, 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 \texttt{\dospecial}. \TeX{} adds another macro called \texttt{@sanitize} representing the same character set, but without the curly braces. The macros \texttt{\bbl@add@special\langle char\rangle} and \texttt{\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.\footnote{This mechanism was introduced by Bernd Raichle.}

- \texttt{\babel@save}\ To save the current meaning of any control sequence, the macro \texttt{\babel@save} is provided. It takes one argument, \langle csname\rangle, the control sequence for which the meaning has to be saved.
- \texttt{\babel@savevariable}\ A second macro is provided to save the current value of a variable. In this context, anything that is allowed after the \texttt{\the} primitive is considered to be a variable. The macro takes one argument, \langle variable\rangle.

The effect of the preceding macros is to append a piece of code to the current definition of \texttt{\originalTeX}. When \texttt{\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

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

3.7 Macros common to a number of languages

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

The `\langle category \rangle` is either captions, date or extras. You must stick to these three categories, even if no error is raised when using other name. 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\monthiiname{Februar}
  \SetString\monthiiiname{M"{a}rz}
  \SetString\monthivname{April}
```

\footnotesize

\footnotesize
27 In future releases further categories may be added.

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

\StartBabelCommands \*\langle language-list\rangle\langle category\rangle\langle selector\rangle
The starred version just forces strings to take a value – if not set as package option, then the default for the engine is used. This is not done by default to prevent backward incompatibilities, but if you are creating a new language this version is better. It’s up to the maintainers of the current languages to decide if using it is appropriate.\footnote{\textsuperscript{28}}
\EndBabelCommands

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

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

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

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

#1 is replaced by the roman numeral.

\footnote{\textsuperscript{28}}This replaces in 3.9g a short-lived \UseStrings which has been removed because it did not work.
Sets globally code to be executed at \MakeUppercase and \MakeLowercase. The code would typically be things like \let\BB\bb and \uccode or \lccode (although for the reasons explained above, changes in lc/uc codes may not work). A \langle map-list \rangle is a series of macros using the internal format of \@ucclist (eg, \bb\BB\cc\CC). The mandatory arguments take precedence over the optional one. This command, unlike \SetString, is executed always (even without strings), and it is intended for minor readjustments only. For example, as T1 is the default case mapping in \LaTeX, we could set for Turkish:

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

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

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

(Note the mapping for OT1 is not complete.)

\SetHyphenMap
 \langle to-lower-macros \rangle

\BabelLower{\langle uccode \rangle}{\langle lccode \rangle}

\BabelLowerMM{\langle uccode-from \rangle}{\langle uccode-to \rangle}{\langle step \rangle}{\langle lccode-from \rangle}

\BabelLowerMO{\langle uccode-from \rangle}{\langle uccode-to \rangle}{\langle step \rangle}{\langle lccode \rangle}

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 xetex and luatex) – if an assignment is wrong, fix it directly.

4 Changes

4.1 Changes in babel version 3.9

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

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

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

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

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

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

- \textormath raised an error with a conditional.

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

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

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

Part II

Source code

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

6 locale directory

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

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

This is a preliminary documentation.

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

Most keys are self-explanatory.

- **charset** the encoding used in the ini file.
- **version** of the ini file
- **level** “version” of the ini specification, which keys are available (they may grow in a compatible way) and how they should be read.
- **encodings** a descriptive list of font encodings.
- **[captions]** section of captions in the file charset
- **[captions.lcr]** same, but in pure ASCII using the LICR
- **date.long** fields are as in the CLDR, but the syntax is different. Anything inside brackets is a date field (eg, MMMM for the month name) and anything outside is text. In addition, [ ] is a non breakable space and [ . ] is an abbreviation dot.

Keys may be further qualified in a particular language with a suffix starting with a uppercase letter. It can be just a letter (eg, babel.name.A, babel.name.B) or a name (eg, date.long.Nominative, date.long.Formal, but no language is currently using the latter).

Multi-letter qualifiers are forward compatible in the sense they won’t conflict with new “global” keys (all lowercase).

7 Tools

1 ⟨⟨version=3.42⟩⟩
2 ⟨⟨date=2020/03/22⟩⟩
Do not use the following macros in \lbf 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}
3 ⟨⟨ Basic macros ⟩⟩ ≡
4 \bbl@trace{Basic macros}
5 \def\bbl@stripslash{\expandafter\@gobble\string}
6 \def\bbl@add#1#2{%
7    \bbl@ifunset{\bbl@stripslash#1}%
8    {\def#1{#2}}%
9    {\expandafter\def\expandafter#1\expandafter{#1#2}}
10 \def\bbl@xin@{\@expandtwoargs\in@}
11 \def\bbl@csarg#1#2{\expandafter#1\csname bbl@#2\endcsname}%
12 \def\bbl@cs#1{\csname bbl@#1\endcsname}
13 \def\bbl@cl#1{\csname bbl@#1@\languagename\endcsname}
14 \def\bbl@loop#1#2#3{\bbl@@loop#1{#3}#2,\@nnil,}
15 \def\bbl@loopx#1#2{\expandafter\bbl@loop\expandafter#1\expandafter{#2}}
16 \def\bbl@loop@loop#1#2#3,{%
17    \ifx\@nnil#3\relax\else
18    \def#1{#3}#2\bbl@afterfi\bbl@@loop#1{#2}%
19 \fi}
20 \def\bbl@for#1#2#3{\bbl@loopx#1{#2}{\ifx#1\@empty\else#3\fi}}
\end{verbatim}

\bbl@addlist 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.

\begin{verbatim}
21 \def\bbl@addlist#1#2{%
22   \edef#1%{
23      \bbl@ifunset{\bbl@stripslash#1}{%}{\ifx#1\@empty\else#1,\fi}{%}
24        {ifs#1@empty\else1,\fi}%
25   #2})
\end{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[29]. These macros will break if another \if...\fi statement appears in one of the arguments and it is not enclosed in braces.

\begin{verbatim}
27 \long\def\bbl@afterelse#1\else#2\fi{\fi#1}
28 \long\def\bbl@afterfi#1\fi{\fi#1}
\end{verbatim}

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

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

[29] This code is based on code presented in TUGboat vol. 12, no2, June 1991 in “An expansion Power Lemma” by Sonja Maus.
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.

```latex
\def\bbl@tempa#1{\long\def\bbl@trim##1##2{\futurelet\bbl@trim@a\bbl@trim@c##2\@nil\@nil#1\@nil\relax{##1}}\def\bbl@trim@c{\ifx\bbl@trim@a\@sptoken\expandafter\bbl@trim@b\else\expandafter\bbl@trim@b\expandafter#1\fi}}\bbl@tempa{ }
\long\def\bbl@trim@i#1\@nil#2\relax#3{#3{#1}}\long\def\bbl@trim@def#1{\bbl@trim{\def#1}}
```

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.

```latex
\begingroup\gdef\bbl@ifunset#1{\expandafter\ifx\csname#1\endcsname\relax\expandafter\@firstoftwo\else\expandafter\@secondoftwo\fi}
\bbl@ifunset{ifcsname}{}
\endgroup
```

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

```latex
\def\bbl@ifblank#1{\bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
\long\def\bbl@ifblank@i#1#2\@nil#3#4#5\@nil{#4}
```

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

```latex
\def\bbl@forkv#1#2{% \def\bbl@kvcmd##1##2##3{\@empty\ifx\bbl@kvcmd\@nil\@nil\else\bbl@kvnext##1,\@nil,\fi} \def\bbl@kvnext#1,{% \ifx\@nil\relax\else
```
A for loop. Each item (trimmed), is \#1. It cannot be nested (it's doable, but we don't need it).

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

An extension to the previous macro. It takes into account the parameters, and it is string based (ie, if you replace \relax by ho, then \relax becomes \rho). No checking is done at all, because it is not a general purpose macro, and it is used by babel only when it works (an example where it does not work is in \bbl@TG@@date, and also fails if there are macros with spaces, because they are retokenized). It may change! (or even merged with \bbl@replace; I'm not sure checking the replacement is really necessary or just paranoia).

\edef\bbl@exp\detokenize\@undefined% Unused macros if old Plain TeX
\def\bbl@exp{\detokenize{macro:}}#2->#3%
Two further tools. \texttt{bbl@samestring} first expand its arguments and then compare their expansion (sanitized, so that the catcodes do not matter). \texttt{bbl@engine} takes the following values: 0 is \TeX, 1 is \LaTeX, and 2 is \texttt{xetex}. You may use the latter in your language style if you want.

\begin{verbatim}
\def\bbl@ifsamestring#1#2{% 
\begingroup 
\protected@edef\bbl@tempb{#1}% 
\edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}% 
\protected@edef\bbl@tempc{#2}% 
\edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}% 
\ifx\bbl@tempb\bbl@tempc 
\aftergroup\@firstoftwo 
\else 
\aftergroup\@secondoftwo 
\fi 
\endgroup}
\end{verbatim}

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

\begin{verbatim}
\if\directlua\@undefined\else
\if\bbl@luapatterns\@undefined
\input luababel.def
\fi\fi
\end{verbatim}

The following code is used in \texttt{babel.def} and \texttt{switch.def}.

\begin{verbatim}
\if\AtBeginDocument\@undefined
\fi
\end{verbatim}
7.1 Multiple languages

\language \input plain.def \relax
\fi
\langle\langle \text{Load macros for plain if not \LaTeX}\rangle\rangle

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.

\langle\langle \text{Define core switching macros}\rangle\rangle ≡
\begin{verbatim}
\ifx\language\@undefined
\csname newcount\endcsname\language
\fi
\langle\langle \text{Define core switching macros}\rangle\rangle
\end{verbatim}

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

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

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

\langle\langle \text{Define core switching macros}\rangle\rangle ≡
\begin{verbatim}
\ifx\newlanguage\@undefined
\csname newcount\endcsname\last@language
\def\addlanguage#1{%
\global\advance\last@language\@ne
\ifnum\last@language<\@cclvi
\else
  \errmessage{No room for a new \string\language!}%
  \fi
\global\chardef#1\last@language
\wlog{\string#1 = \string\language\the\last@language}}
\else
  \global\chardef\last@language=19
  \def\addlanguage{%
  \global\chardef\@cclvi
  \wlog{\string#1 = \string\language\the\last@language}}
\else
\fi
\langle\langle \text{Define core switching macros}\rangle\rangle
\end{verbatim}

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

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

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

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

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

8.1 base

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

If the format created a list of loaded languages (in \texttt{\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.
8.2 key=value options and other general option

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

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.
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.

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

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

```latex
\def\bbl@tempa#1=#2\bbl@tempa{% 
  \bbl@csarg\ifx{opt@#1}@nil
  \bbl@csarg\edef{opt@#1}{#2}%
  \else
  \bbl@error{Bad option `#1=#2'. Either you have misspelled the\%
    key or there is a previous setting of `#1'}{Valid keys are `shorthands', `config', `strings', `main', `headfoot', `safe', `math', among others.}
  \fi}
```

Now the option list is processed, taking into account only currently declared options (including those declared with a =), and `<key>=<value>` options (the former take precedence). Unrecognized options are saved in \bbl@language@opts, because they are language options.

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

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

8.3 Conditional loading of shorthands

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

```latex
\bbl@trace{Conditional loading of shorthands}
\def\bbl@sh@string#1{\ifx#1\@empty\else
  \ifx#1t\string~%
  \else\ifx#1c\string,\%
  \else\string#1\fi\fi
  \expandafter\bbl@sh@string
  \fi}
\ifx\bbl@opt@shorthands@nil
  \def\bbl@ifshorthand#1#2#3(#2)%
```
The following macro tests if a shorthand is one of the allowed ones.

```
\def\bbl@ifshorthand#1#2#3{%
  \ifx#1\empty
    \def\bbl@ifshorthand#1{%
      \bbl@xin@{\string#1}{\bbl@opt@shorthands}%
      \ifin@
      \expandafter\@firstoftwo
    \else
      \expandafter\@secondoftwo
    \fi}
  \fi
}
```

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{%
  \expandafter\bbl@sh@string\bbl@opt@shorthands\@empty}%
```

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

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

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{%
    \set@typeset@protect
    \select@language@x\expandafter{\bbl@opt@headfoot}%
    \let\protect\noexpand
  }%
\fi
```

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

```
\ifx\bbl@opt@safe\@undefined
  \edef\bbl@opt@safe{BR}
\fi
```

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

```
\bbl@trace{Defining IfBabelLayout}
```

```
\newcommand\IfBabelLayout[3]{#3}%
```

```
\ifx\bbl@opt@layout\@nnil
  \newcommand\IfBabelLayout[3]{#3}%
  \else
  \newcommand\IfBabelLayout[1]{%}
  \expandafter\@firstoftwo
  \else
  \expandafter\@secondoftwo
  \fi
\fi
```
8.4 Language options

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

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

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

\bbl@try@load@lang#1#2#3{%
  \IfFileExists{\CurrentOption.ldf}{%
    \bbl@load@language{\CurrentOption}{#1}{#2}{#3}}{%
    \DeclareOption{afrikaans}{\bbl@try@load@lang{}{dutch}{}}}
  \DeclareOption{brazil}{\bbl@try@load@lang{}{portuges}{}}
  \DeclareOption{brazilian}{\bbl@try@load@lang{}{portuges}{}}
  \DeclareOption{hebrew}{%
    \input{rlbabel.def}%
    \bbl@load@language{hebrew}}
  \DeclareOption{hungarian}{\bbl@try@load@lang{}{magyar}{}%}
  \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}{}%}
  \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}{}%}
  \DeclareOption{polutonikogreek}{%
    \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}%
  }\DeclareOption{portuguese}{\bbl@try@load@lang{}{portuges}{}%}
  \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}{}%}
  \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}{}%}
  \DeclareOption{uppersorbian}{\bbl@try@load@lang{}{usorbian}{}%}
}

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

\ifx\bbl@opt@config\@nnil
  \@ifpackagewith{babel}{noconfigs}{}%
  \InputIfFileExists{bblopts.cfg}{%
    \typeout{*************************************^^J%
      * Local config file bblopts.cfg used^^J%
    }%
    *
    *}
  {}}%
Recognizing global options in packages not having a closed set of them is not trivial, as for them to be processed they must be defined explicitly. So, package options not yet taken into account and stored in \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 names 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 \LaTeX{} 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{}}.

69
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.

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

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

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

9.1 Tools

\iffalse
\ldf@quit
\fi
\ProvidesFile{babel.def}[^{\langle \langle \text{Make sure ProvidesFile is defined} \rangle \rangle \text{Babel common definitions}}]

The file babel.def expects some definitions made in the \LaTeX2e style file. So, in \LaTeX2e 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 are no package options, and therefore and alternative mechanism is provided. For the moment, only \texttt{\textbackslash babeloptionstrings} and \texttt{\textbackslash babeloptionmath} are provided, which can be defined before loading babel. \texttt{\textbackslash 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{#2}\
\let\bbl@language@opts\@empty
\ifx\babeloptionstrings\@undefined
\let\bbl@opt@strings\@nnil
\else
  \let\bbl@opt@strings\babeloptionstrings
\fi
\def\BabelStringsDefault{generic}
\def\bbl@tempa{normal}
\ifx\babeloptionmath\bbl@tempa
\def\bbl@mathnormal{\noexpand\textormath}
\fi
\def\AfterBabelLanguage#1#2{}
\ifx\BabelModifiers\@undefined\let\BabelModifiers\relax\fi
\let\bbl@afterlang\relax
\let\bbl@opt@safe{BR}
\ifx\@uclclist\@undefined\let\@uclclist\@empty\fi
\ifx\bbl@trace\@undefined\def\bbl@trace#1{}\fi
\expandafter\newif\csname ifbbl@single\endcsname
\fi
\end{verbatim}

And continue.

\begin{verbatim}
\ifx\bbl@switchflag\@undefined % Prevent double input
  \let\bbl@switchflag\relax
  \input switch.def\relax
\fi
\ifx\bbl@languages\@undefined
  \ifx\directlua\@undefined
    \openin1 = language.def
    \ifeof1
      \closein1
      \message{I couldn't find the file language.def}
    \else
      \closein1
      \begingroup
        \def\addlanguage#1#2#3#4#5{%
          \expandafter\ifx\csname lang@#1\endcsname\relax\else
            \global\expandafter\let\csname l@#1\expandafter\endcsname
            \csname lang@#1\endcsname
          \fi}%
        \def\uselanguage#1{}%
        \input language.def
      \endgroup
    \fi
  \fi
\end{verbatim}
For each language four control sequences have to be defined that control the language-specific definitions. To be able to add something to these macro once they have been defined the macro \addto is introduced. It takes two arguments, a \textit{control sequence} and TeX-code to be added to the \textit{control sequence}.

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

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

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 E\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@…

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

This version of \bbl@redefine can be used to redefine \texttt{\long} commands such as \texttt{\ifthenelse}.
\bbl@redefinerobust  For commands that are redefined, but which might be robust we need a slightly more intelligent macro. A robust command \texttt{foo} is defined to expand to \texttt{\protect\foo}. So it is necessary to check whether \texttt{\foo} exists. The result is that the command that is being redefined is always robust afterwards. Therefore all we need to do now is define \texttt{\foo}.

\begin{verbatim}
\def\bbl@redefinerobust#1{\
  \edef\bbl@tempa{\bbl@stripslash#1}\
  \bbl@ifunset{\bbl@tempa}{\expandafter\let\csname org@\bbl@tempa\endcsname#1 \bbl@exp{\def\#1{\protect<\bbl@tempa}}}\
  \@namedef{\bbl@tempa}
\}
\end{verbatim}

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

9.2 Hooks

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

\begin{verbatim}
\def\bbl@evargs{,%
  everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,\%
  adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,\%
  beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,\%
  hyphenation=2,initiateactive=3,afterreset=0,foreign=0,foreign*=0,\%
  beforestart=0,languagename=0}
\end{verbatim}

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).

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

\bblensure The user command just parses the optional argument and creates a new macro named \texttt{\bbl@e@⟨language⟩}. We register a hook at the afterextras event which just executes this macro in a “complete” selection (which, if undefined, is \texttt{\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 \texttt{\textbackslash babelensure{\{\texttt{include}\}}{\{\texttt{exclude}\}}{\{\texttt{fontenc}\}}}, which in turn loops over the macros names in \texttt{\textbackslash captionslist}, excluding (with the help of \texttt{\textbackslash in@}) those in the exclude list. If the fontenc is given (and not \texttt{\relax}), the \texttt{\textbackslash fontencoding} is also added. Then we loop over the include list, but if the macro already contains \texttt{\textbackslash foreignlanguage}, nothing is done. Note this macro (1) is not restricted to the preamble, and (2) changes are local.

\begin{verbatim}
\__bbl@trace{Defining babelensure}
\newcommand\babelensure[2][3]{{% TODO - revise test files
  \ifcase\bbl@cl\% #1
    \begingroup
      \let\bbl@ens@include\@empty
      \let\bbl@ens@exclude\@empty
      \bbl@tempb\% elt for (excluding) \bbl@captionslist list
      \ifx\#1\@undefined % 3.32 - Don't assume the macros exists
        \edef\#1{\noexpand\bbl@nocaption{\bbl@stripslash\#1}{\languagename\bbl@stripslash\#1}}% 
      \else
        \ifx\#1\@empty
          \in@{\#1}{#2}%
          \ifin@%
            \bbl@ifunset{bbl@ensure@\languagename}\
            \bbl@foreach\bbl@tempa{% #1}
            \def\bbl@tempc{\bbl@ensure}%
            \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{\bbl@ens@include}%
            \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{\bbl@ens@exclude}%
            \toks@\expandafter{\bbl@tempc}\
            \bbl@exp%
          \else
            \edef\#1{\bbl@csarg\noexpand{\bbl@ens@include}{\the\toks@}}%
          \fi
        \fi
      \fi
    \endgroup
    \def\<bbl@e@#2>{\the\toks@{\bbl@ens@fontenc}}}}}%
\def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
  \def\bbl@tempb\% elt for (excluding) \bbl@captionslist list
  \ifx\#1\@undefined % 3.32 - Don't assume the macros exists
    \edef\#1{\noexpand\bbl@nocaption{\bbl@stripslash\#1}{\languagename\bbl@stripslash\#1}}%
  \else
    \ifx\#1\@empty
      \in@{\#1}{#2}%
      \ifin@%
        \bbl@ifunset{bbl@ensure@\languagename}\
        \bbl@foreach\bbl@tempa{% #1}
        \def\bbl@tempc{\bbl@ensure}%
        \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{\bbl@ens@include}%
        \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{\bbl@ens@exclude}%
        \toks@\expandafter{\bbl@tempc}%
        \bbl@exp%
      \else
        \edef\#1{\bbl@csarg\noexpand{\bbl@ens@include}{\the\toks@}}%
      \fi
    \fi
  \fi
\end{verbatim}
9.3 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{\ldf@quit} to set the main language, restore the category code of the @-sign and call \texttt{\endinput}.

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

Finally we check \texttt{\originalTeX}.

\texttt{\bbl@trace(Macros for setting language files up)}
\texttt{\def\bbl@ldfinit{}}
\texttt{\let\bbl@screset\@empty}
\texttt{\let\BabelStrings\bbl@opt@string}
\texttt{\let\BabelOptions\@empty}
\texttt{\let\BabelLanguages\relax}
\texttt{\ifx\originalTeX\@undefined}
\texttt{\let\originalTeX\@empty}
\texttt{\else}
\texttt{\originalTeX}
\texttt{\fi}
\ldf@quit This macro interrupts the processing of a language definition file.

\def\ldf@quit#1{% 
\expandafter\main@language\expandafter{#1}%
\catcode`@=\atcatcode \let\atcatcode\relax
\catcode`==\eqcatcode \let\eqcatcode\relax
\endinput}

\ldf@finish This macro takes one argument. It is the name of the language that was defined in the language definition file.
We load the local configuration file if one is present, we set the main language (taking into account that the argument might be a control sequence that needs to be expanded) and reset the category code of the @-sign.

\def\bbl@afterldf#1{% 
\bbl@afterlang
\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
\endinput

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

\main@language \bbl@main@language

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

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

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

\def\bbl@beforestart{% 
\bbl@usehooks{beforestart}{}
\global\let\bbl@beforestart\relax
\AtBeginDocument{%
\bbl@cs{beforestart}%
\ife@fi
A bit of optimization. Select in heads/feet the language only if necessary.

9.4 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 \TeX is used). It is used only at one place, namely when \initiate@active@char is called (which is ignored if the char has been made active before). Because \@sanitize can be undefined, we put the definition inside a conditional. Items are added to the lists without checking its existence or the original catcode. It does not hurt, but should be fixed. It’s already done with nfss@catcodes, added in 3.10.

The companion of the former macro is \bbl@remove@special. It removes a character from the set macros \dospecials and \@sanitize, but it is not used at all in the babel core.
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{\textlangle char \textrangle} to expand to the character in its ‘normal state’ and it defines the active character to expand to \normal@char{\textlangle char \textrangle} by default (\textlangle char \textrangle being the character to be made active). Later its definition can be changed to expand to \active@char{\textlangle char \textrangle} by calling \bbl@activate{\textlangle char \textrangle}.

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 {"} (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, \langle level\rangle@group, \langle level\rangle@active and \langle next-level\rangle@active (except in system).

\def\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@afterfi\csname#2@sh@#1\endcsname  
\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.

\def\initiate@active@char#1{  
\bbl@ifunset{active@char\string#1}{}  
{\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{%  
\expandafter\ifx\csname#2@sh@#1\endcsname\relax  
\bbl@csarg{\def{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}}%  
{\expandafter\@initiate@active@char\expandafter}{\string#1\string#1#1}%  
{}}
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 math code is set to “8000 a posteriori”).

\begin{verbatim}
\let\noexpand#1\%
\expandafter\noexpand\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 math code is set to “8000 a posteriori”).

\begin{verbatim}
\let\noexpand#1\%
\expandafter\noexpand\csname bbl@oridef@@#2\endcsname}%
\fi
\end{verbatim}

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).

\begin{verbatim}
\let\bbl@tempa\@firstoftwo
\if\string^#2%
\def\bbl@tempa{\noexpand\textormath}
\else
\ifx\bbl@mathnormal@undefined\@undefined
\let\bbl@tempa\bbl@mathnormal
\fi
\expandafter\edef\csname active@char#2\endcsname{\bbl@tempa{\noexpand\if\@safe@actives
\noexpand\expandafter\expandafter\expandafter\csname normal@char#2\endcsname
\noexpand\else
\noexpand\expandafter\expandafter\expandafter\csname normal@char#2\endcsname
\noexpand\fi}}
\else
\expandafter\edef\csname normal@char#2\endcsname{\#2}%
\fi
\fi
\end{verbatim}

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).

\begin{verbatim}
\let\bbl@tempa\@firstoftwo
\if\string^#2%
\def\bbl@tempa{\noexpand\textormath}
\else
\ifx\bbl@mathnormal@undefined\@undefined
\let\bbl@tempa\bbl@mathnormal
\fi
\expandafter\edef\csname active@char#2\endcsname{\bbl@tempa{\noexpand\if\@safe@actives
\noexpand\expandafter\expandafter\expandafter\csname normal@char#2\endcsname
\noexpand\else
\noexpand\expandafter\expandafter\expandafter\csname normal@char#2\endcsname
\noexpand\fi}}
\else
\expandafter\edef\csname normal@char#2\endcsname{\#2}%
\fi
\fi
\end{verbatim}

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}).

\begin{verbatim}
\let\bbl@tempa@firstoftwo\@ifundefined
\if\string^#2%
\def\bbl@tempa{\noexpand\textormath}
\else
\ifx\bbl@mathnormal@undefined\@undefined
\let\bbl@tempa\bbl@mathnormal
\let\bbl@tempa\@ifundefined
\fi
\expandafter\edef\csname active@char#2\endcsname{\bbl@tempa{\noexpand\if\@safe@actives
\noexpand\expandafter\expandafter\expandafter\csname normal@char#2\endcsname
\noexpand\else
\noexpand\expandafter\expandafter\expandafter\csname normal@char#2\endcsname
\noexpand\fi}}
\else
\expandafter\edef\csname normal@char#2\endcsname{\#2}%
\fi
\fi
\end{verbatim}

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}).
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

\[ \text{\texttt{\textbackslash active@prefix \langle char \rangle \textnormal{\textbackslash normal@char\langle char \rangle}}}. \]

(\textit{where \texttt{\texttt{\textbackslash active@char\langle char \rangle}} is one control sequence}).

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

In order to do the right thing when a shorthand with an argument is used by itself at the end of the line we provide a definition for the case of an empty argument. For that case we let the shorthand character expand to its non-active self. Also, When a shorthand combination such as ' ' ends up in a heading \TeX{} would see \texttt{\textbackslash protect \textbackslash 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 \texttt{\textbackslash 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.

Initiating a shorthand makes active the char. That is not strictly necessary but it is still done for backward compatibility. So we need to restore the original catcode at the end of package \texttt{and} and the end of the \texttt{lfd}.
\bbl@sh@select This command helps the shorthand supporting macros to select how to proceed. Note that this macro needs to be expandable as do all the shorthand macros in order for them to work in expansion-only environments such as the argument of \hyphenation. This macro expects the name of a group of shorthands in its first argument and a shorthand character in its second argument. It will expand to either \bbl@firstcs or \bbl@scndcs. Hence two more arguments need to follow it.

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

\begingroup
\bbl@ifunset{ifincsname}\
{\gdef\active@prefix#1{% 
  \ifx\protect\@typeset@protect
  \else
    \ifx\protect\@unexpandable@protect
      \noexpand#1
    \else
      \protect#1
    \fi
    \expandafter\@gobble
  \fi
  \fi}
}{\gdef\active@prefix#1{% 
  \ifincsname
    \string#1
  \else
    \protect#1
  \fi
  \expandafter\@gobble
  \fi}}
{\gdef\active@prefix#1{% 
  \ifincsname
    \string#1
  \else
    \protect#1
  \fi
  \expandafter\@gobble
  \fi}}
\endgroup
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.

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'.

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.

Currently we only support two groups of user level shorthands, named internally \verb|user| and \verb|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 \verb|defineshorthand|) a new level is inserted for it (user@generic, done by \verb|\bbl@set@user@generic|); we make also sure {} and \verb|protect| are taken into account in this new top level.
\defineshorthand[3][\text{user}]{%
\edef\bbl@tempa{\zap@space#1 \@empty}%
\bbl@for\bbl@tempb\bbl@tempa{
\if*\expandafter\@car\bbl@tempb\@nil
  \edef\bbl@tempb{\text{user}@\expandafter\@gobble\bbl@tempb}%
\expandafter\@expandtwoargs
  \bbl@set@user@generic{\expandafter\string\@car#2\@nil}\bbl@tempb
\fi
\declare@shorthand{\bbl@tempb}{#2}{#3}}}

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

\aliasshorthand  First the new shorthand needs to be initialized,\def\aliasshorthand#1#2{%\bbl@ifshorthand{#2}%  {\expandafter\ifx\csname active@char\string#2\endcsname\relax
    \if\document\@notprerr
      \@notshorthand{#2}%
    \else
      \initiate@active@char{#2}%
    \fi
  \else\initiate@active@char{#2}%
  \fi}

Then, we define the new shorthand in terms of the original one, but note with \aliasshorthand{"}{/} is \texttt{\active@prefix} \texttt{/\active@char/}, so we still need to let the latest to \texttt{\active@char"}.
\expandafter\let\csname active@char\string#2\endcsname\csname active@char\string#1\endcsname
\expandafter\let\csname normal@char\string#2\endcsname\csname normal@char\string#1\endcsname
\bbl@activate{#2}%
\fi
\fi}%
\bbl@iferror{%
  {\bbl@error{\texttt{The character `\string #1'} should be made a shorthand character;\texttt{\%}}
    \add the command \texttt{\string\useshorthands\string\#1\string} to
    the preamble.\texttt{\%}}
  {I will ignore your instruction}\%}
  {You may proceed, but expect unexpected results}}\notshorthand  The character `\string #1' should be made a shorthand character;\texttt{\%} add the command \texttt{\string\useshorthands\string\#1\string} to
the preamble.\texttt{\%} I will ignore your instruction}\%
  {You may proceed, but expect unexpected results}}\shorthandoff  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.
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.

```
1039 \def\bbl@switch@sh#1#2{%  
1040 \ifx#2@@nil\else  
1041 \bbl@ifunset{\bbl@active@\string#2}{%  
1042 \bbl@error{I cannot switch \string#2 on or off--not a shorthand}%  
1043 \bbl@ifshorthand{#2}{%  
1044 \csname bbl@oricat@\string#2@endcsname  
1045 \csname bbl@oridef@\string#2@endcsname  
1046 \fi}%  
1047 \catcode#212\relax  
1048 \or  
1049 \catcode#2\active  
1050 \or  
1051 \csname bbl@active@\string#1@endcsname  
1052 \csname bbl@active@\string#1@endcsname  
1053 \fi}%  
1054 \bbl@afterfi\bbl@switch@sh#1%  
1055 \fi}
```

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

```
1056 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}  
1057 \def\bbl@putsh#1{%  
1058 \bbl@ifunset{\bbl@active@\string#1}{%  
1059 \bbl@putsh@i#1@@nil\@nil\@nil{%  
1060 \csname bbl@active@\string#1@endcsname}  
1061 \def\bbl@putsh#1@@nil\@nil{%  
1062 \csname languagename@sh@\string#1@endcsname  
1063 \ifx\@empty#1\else\string#1@\fi\endcsname}  
1064 \ifx\bbl@opt@shorthands@@nil\else  
1065 \let\bbl@s@initiate@active@char\initiate@active@char  
1066 \def\initiate@active@char#1{%  
1067 \bbl@ifshorthand{#1}{%  
1068 \let\bbl@active@\string#1\bbl@switch@sh  
1069 \def\bbl@switch@sh#1@#2{%  
1070 \ifx#2@@nil\else  
1071 \bbl@afterfi\bbl@switch@sh#1@#2%  
1072 \bbl@ifshorthand{#2}{%  
1073 \bbl@switch@sh#1@#2}%  
1074 \let\bbl@active@\bbl@activate  
1075 \def\bbl@activate#1{%  
1076 \bbl@ifshorthand{#1}{%  
1077 \let\bbl@active@\bbl@deactivate  
1078 \def\bbl@deactivate#1{%  
1079 \bbl@ifshorthand{#1}{%  
1080 \fi
```

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

```
1081 \newcommand\ifbabelshorthand[3]{\bbl@ifunset{\bbl@active@\string#1}{#3}{#2}}
```
One of the internal macros that are involved in substituting \prime for each right quote in mathmode is \pr@ms. 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.

\begin{verbatim}
1082 \def\bbl@prim@s{% 
1083 \prime\futurelet\@let@token\bbl@pr@m@s} 
1084 \def\bbl@if@primes#1#2{% 
1085 \ifx#1\@let@token 
1086 \expandafter\@firstoftwo 
1087 \else\ifx#2\@let@token 
1088 \bbl@afterelse\expandafter\@firstoftwo 
1089 \else 
1090 \bbl@afterfi\expandafter\@secondoftwo 
1091 \fi\fi} 
1092 \begin{group}
1093 \catcode`\^=7 \catcode`*=\active \lccode`*=`\^ 
1094 \catcode`\^=12 \catcode`*=\active \lccode`*=`\'
1095 \lowercase{% 
1096 \gdef\bbl@pr@m@s{% 
1097 \bbl@if@primes*`\pr@@@s 
1098 \bbl@if@primes*`\pr@@@t\egroup}}
1099 \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 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{verbatim}
1101 \initiate@active@char{-} 
1102 \declare@shorthand{system}{-}{\leavevmode\nobreak\ } 
1103 \bbl@activate{-} 
\end{verbatim}

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

\begin{verbatim}
1104 \expandafter\def\csname OT1dqpos\endcsname{127} 
1105 \expandafter\def\csname T1dqpos\endcsname{4}
\end{verbatim}

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

\begin{verbatim}
1106 \ifx\f@encoding@undefined 
1107 \def\f@encoding{OT1} 
1108 \fi
\end{verbatim}

\section{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.

\begin{verbatim}
1109 \bbl@trace{Language attributes} 
1110 \newcommand\languageattribute[2]{% 
\end{verbatim}
We want to make sure that each attribute is selected only once; therefore we store the already selected attributes in \bbl@known@attribs. When that control sequence is not yet defined this attribute is certainly not selected before.

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

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

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}.

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

```latex
\begin{verbatim}
\ifx\bbl@known@attribs\@undefined
\in@false
\else
\fi

\textbf{The we need to check the list of known attributes.}

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

\textbf{We first assume the attribute is unknown.}

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

\textbf{Finally we execute} \texttt{\bbl@tempa}.

\textbf{This macro removes all the attribute code from \LaTeX{}'s memory at \texttt{\begin{document}} time (if any is present).}
\end{verbatim}
```

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

\texttt{\bbl@clear@ttrib} This macro removes all the attribute code from \LaTeX{}’s memory at \texttt{\begin{document}} time (if any is present).
9.6 Support for saving macro definitions

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

\begin{verbatim}
\def\babel@beginsave{\def\babel@savecnt\z@}
\def\babel@save#1{\let\csname babel@\number\babel@savecnt\endcsname#1=\relax
\toks0={\originalTeX \let#1=}\def\originalTeX#1\relax{\the\toks0\<\babel@number\babel@savecnt\><\relax}}
\advance\babel@savecnt\@ne}
\def\babel@savevariable#1{\toks0={\originalTeX #1=}\def\originalTeX#1\relax{\the\toks0\the#1\relax}}
\def\babel@save{\let\csname babel@\number\babel@savecnt\endcsname\relax
\toks0={\originalTeX \let#1=}\def\originalTeX#1\relax{\the\toks0\<\babel@number\babel@savecnt\><\relax}}
\def\ babel@savevariable#1{\toks0={\originalTeX #1=}\def\originalTeX#1\relax{\the\toks0\the#1\relax}}
\def\babel@beginsave{\def\babel@savecnt\z@}
\def\babel@save#1{\let\csname babel@\number\babel@savecnt\endcsname#1=\relax
\toks0={\originalTeX \let#1=}\def\originalTeX#1\relax{\the\toks0\<\babel@number\babel@savecnt\><\relax}}
\advance\babel@savecnt\@ne}
\def\babel@savevariable#1{\toks0={\originalTeX #1=}\def\originalTeX#1\relax{\the\toks0\the#1\relax}}
\def\babel@save{\let\csname babel@\number\babel@savecnt\endcsname\relax
\toks0={\originalTeX \let#1=}\def\originalTeX#1\relax{\the\toks0\<\babel@number\babel@savecnt\><\relax}}
\def\ babel@savevariable#1{\toks0={\originalTeX #1=}\def\originalTeX#1\relax{\the\toks0\the#1\relax}}
\def\babel@beginsave{\def\babel@savecnt\z@}
\def\babel@save#1{\let\csname babel@\number\babel@savecnt\endcsname#1=\relax
\toks0={\originalTeX \let#1=}\def\originalTeX#1\relax{\the\toks0\<\babel@number\babel@savecnt\><\relax}}
\advance\babel@savecnt\@ne}
\def\babel@savevariable#1{\toks0={\originalTeX #1=}\def\originalTeX#1\relax{\the\toks0\the#1\relax}}
\end{verbatim}

9.7 Short tags

These macros are straightforward. After zapping spaces, we loop over the list and define the macros \texttt{\text{⟨tag⟩}} and \texttt{\langle tag⟩}. Definitions are first expanded so that they don’t contain \texttt{\csname} but the actual macro.

\begin{verbatim}
\def\babeltags#1{\edef\bbl@tempa{\zap@space#1 \@empty}\def\bbl@tempb##1=##2\@@{\let\bbl@frenchspacing\relax
\let\bbl@nonfrenchspacing\relax\def\bbl@frenchspacing{\ifnum\the\sfcode`\relax=\@m\relax\let\bbl@frenchspacing\relax}
\let\bbl@frenchspacing\relax\let\bbl@nonfrenchspacing\relax}
\end{verbatim}
9.8 Hyphens

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

\texttt{\textbackslash babelhyphen}  
Macrosto insert common hyphens. Notethespace before @ in \texttt{\textbackslash babelhyphen}. Instead of protecting it with \texttt{\textbackslash DeclareRobustCommand}, which could insert a \texttt{\textbackslash relax}, we use the same procedure as shorthands, with \texttt{\active@prefix}.
The following two commands are used to wrap the “hyphen” and set the behavior of the rest of the word – the version with a single @ is used when further hyphenation is allowed, while that with @@ if no more hyphens are allowed. In both cases, if the hyphen is preceded by a positive space, breaking after the hyphen is disallowed. There should not be a discretionary after a hyphen at the beginning of a word, so it is prevented if preceded by a skip. Unfortunately, this does handle cases like “(-suffix)”. \nobreak is always preceded by \leavevmode, in case the shorthand starts a paragraph.

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.

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

9.9 Multiencoding strings
The aim following commands is to provide a common interface for strings in several encodings. They also contains several hooks which can be used by luatex and xetex. The code is organized here with pseudo-guards, so we start with the basic commands.
Tools  But first, a couple of tools. The first one makes global a local variable. This is not the best solution, but it works.

1276 \bbl@trace{Multiencoding strings}
1277 \def\bbl@toglobal#1{\global\let#1#1}
1278 \def\bbl@recatcode#1{\
1279 \@tempcnta="7F
1280 \def\bbl@tempa{\
1281 \ifnum\@tempcnta>"FF\else
1282 \catcode\@tempcnta="#1\relax
1283 \advance\@tempcnta\@ne
1284 \expandafter\bbl@tempa
1285 \fi}\
1286 \bbl@tempa}
1287 The second one. We need to patch \@ucclist, but it is done once and only if \SetCase is used or if strings are encoded. The code is far from satisfactory for several reasons, including the fact \@ucclist 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@ucld. The parser is restarted inside \langle{lang}\rangle\bbl@ucld because we do not know how many expansions are necessary (depends on whether strings are encoded). The last part is tricky – when upercasing, we have:

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

and starts over (and similarly when lowercasing).

1288 \ifpackagewith{babel}{nocase}\
1289 \let\bbl@patchucld@relax\
1290 \def\bbl@patchucld{\
1291 \global\let\bbl@patchucld@relax
1292 \g@addto@macro\@ucclist{\reserved@b{\reserved@b\bbl@ucld}}\%
1293 \gdef\bbl@ucld##1{\
1294 \let\bbl@encoded\bbl@encoded@ucld
1295 \bbl@ifunset{\languagename \bbl@ucld}{}{\%}
1296 \{\let\bbl@tempa##1\relax % Used by LANG@bbl@ucld
1297 \csname\languagename \bbl@ucld\endcsname\%}
1298 \{\let\bbl@toppler@empty\{\bbl@toppler@empty\}\%
1299 \gdef\bbl@toppler{\csname\languagename \bbl@uc@endcsname\%}
1300 \gdef\bbl@toppler{\csname\languagename \bbl@uc@endcsname\%}}
1301 \let\bbl@opt@strings\@nnil % accept strings=value
1302 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1303 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1304 \def\BabelStringsDefault{generic}
1305 \let\bbl@opt@strings\@nnil % accept strings=value
1306 \let\bbl@opt@strings\@nnil % accept strings=value
1307 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1308 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1309 \let\bbl@opt@strings\@nnil % accept strings=value
1310 \let\bbl@opt@strings\@nnil % accept strings=value

The following package options control the behavior of \SetString.

1311 \let\bbl@opt@strings\@nnil % accept strings=value
1312 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1313 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1314 \let\bbl@opt@strings\@nnil % accept strings=value
1315 \let\bbl@opt@strings\@nnil % accept strings=value
1316 \let\bbl@opt@strings\@nnil % accept strings=value
1317 \let\bbl@opt@strings\@nnil % accept strings=value
1318 \let\bbl@opt@strings\@nnil % accept strings=value
1319 \let\bbl@opt@strings\@nnil % accept strings=value
1320 \let\bbl@opt@strings\@nnil % accept strings=value
1321 \let\bbl@opt@strings\@nnil % accept strings=value
1322 \let\bbl@opt@strings\@nnil % accept strings=value
1323 \let\bbl@opt@strings\@nnil % accept strings=value
1324 \let\bbl@opt@strings\@nnil % accept strings=value
1325 \let\bbl@opt@strings\@nnil % accept strings=value
1326 \let\bbl@opt@strings\@nnil % accept strings=value
1327 \let\bbl@opt@strings\@nnil % accept strings=value
1328 \let\bbl@opt@strings\@nnil % accept strings=value
1329 \let\bbl@opt@strings\@nnil % accept strings=value
1330 \let\bbl@opt@strings\@nnil % accept strings=value
1331 \let\bbl@opt@strings\@nnil % accept strings=value
1332 \let\bbl@opt@strings\@nnil % accept strings=value
1333 \let\bbl@opt@strings\@nnil % accept strings=value
1334 \let\bbl@opt@strings\@nnil % accept strings=value
1335 \let\bbl@opt@strings\@nnil % accept strings=value
1336 \let\bbl@opt@strings\@nnil % accept strings=value
1337 \let\bbl@opt@strings\@nnil % accept strings=value
1338 \let\bbl@opt@strings\@nnil % accept strings=value
1339 \let\bbl@opt@strings\@nnil % accept strings=value
1340 \let\bbl@opt@strings\@nnil % accept strings=value
1341 \let\bbl@opt@strings\@nnil % accept strings=value
1342 \let\bbl@opt@strings\@nnil % accept strings=value
1343 \let\bbl@opt@strings\@nnil % accept strings=value
1344 \let\bbl@opt@strings\@nnil % accept strings=value
1345 \let\bbl@opt@strings\@nnil % accept strings=value
1346 \let\bbl@opt@strings\@nnil % accept strings=value
1347 \let\bbl@opt@strings\@nnil % accept strings=value
1348 \let\bbl@opt@strings\@nnil % accept strings=value
1349 \let\bbl@opt@strings\@nnil % accept strings=value
1350 \let\bbl@opt@strings\@nnil % accept strings=value
1351 \let\bbl@opt@strings\@nnil % accept strings=value
1352 \let\bbl@opt@strings\@nnil % accept strings=value
1353 \let\bbl@opt@strings\@nnil % accept strings=value
1354 \let\bbl@opt@strings\@nnil % accept strings=value
1355 \let\bbl@opt@strings\@nnil % accept strings=value
1356 \let\bbl@opt@strings\@nnil % accept strings=value
1357 \let\bbl@opt@strings\@nnil % accept strings=value
1358 \let\bbl@opt@strings\@nnil % accept strings=value
1359 \let\bbl@opt@strings\@nnil % accept strings=value

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.

1360 \onlypreamble\StartBabelCommands
\def\StartBabelCommands{%
\begingroup
\bbl@recatcode{11}%
\providecommand\bbl@provstring##1##2{%\providecommand##1{##2}%\bbl@toglobal##1}%
\global\let\bbl@scafter\@empty
\let\StartBabelCommands\bbl@startcmds
\ifx\BabelLanguages\relax\let\BabelLanguages\CurrentOption\fi
\begingroup
\let\bbl@screset\@nnil % local flag - disable 1st stopcommands
\StartBabelCommands}
\def\bbl@startcmds{%\ifx\bbl@screset\@nnil\else\bbl@usehooks{stopcommands}{}%\fi\endgroup
\begingroup
\@ifstar
{\ifx\bbl@opt@strings\@nnil
\let\bbl@opt@strings\BabelStringsDefault
\fi\bbl@startcmds@i}\
\bbl@startcmds@i}{%\if\bbl@sc@label\generic\def\bbl@encstring##1##2{%\ProvideTextCommandDefault##1{##2}%\bbl@toglobal##1}%\let\bbl@sctest\in@true\else\let\bbl@sc@charset\space % <- zapped below\let\bbl@sc@fontenc\space % <- " "\def\bbl@tempa##1=##2\@nil{%\ProvideTextCommandDefault##1{##2}%\bbl@toglobal##1}%\expandafter\bbl@toglobal\csname\string\##1\endcsname}\let\bbl@sctest\in@true\else
\let\bbl@screset\@nnil % local flag - disable 1st stopcommands
\StartBabelCommands}
\def\bbl@startcmds{%\ifx\bbl@screset\@nnil\else\bbl@usehooks{stopcommands}{}%\fi\endgroup
\begingroup
\@ifstar
{\ifx\bbl@opt@strings\@nnil
\let\bbl@opt@strings\BabelStringsDefault
\fi\bbl@startcmds@i}\
\bbl@startcmds@i}{%\if\bbl@sc@label\generic\def\bbl@encstring##1##2{%\ProvideTextCommandDefault##1{##2}%\bbl@toglobal##1}%\let\bbl@sctest\in@true\else\let\bbl@sc@charset\space % <- zapped below\let\bbl@sc@fontenc\space % <- " "\def\bbl@tempa##1=##2\@nil{%\ProvideTextCommandDefault##1{##2}%\bbl@toglobal##1}%\expandafter\bbl@toglobal\csname\string\##1\endcsname}\let\bbl@sctest\in@true\else
Parsetheencodinginfotogetthelabel, input, and font parts.
Selectthebehaviorof\SetString. Threear two main cases, dependingof 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.\newcommand\bbl@startcmds@i\bbl@startcmds@i\@empty{\empty}{%\let\SetString\@gobbletwo
\let\bbl@stringdef\@gobbletwo\let\AfterBabelCommands\@gobble
\if\bbl@sc@label\generic\def\bbl@encstring##1##2{%\ProvideTextCommandDefault##1{##2}%\bbl@toglobal##1}%\let\bbl@sctest\in@true\else
\let\bbl@screset\@nnil % local flag - disable 1st stopcommands
\StartBabelCommands
does nothing.
\let\bbl@sc@charset\space % <- zapped below
\let\bbl@sc@fontenc\space % <- " "
\def\bbl@tempa##1=##2\@nil{%\ProvideTextCommandDefault##1{##2}%\bbl@toglobal##1}%\expandafter\bbl@toglobal\csname\string\##1\endcsname}\let\bbl@sctest\in@true\else
\let\bbl@screset\@nnil % local flag - disable 1st stopcommands
\StartBabelCommands
There are two versions of \texttt{\bbl@scswitch}. The first version is used when \texttt{\ldf}s are read, and it makes sure that the group/\texttt{\language} is reset, but only once (\texttt{\bbl@screset} is used to keep track of this). The second version is used in the preamble and packages loaded after \texttt{babel} and does nothing. The macro \texttt{\bbl@forlang} loops but its body is executed only if the value is in \texttt{\BabelLanguages} (inside \texttt{babel}) or \texttt{\date/\language} is defined (after \texttt{babel} has been loaded). There are also two versions of \texttt{\bbl@forlang}. The first one skips the current iteration if the language is not in \texttt{\BabelLanguages} (used in \texttt{\ldf}s), and the second one skips undefined languages (after \texttt{babel} has been loaded).
Now we define commands to be used inside \StartBabelCommands.

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

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

\newcommand\SetCase[3][]{\bbl@forlang\bbl@tempa{\expandafter\bbl@encstring\csname\bbl@tempa @bbl@uc\endcsname{##2} \expandafter\bbl@encstring\csname\bbl@tempa @bbl@lc\endcsname{##3}}}\
\newcommand\SetHyphenMap[1][]{\bbl@forlang\bbl@tempa{\expandafter\bbl@stringdef\csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}\

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.
\lc{\textcode{#1}=#2}\relax
\fi}\newcommand\BabelLowerMM[4]{% many-to-many
  \@tempcnta=#1\relax
  \@tempcntb=#4\relax
  \def\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{%
    \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.
\[\text{\texttt{\texttt{⟨⟨∗}} \text{\texttt{More package options}} \text{\texttt{⟩⟩ ≡}}]}\]
\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}
\[\text{\texttt{⟨⟨}} / \text{\texttt{More package options}} \text{\texttt{⟩⟩}}\]

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}

9.10 Macros common to a number of languages
\set@low@box The following macro is used to lower quotes to the same level as the comma. It prepares its argument in box register 0.
\bbl@trace{Macros related to glyphs}
\set@low@box{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}\
  \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@\%\setbox\z@\hbox{\lower\dimen\z@ \box\z@}\ht\z@\ht\tw@ \dp\z@\dp\tw@}
\save@sf@q The macro \save@sf@q is used to save and reset the current space factor.
\save@sf@q{\set@low@box{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}\
  \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@\%\setbox\z@\hbox{\lower\dimen\z@ \box\z@}\ht\z@\ht\tw@ \dp\z@\dp\tw@}
\save@sf@q{\set@low@box{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}\
  \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@\%\setbox\z@\hbox{\lower\dimen\z@ \box\z@}\ht\z@\ht\tw@ \dp\z@\dp\tw@}

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

\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}
1534 \ProvideTextCommand{\quotedblbase}{OT1}{%
1535  \save@sf@q{\set@low@box{\textquotedblright}/}%
1536  \box\z@\kern-.04em\bbl@allowhyphens}}
\end{verbatim}

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

\begin{verbatim}
1537 \ProvideTextCommandDefault{\quotedblbase}{%
1538  \UseTextSymbol{OT1}{\quotedblbase}}
\end{verbatim}

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

\begin{verbatim}
1539 \ProvideTextCommand{\quotesinglbase}{OT1}{%
1540  \save@sf@q{\set@low@box{\textquoteright}/}%
1541  \box\z@\kern-.04em\bbl@allowhyphens}}
\end{verbatim}

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

\begin{verbatim}
1542 \ProvideTextCommandDefault{\quotesinglbase}{%
1543  \UseTextSymbol{OT1}{\quotesinglbase}}
\end{verbatim}

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

\begin{verbatim}
1544 \ProvideTextCommand{\guillemotleft}{OT1}{%
1545  \ifmmode
1546    \ll
1547  \else
1548    \save@sf@q{\nobreak
1549    \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bbl@allowhyphens}%
1550  \fi}
1551 \ProvideTextCommand{\guillemotright}{OT1}{%
1552  \ifmmode
1553    \gg
1554  \else
1555    \save@sf@q{\nobreak
1556    \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bbl@allowhyphens}%
1557  \fi}
\end{verbatim}

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

\begin{verbatim}
1558 \ProvideTextCommandDefault{\guillemotleft}{%
1559  \UseTextSymbol{OT1}{\guillemotleft}}
1560 \ProvideTextCommandDefault{\guillemotright}{%
1561  \UseTextSymbol{OT1}{\guillemotright}}
\end{verbatim}

\texttt{\textbackslash guilsingleleft} The single guillemets are not available in OT1 encoding. They are faked.

\begin{verbatim}
1562 \ProvideTextCommand{\guilsingleleft}{OT1}{%
1563  \ifmmode
1564    <
1565  \else
1566    \save@sf@q{\nobreak
1567    \raise.2ex\hbox{$\scriptscriptstyle<$}\bbl@allowhyphens}%
1568  \fi}
1569 \ProvideTextCommand{\guilsingleright}{OT1}{%
1570  \ifmmode
\end{verbatim}
9.11.2 Letters

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

\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 Stipčević Mario, (stipcevic@olimp.irb.hr).

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

\def\crrtic@{\hrule height0.1ex width0.3em}
\def\crttic@{\hrule height0.1ex width0.33em}
\def\ddj@{\setbox0\hbox{d}\dimen@=.55\ht0
\dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@\advance\dimen@ii.15ex % correction for the dash position
\advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
\leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
\def\DDJ@{\setbox0\hbox{D}\dimen@=.55\ht0
\dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@\advance\dimen@ii.15ex % correction for the dash position
\advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
\leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}
For the \texttt{T1} encoding \texttt{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.

\texttt{SS} \footnote{\texttt{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.}

\providecommand{\SS}{\texttt{SS}}
\providecommanddefault{\SS}{\texttt{SS}}

\section{Shorthands for quotation marks}

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

\texttt{glq} The ‘german’ single quotes.
\texttt{grq} The ‘german’ double quotes.

\providecommand{\glq}{\texttt{glq}}
\providecommanddefault{\glq}{\texttt{glq}}
\providecommand{\grq}{\texttt{grq}}
\providecommanddefault{\grq}{\texttt{grq}}

\texttt{glqq} The ‘german’ double quotes.
\texttt{grqq} The ‘german’ double quotes.

\providecommand{\glqq}{\texttt{glqq}}
\providecommanddefault{\glqq}{\texttt{glqq}}
\providecommand{\grqq}{\texttt{grqq}}
\providecommanddefault{\grqq}{\texttt{grqq}}

\texttt{flq} The ‘french’ single guillemets.
\texttt{frq} The ‘french’ single guillemets.

\providecommand{\flq}{\texttt{flq}}
\providecommanddefault{\flq}{\texttt{flq}}
\providecommand{\frq}{\texttt{frq}}
\providecommanddefault{\frq}{\texttt{frq}}
The ‘french’ double guillemets.

\text{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.

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

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.

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

Then we compute the new x-height in such a way that the umlaut character is lowered to the base character. The value of \texttt{.45ex} depends on the 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{accent} primitive, reset the old x-height and insert the base character in the argument.

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

```latex
\AtBeginDocument{% 
\DeclareTextCompositeCommand{"}{OT1}{a}{\texttt{\textbackslash{umlaut}a}{a}}% 
\DeclareTextCompositeCommand{"}{OT1}{e}{\texttt{\textbackslash{umlaut}e}{e}}% 
\DeclareTextCompositeCommand{"}{OT1}{i}{\texttt{\textbackslash{umlaut}e}{i}}% 
\DeclareTextCompositeCommand{"}{OT1}{\i}{\texttt{\textbackslash{umlaut}e}\{\i\}}% 
\DeclareTextCompositeCommand{"}{OT1}{o}{\texttt{\textbackslash{umlaut}o}{o}}% 
\DeclareTextCompositeCommand{"}{OT1}{u}{\texttt{\textbackslash{umlaut}u}{u}}% 
\DeclareTextCompositeCommand{"}{OT1}{A}{\texttt{\textbackslash{umlaut}A}{A}}% 
\DeclareTextCompositeCommand{"}{OT1}{E}{\texttt{\textbackslash{umlaut}E}{E}}% 
\DeclareTextCompositeCommand{"}{OT1}{I}{\texttt{\textbackslash{umlaut}E}{I}}% 
\DeclareTextCompositeCommand{"}{OT1}{O}{\texttt{\textbackslash{umlaut}O}{O}}% 
\DeclareTextCompositeCommand{"}{OT1}{U}{\texttt{\textbackslash{umlaut}U}{U}}% 
}%
```

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

```latex
\if\l@english\@undefined
\chardef\l@english\z@
\fi
\main@language{english}
```

## 9.12 Layout

**Work in progress.**

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

```latex
\bbl@trace{Bidi layout}
\providecommand\IfBabelLayout[3]{#3}
\newcommand\BabelPatchSection[1]{% 
  \@ifundefined{#1}{}{\
    \bbl@exp{\let\langle\bbl@ss@#1\rangle<#1>}
    \namedef{#1}{\
      \@ifstar{\bbl@presec@s{#1}}{}{\@dblarg{\bbl@presec@x{#1}}}}}
  \IfBabelLayout{sectioning}{% 
    \BabelPatchSection{part}{% 
      \BabelPatchSection{chapter}{% 
        \BabelPatchSection{section}{% 
          \BabelPatchSection{subsection}{% 
            \bbl@exp{\
              \select@language@x{\bbl@main@language}{\bbl@cs{sspre@#1}{\bbl@cs{ss@#1}}{#2}}{\bbl@cs{ss@#1}}{#3}}% 
            \bbl@exp{% 
              \select@language@x{\bbl@main@language}{\bbl@cs{sspre@#1}{\bbl@cs{ss@#1}}{#2}}{\bbl@cs{ss@#1}}{#3}}% 
            \foreignlanguage{\languagename}{\unexpanded{#2}}% 
            \foreignlanguage{\languagename}{\unexpanded{#3}}}}
          \select@language@x{\languagename}{x{\languagename}}}% 
        \endgroup}}
      \endgroup}}
  \endgroup}}
\endgroup}}
\if\l@english\@undefined
\chardef\l@english\@undefined
\fi
\main@language{english}
```
9.13 Load engine specific macros

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

9.14 Creating languages

\bbl@trace{Creating languages and reading ini files}
\newcommand{\bbl@provide}[2][]{%
\let\bbl@savelangname\languagename
\edef\bbl@savelocaleid{\the\localeid}%
% Set name and locale id
\edef\languagename{#2}%
% \global\@namedef{bbl@lcname@#2}{#2}%
\bbl@id@assign
\let\bbl@KVP@captions\@nil
\let\bbl@KVP@import\@nil
\let\bbl@KVP@main\@nil
\let\bbl@KVP@script\@nil
\let\bbl@KVP@language\@nil
\let\bbl@KVP@hyphenrules\@nil % only for provide@new
\let\bbl@KVP@mapfont\@nil
\let\bbl@KVP@maparabic\@nil
\let\bbl@KVP@mapdigits\@nil
\let\bbl@KVP@intraspace\@nil
\let\bbl@KVP@intrapenalty\@nil
\let\bbl@KVP@onchar\@nil
\let\bbl@KVP@alph\@nil
\let\bbl@KVP@Alph\@nil
\let\bbl@KVP@info\@nil % Ignored with import? Or error/warning?
\bbl@forkv{#1}{% TODO - error handling
  \in@{\}/{##1}{##1}%
  \elifin@
    \bbl@renewinikey{#1}{##2}%
  \else
    \bbl@csarg\edef{KVP@@}{##2}%
  \fi}%
% == import, captions ==
\iffx\bbl@KVP@import\@nil\else
  \bbl@exp{\bbl@ifblank{\bbl@KVP@import}{}}%
\def\BabelBeforeIni##1##2{$\texttt{gdef}\ bbl@KVP@import{##1}\texttt{endinput}$}
\InputIfFileExists{babel-#2.tex}{}}{}
\endgroup{}
{}
{}
\if\bbl@KVP@captions\@nil
{\let\bbl@KVP@captions\bbl@KVP@import}
\fi
%
% Load ini
\bbl@ifunset{date#2}{}
{\bbl@provide@new{#2}}{}
{\bbl@ifblank{#1}{\bbl@error
{If you want to modify `#2' you must tell how in\% the optional argument. See the manual for the\% available options.}{}
{Use this macro as documented}}{}
{\bbl@provide@renew{#2}}}{}
%
% Post tasks
\bbl@exp{\\babelensure[exclude=\\today]{#2}}{}
\bbl@funset{\bbl@ensure@languagename}{}
{\bbl@exp{\\DeclareRobustCommand\bbl@ensure@languagename[1]{\\foreignlanguage{\languagename}{####1}}}}{}
{}
% At this point all parameters are defined if 'import'. Now we
% execute some code depending on them. But what about if nothing was
% imported? We just load the very basic parameters: ids and a few
% more.
\bbl@ifunset{\bbl@lname@#2}{}
{\def\BabelBeforeIni##1##2{\begingroup
\catcode`\[=12 \catcode`\]=12 \catcode`\;=12 %
\let\bbl@ini@captions@aux\@gobbletwo
\def\bbl@inidate ####1.####2.####3.####4\relax ####5####6{}}{}
\bbl@read@ini{##1}{basic data}{}
\bbl@exportkey{chrng}{characters.ranges}{}
\bbl@exportkey{dgnat}{numbers.digits.native}{}
\bbl@exportkey{prehc}{typography.prehyphenchar}{}
\bbl@exportkey{lnbrk}{typography.linebreaking}{h}{}
\bbl@exportkey{hyphr}{typography.hyphenrules}{}
\bbl@exportkey{hyoth}{typography.hyphenate.other}{}
\bbl@exportkey{intsp}{typography.intraspace}{}
\endinput
\endgroup}% boxed, to avoid extra spaces:
{\setbox\z@hbox{\InputIfFileExists{babel-#2.tex}{}}}}{}
{}
% -
% == script, language ==
% Override the values from ini or defines them
\if\bbl@KVP@script\@nil\else
{\bbl@csarg\edef{sname@#2}{\bbl@KVP@script}}{}
\fi
\if\bbl@KVP@language\@nil\else
{\bbl@csarg\edef{lname@#2}{\bbl@KVP@language}}{}
\fi
%
% == onchar ==
%
\directlua{
if Babel.locale_mapped == nil then
  Babel.locale_mapped = true
  Babel.linebreaking.add_before(Babel.locale_map)
  Babel.loc_to_scr = {}
  Babel.chr_to_loc = Babel.chr_to_loc or {}
end}
\directlua{
if Babel.script_blocks['\bbl@cl{sbcp}'] then
  Babel.loc_to_scr[\the\localeid] = Babel.script_blocks['\bbl@cl{sbcp}']
  Babel.locale_props[\the\localeid].lc = \the\localeid\space
  Babel.locale_props[\the\localeid].lg = \the\@nameuse{l@\languagename}\space
end
}
\directlua{
if Babel.script_blocks['\bbl@cl{sbcp}'] then
  Babel.loc_to_scr[\the\localeid] = Babel.script_blocks['\bbl@cl{sbcp}']
end}
\AtBeginDocument{%
\expandafter\bbl@add\csname selectfont \endcsname{\bbl@mapselect}{selectfont}}
\def\bbl@mapselect{%
\let\bbl@mapselect\relax
\edef\bbl@prefontid{\fontid\font}
\def\bbl@mapdir##1{%
  \def\languagename{##1}
  \let\bbl@ifrestoring\@firstoftwo % To avoid font warning
  \bbl@switchfont
  \directlua{
    Babel.locale_props[\the\csname bbl@id@@##1\endcsname\space]
    ['/'\bbl@prefontid'] = \fontid\space}}%}
\fi
\bbl@exp{\bbl@add\bbl@mapselect\bbl@mapdir{\languagename}}%
% == mapfont ==
% For bidi texts, to switch the font based on direction
\ifx\bbl@KVP@mapfont\@nil\else
  \bbl@ifsamestring{\bbl@KVP@mapfont}{direction}{}%\bbl@error{Option \bbl@KVP@mapfont unknown for %
  ['/'\bbl@prefontid'] = \fontid\space}}%
  \bbl@switchfont
  \directlua{
    Babel.locale_props[\the\csname bbl@id@@##1\endcsname\space]
    ['/'\bbl@prefontid'] = \fontid\space}}%}
\fi
% TODO - catch non-valid values
\if\bbl@mapfont ==
  \bbl@error{Option \bbl@KVP@mapfont' unknown for \%}
  \bbl@switchfont
  \directlua{
    Babel.locale_props[\the\csname bbl@id@@##1\endcsname\space]
    ['/'\bbl@prefontid'] = \fontid\space}}%}
\fi
% == mapfont ==
% For bidi texts, to switch the font based on direction
\if\bbl@KVP@mapfont\@nil\else
  \bbl@ifsamestring{\bbl@KVP@mapfont}{direction}{}%\bbl@error{Option \bbl@KVP@mapfont unknown for %
  ['/'\bbl@prefontid'] = \fontid\space}}%
  \bbl@switchfont
  \directlua{
    Babel.locale_props[\the\csname bbl@id@@##1\endcsname\space]
    ['/'\bbl@prefontid'] = \fontid\space}}%}
\fi
% TODO - catch non-valid values
\if\bbl@mapfont ==
  \bbl@error{Option \bbl@KVP@mapfont' unknown for \%}
  \bbl@switchfont
  \directlua{
    Babel.locale_props[\the\csname bbl@id@@##1\endcsname\space]
    ['/'\bbl@prefontid'] = \fontid\space}}%}
(See the manual for details.)% 
\bbl@ifunset{bbl@lsys@\languagename}\{\bbl@provide@lsys@\languagename}\%
\bbl@ifunset{bbl@wdir@\languagename}\{\bbl@provide@dirs@\languagename}\%
\ifax\bbl@mapselect@undefined
  \AtBeginDocument(%
    \expandafter\bbl@add\csname selectfont \endcsname{\bbl@mapselect}%
    \{\selectfont}%
  \def\bbl@mapselect{%
    \let\bbl@mapselect\relax
    \edef\bbl@prefontid{\fontid\font}}%
  \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\languagename}}}%
  \fi
\bbl@exp{\\bbl@add\\bbl@mapselect{\\bbl@mapdir{\languagename}}}%
  \fi
% == intraspace, intrapenalty ==
% For CJK, East Asian, Southeast Asian, if interspace in ini
\ifax\bbl@KVP@intraspace@nil\else % We can override the ini or set
  \bbl@csarg\edef{intsp@#2}{\bbl@KVP@intraspace}%
\fi
\bbl@provide@intraspace
% == hyphenate.other ==
\bbl@ifunset{bbl@hyoth@\languagename}{}%
{\bbl@csarg@replace@hyoth@\languagename}{ , }%
\bbl@provide@commands*{\languagename}%
\ifcase\bbl@engine
  \ifnum##1<257
    \SetHyphenMap{\BabelLower{##1}{##1}}%
  \fi
\else
  \SetHyphenMap{\BabelLower{##1}{##1}}%
\fi%
\bbl@endcommands%
% == maparabic ==
% Native digits, if provided in ini (TeX level, xe and lua)
\ifcase\bbl@engine\else
  \bbl@ifunset{bbl@dgnat@\languagename}{}%
  {\bbl@csarg@\bbl@setdigits@{\languagename}{}
    \ifx\bbl@KVP@maparabic@nil\else
      \ifx\bbl@latinarabic@undefined
        \let\bbl@latinarabic@\csname bbl@counter@\languagename\endcsname
      \else
        \let\bbl@latinarabic@\csname bbl@counter@\languagename\endcsname
      \fi
    \fi
  }%
\fi%
\ifodd\bbl@engine
\ifx\bbl@KVP@mapdigits@nil\else
\bbl@ifunset{bbl@dgnat@\languagename}{}
\directlua{
Babel = Babel or {}  %%% -> presets in luababel
Babel.digits_mapped = true
Babel.digits = Babel.digits or {}
Babel.digits[\the/localeid] =
table.pack(string.utfvalue(\bbl@cl{dgnat}'))
if not Babel.numbers then
function Babel.numbers(head)
  local LOCALE = luatexbase.registernumber'\bbl@attr@locale'
  local GLYPH = node.id'glyph'
  local inmath = false
  for item in node.traverse(head) do
    if not inmath and item.id == GLYPH then
      local temp = node.get_attribute(item, LOCALE)
      if Babel.digits[temp] then
        local chr = item.char
        if chr > 47 and chr < 58 then
          item.char = Babel.digits[temp][chr-47]
        end
      end
      elseif item.id == node.id'math' then
        inmath = (item.subtype == 0)
      end
      end
      return head
    end
  end
}\i
\fi
% == alph, Alph ==
% What if extras<lang> contains a \bbl@save\@alph? It won't be
% restored correctly when exiting the language, so we ignore
% this change with the \bbl@alph@saved trick.
\ifx\bbl@KVP@alph@nil\else
\toks@expandafter\expandafter\expandafter\expandafter\%
\csname extras\languagename\endcsname\%
\bbl@exp(\%
\edef<extras\languagename>{\%
\let\\bbl@alph@saved\\@alph
\the\toks@
\let\@alph\\bbl@alph@saved
\bbl@save\\@alph
\let\@alph<\bbl@cntr@\bbl@KVP@alph @\languagename>}}\%
\fi
\fi\bbl@KVP@Alph@nil\else
\toks@expandafter\expandafter\expandafter\expandafter\%
\csname extras\languagename\endcsname\%
\bbl@exp(\%
\edef<extras\languagename>{\%
\let\\bbl@Alph@saved\\@Alph
\the\toks@
\let\@Alph\\bbl@Alph@saved
\bbl@save\\@Alph
\let\@Alph<\bbl@cntr@\bbl@KVP@Alph @\languagename>}}\%
\fi

A tool to define the macros for native digits from the list provided in the ini file. Somewhat convoluted because there are 10 digits, but only 9 arguments in TeX.

Depending on whether or not the language exists, we define two macros.
The hyphenrules option is handled with an auxiliary macro.

The hyphenrules option is handled with an auxiliary macro.
The reader of ini files. There are 3 possible cases: a section name (in the form [...]), a comment (starting with ;) and a key/value pair.
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. The secpost "hook" is used only by 'identification', while secpre only by date.gregorian.licr.

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. Note \bbl@secpost@identification is called always (via \bbl@iniexec), while \bbl@after@ini must be called explicitly after \bbl@read@ini if necessary.
Now captions and captions.licr, depending on the engine. And below also for dates. They rely on a few auxiliary macros. It is expected the ini file provides the complete set in Unicode and LICR, in that order.

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

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

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

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

\ifcase\bbl@engine\bbl@csarg\def{inikv@date.gregorian}#1=#2\@@{% for defaults\bbl@inidate#1...\relax{#2}}\else\bbl@csarg\def{inikv@date.gregorian.licr}#1=#2\@@{% override\bbl@inidate#1...\relax{#2}}\fi
\bbl@csarg\def{secpre@date.gregorian.licr}{% discard uni\ifcase\bbl@engine\let\bbl@savedate\@empty\fi}

% eg: 1=months, 2=wide, 3=1, 4=dummy
\def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'\bbl@trim\bbl@inidate#1...\relax{#2}{#3}{#4}{#5}{#6}}

\ifcase\bbl@engine\bbl@csarg\def{inikv@date.gregorian.licr}#1=#2\@@{% override\bbl@inidate#1...\relax{#2}}\else\bbl@csarg\def{secpre@date.gregorian.licr}{% discard uni\ifcase\bbl@engine\let\bbl@savedate\@empty\fi}
\fi

\def\bbl@csarg\def{secpre@date.gregorian.licr}{% discard uni\ifcase\bbl@engine\let\bbl@savedate\@empty\fi}

\ifcase\bbl@engine\bbl@csarg\def{inikv@date.gregorian}#1=#2\@@{% for defaults\bbl@inidate#1...\relax{#2}}\else\bbl@csarg\def{inikv@date.gregorian.licr}#1=#2\@@{% override\bbl@inidate#1...\relax{#2}}\fi
\bbl@csarg\def{secpre@date.gregorian.licr}{% discard uni\ifcase\bbl@engine\let\bbl@savedate\@empty\fi}

% eg: 1=months, 2=wide, 3=1, 4=dummy
\def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'\bbl@trim\bbl@inidate#1...\relax{#2}{#3}{#4}{#5}{#6}}
\bbl@ifsamestring\bbl@tempa{months.wide}\% to savedate
{\bbl@trim@def\bbl@tempa{#3}\%
\bbl@trim\toks@{#5}\%
\bbl@exp{\\\bbl@add\\bbl@savedate{\%
\SetString\month\romannumeral\bbl@tempa#6name\{\the\toks@\})){
{\bbl@ifsamestring\bbl@tempa{date.long}\% defined now
{\bbl@trim@def\bbl@toreplace{#5}\%
\bbl@TG@\date
\global\bbl@csarg\let{date@\languagename}\bbl@toreplace
\bbl@exp{\gdef\languagename date\{}{\protect\languagename date }\%
\gdef\languagename date>####1####2####3{\bbl@usedategrouptrue
\bbl@ensure\languagename\%
\bbl@date\languagename\####1\####2\####3}%
\bbl@add\bbl@savetoday{\%
\SetString\today{\the\year}{\the\month}{\the\day}}}}}%

{}\}

Dates will require some macros for the basic formatting. They may be redefined by
language, so “semi-public” names (camel case) are used. Oddly enough, the CLDR places
particles like “de” inconsistently in either the date or in the month name.
\let\bbl@calendar@\empty
\newcommand\BabelDateSpace{\nobreakspace}
\newcommand\BabelDateDot{.\@}
\newcommand\BabelDated[1]{\number#1}
\newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
\newcommand\BabelDateM[1]{\number#1}
\newcommand\BabelDateMM[1]{\number#1}
\newcommand\BabelDateMMMM[1]{\csname month\romannumeral#1\bbl@calendar name\endcsname}}\%
\newcommand\BabelDatey[1]{\number#1}\%
\newcommand\BabelDateyy[1]{\ifnum#1<10 0\fi\number#1 \%
\else\ifnum#1<100 \number#1 \%
\else\ifnum#1<1000 \expandafter@ gobble\number#1 \%
\else\ifnum#1<10000 \expandafter@ gobble\two\number#1 \%
\else\bbl@error
\bbl@error{Currently two-digit years are restricted to the\%
\range 0-9999.}\%
\bbl@error{There is little you can do. Sorry.}\%
\bbl@error{The two-digit years are restricted to the\%
\range 0-9999.}\%
\bbl@error{There is little you can do. Sorry.}\%}

\newcommand\BabelDateyyyy[1]{\number#1} % FIXME - add leading 0
\def\bbl@replace@finish@iii#1{\def#1####1####2####3{\the\toks@}}
\def\bbl@TG@\date{\bbl@replace\bbl@toreplace{[ ]}{\BabelDateSpace{}}\%
\bbl@replace\bbl@toreplace{[]}{\BabelDateDot{}}\%
\bbl@replace\bbl@toreplace{[d]}{\BabelDated{####3}}\%
\bbl@replace\bbl@toreplace{[dd]}{\BabelDatedd{####3}}\%
\bbl@replace\bbl@toreplace{[M]}{\BabelDateM{####2}}\%
\bbl@replace\bbl@toreplace{[MM]}{\BabelDateMM{####2}}\%
\bbl@replace\bbl@toreplace{[MMMM]}{\BabelDateMMMM{####2}}\%
\bbl@replace\bbl@toreplace{[y]}{\BabelDatey{####1}}\%
\bbl@replace\bbl@toreplace{[yy]}{\BabelDateyy{####1}}\%
\bbl@replace\bbl@toreplace{[yyyy]}{\BabelDateyyyy{####1}}\%
Language and Script values to be used when defining a font or setting the direction are set with the following macros.

```latex
\def\bbl@provide@lsys#1{% \\
  \bbl@ifunset{bbl@lname@#1}{}% \\
  \bbl@csarg\let{lsys@#1}\@empty% \\
  \bbl@ifunset{bbl@sname@#1}{\bbl@csarg\gdef{sname@#1}{Default}}{}% \\
  \bbl@ifunset{bbl@sotf@#1}{\bbl@csarg\gdef{sotf@#1}{DFLT}}{}% \\
  \bbl@csarg\bbl@add@list{lsys@#1}{Script=\bbl@cs{sname@#1}}% \\
  \bbl@ifunset{bbl@lname@#1}{}% \\
  \bbl@csarg\bbl@add@list{lsys@#1}{Language=\bbl@cs{lname@#1}}% \\
  \ifcase\bbl@engine\or\or% \\
    \bbl@ifunset{bbl@prehc@#1}{}% \\
    \exp{\bbl@ifblank{\bbl@cs{prehc@#1}}}{}% \\
  \fi% \\
  \bbl@csarg\bbl@toglobal{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 (which means any code in it must be skipped, too).

```latex
\def\bbl@ini@basic#1{% \\
  \def\BabelBeforeIni##1##2{\begingroup\bbl@add\bbl@secpost@identification{\closein\bbl@readstream} \catcode`\=1 \catcode`\|=12 \catcode`\=12 \catcode`\;=12 \catcode`\,=12} \bbl@read@ini{##1}{font and identification data}\endinput % babel-.tex may contain only preamble's \endgroup}% boxed, to avoid extra spaces:% \\
\setbox\z@\hbox{\InputIfFileExists{babel-#1.tex}{}{}}}
```

Alphabetic counters must be converted from a space separated list to an \texttt{ifcase} structure.

```latex
\def\bbl@buildifcase#1 {% Returns \bbl@tempa, requires \toks@={% \\
  \bbl@add\bbl@secpost@identification{\closein\bbl@readstream}% \\
  \def\bbl@exp{\def\bbl@tmpa##1##2{% \\
    \ifcase##1###1\space\the\toks@\else\\@ctrerr<\texttt{if}>\fi}}% \\
  \else% \\
    \toks@\expandafter{\the\toks@\or #1}% \\
  \expandafter\bbl@buildifcase% \\
  \fi% \\
  \def\bbl@buildifcase#1%}
```

The code for additive counters is somewhat tricky and it's based on the fact the arguments just before \texttt{\@} collects digits which have been left 'unused' in previous arguments, the first of them being the number of digits in the number to be converted. This explains the reverse set 76543210. Digits above 10000 are not handled yet. When the key contains the subkey \texttt{.F.}, the number after is treated as a special case. for a fixed form (see babel-he.ini, for example).

```latex
\newcommand\localenumeral[2]{\bbl@cs{cntr@#1\languagename}{#2}}\def\bbl@localecntr#1#2{\localenumeral{#2}{#1}}
```

115
The information in the identification section can be useful, so the following macro just exposes it with a user command.

```
\newcommand{\localeinfo}[1]{% 
  \bbl@ifunset{bbl@\csname bbl@info@#1\endcsname @\languagename}{}% 
  \bbl@error{I've found no info for the current locale.\% 
    The corresponding ini file has not been loaded\% 
    Perhaps it doesn't exist}\% 
  \{\bbl@cs{\csname bbl@info@#1\endcsname @\languagename}\}% 
  \{See the manual for details.\}% 
  \{\bbl@cs{\csname bbl@info@#1\endcsname @\languagename}\}% 
}\% 
\edef{\bbl@info@tag.ini}{lini}\% 
\edef{\bbl@info@tag.bcp47}{lbcp}\% 
\edef{\bbl@info@tag.opentype}{lotf}\% 
\edef{\bbl@info@script.name}{esname}\% 
\edef{\bbl@info@script.name.opentype}{sname}\% 
\edef{\bbl@info@script.tag.bcp47}{sbcp}\% 
\edef{\bbl@info@script.tag.opentype}{sotf}\% 
\let{\bbl@ensureinfo}@gobble\% 
\newcommand{\BabelEnsureInfo}{% 
  \def{\bbl@ensureinfo}{\if\InputIfFileExists\@undefined\else \% 
    \bbl@ifunset{bbl@lname@#1}{\bbl@ini@basic{#1}}{}\% 
  \fi}\% 
\}\% 
```

More general, but non-expandable, is \getlocaleproperty. To inspect every possible loaded ini, we define \LocaleForEach, where \bbl@ini@loaded is a comma-separated list of locales, built by \bbl@read@ini.

```
\newcommand{\getlocaleproperty}[3]{% 
  \let#1@relax\% 
  \def{\bbl@elt}{\let#1@relax\% 
    \ifnum#1>0\% 
      \bbl@error{Alphabetic numeral too large (#1)}\% 
      \{Currently this is the limit.\}% 
    \fi}\% 
  \{\bbl@elt{#2}{#3}\}% 
}\% 
\let{\bbl@ensureinfo}@gobble\% 
\newcommand{\BabelEnsureInfo}{% 
  \def{\bbl@ensureinfo}{\if\InputIfFileExists\@undefined\else \% 
    \bbl@ifunset{bbl@lname@#1}{\bbl@ini@basic{#1}}{}\% 
  \fi}\% 
\}\% 
```

The information in the identification section can be useful, so the following macro just exposes it with a user command.

```
\newcommand{\localeinfo}[1]{% 
  \bbl@ifunset{bbl@\csname bbl@info@#1\endcsname @\languagename}{}% 
  \bbl@error{I've found no info for the current locale.\% 
    The corresponding ini file has not been loaded\% 
    Perhaps it doesn't exist}\% 
  \{\bbl@cs{\csname bbl@info@#1\endcsname @\languagename}\}% 
  \{See the manual for details.\}% 
  \{\bbl@cs{\csname bbl@info@#1\endcsname @\languagename}\}% 
}\% 
\edef{\bbl@info@tag.ini}{lini}\% 
\edef{\bbl@info@tag.bcp47}{lbcp}\% 
\edef{\bbl@info@tag.opentype}{lotf}\% 
\edef{\bbl@info@script.name}{esname}\% 
\edef{\bbl@info@script.name.opentype}{sname}\% 
\edef{\bbl@info@script.tag.bcp47}{sbcp}\% 
\edef{\bbl@info@script.tag.opentype}{sotf}\% 
\let{\bbl@ensureinfo}@gobble\% 
\newcommand{\BabelEnsureInfo}{% 
  \def{\bbl@ensureinfo}{\if\InputIfFileExists\@undefined\else \% 
    \bbl@ifunset{bbl@lname@#1}{\bbl@ini@basic{#1}}{}\% 
  \fi}\% 
\}\% 
```

More general, but non-expandable, is \getlocaleproperty. To inspect every possible loaded ini, we define \LocaleForEach, where \bbl@ini@loaded is a comma-separated list of locales, built by \bbl@read@ini.
10 Adjusting the Babel behavior

A generic high level interface is provided to adjust some global and general settings.

\newcommand\babeladjust[1]{% TODO. Error handling.
\bbl@forkv{#1}{\bbl@cs{ADJ@##1@##2}}}

%\def\bbl@adjust@lua#1#2{%
%\ifvmode
%\ifnum\currentgrouplevel=\z@
%\directlua{ Babel.#2 }%
%\expandafter\expandafter\expandafter\@gobble
%\fi
%\fi
%{\bbl@error % The error is gobbled if everything went ok.
%{Currently, #1 related features can be adjusted only\%}
% in the main vertical list.}%
%{Maybe things change in the future, but this is what it is.}}%
@namedef{bbl@ADJ@bidi.mirroring@on}{%}
@namedef{bbl@ADJ@bidi.mirroring@off}{%}
@namedef{bbl@ADJ@bidi.text@on}{%}
@namedef{bbl@ADJ@bidi.text@off}{%}
@namedef{bbl@ADJ@bidi.mapdigits@on}{%}
@namedef{bbl@ADJ@bidi.mapdigits@off}{%}
@namedef{bbl@ADJ@linebreak.sea@on}{%}
@namedef{bbl@ADJ@linebreak.sea@off}{%}
@namedef{bbl@ADJ@linebreak.cjk@on}{%}
@namedef{bbl@ADJ@linebreak.cjk@off}{%}
\def\bbl@adjust@layout#1{%
\ifvmode
#1%
\expandafter\@gobble
\fi
\fi
{%\bbl@error % The error is gobbled if everything went ok.
%{Currently, #1 related features can be adjusted only\%}
% in the main vertical list.}%
%{Maybe things change in the future, but this is what it is.}}%
11 The kernel of Babel (babel.def for \LaTeX{}Only)

11.1 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.

11.2 Cross referencing macros

The \LaTeX{} book states:

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

When the above quote should still be true when a document is typeset in a language that has active characters, special care has to be taken of the category codes of these characters when they appear in an argument of the cross referencing macros.

When a cross referencing command processes its argument, all tokens in this argument should be character tokens with category `letter' or `other'.

The only way to accomplish this in most cases is to use the trick described in the \TeX{}book [4] (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 The macro \label{} writes a line with a \newlabel{} command into the .aux file to define labels.
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:

\begin{verbatim}
\DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
\DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
\DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
\end{verbatim}

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

An \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.

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

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

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

\ifs\bbl@tempa\bbl@tempb\else \atempswatrue \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@redefine\@citex\[#1\]#2{\@safe@activestrue\edef\@tempa{#2}\@safe@activesfalse\org@@citex\[#1\]\[#2\]{\@tempa}}%
\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{#3}\@safe@activesfalse\org@@citex[#1][#2]{\@tempa}}%
\else\let\org@citex\cite\let\org@pageref\pageref\fi

Unfortunately, the packages natbib and cite need a different definition of \@citex... To begin with, natbib has a definition for \@citex with three arguments... We only know that a package is loaded when \begin{document} is executed, so we need to postpone the different redefinition.

\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{#3}\@safe@activesfalse\org@@citex[#1][#2]{\@tempa}}%

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

\AtBeginDocument{%
\ifpackageloaded{cite}%
\def\@citex[#1][#2]{%
The macro \nocite which is used to instruct BiBTeX to extract uncited references from the database.
\bbl@redefine
ocite#1{%\@safe@activesfalse\cite{#1}\@safe@activesfalse}%
\bibcite
The macro that is used in the .aux file to define citation labels. When packages such as natbib or cite are not loaded its second argument is used to typeset the citation label. In that case, this second argument can contain active characters but is used in an environment where \@safe@activestrue is in effect. This switch needs to be reset inside the \hbox which contains the citation label. In order to determine during .aux file processing which definition of \bibcite is needed we define \bibcite in such a way that it redefines itself with the proper definition. We call \bbl@cite@choice to select the proper definition for \bibcite. This new definition is then activated.
\bbl@redefine\bibcite{%\@bibcite#1#2}{\org@bibcite{#1}{\@safe@activesfalse#2}}
\bbl@cite@choice
The macro \bbl@cite@choice determines which definition of \bibcite is needed. First we give \bibcite its default definition.
\def\bbl@cite@choice{%\global\let\bibcite\bbl@bibcite
Then, when natbib is loaded we restore the original definition of \bibcite. For cite we do the same.
\@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{%\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 \LaTeX 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
11.3 Marks

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

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

\bbl@trace{Marks}
\IfBabelLayout{sectioning}
\{\iffx\bbl@opt@headfoot\@nnil
\g@addto@macro\@resetactivechars{%
\set@typeset@protect\expandafter\select@language@x\expandafter{\bbl@main@language}%
\let\protect@noexpand@境\edef\thepage{%
\noexpand\babelsublr{\unexpanded\expandafter{\thepage}}}}%
\fi\}
\{\iffbbl@single\else
\bbl@ifunset{markright }\bbl@redefine\bbl@redefinerobust\markright#1{%
\bbl@ifblank{#1}%
{\org@markright{}%}
{\toks@{#1}%
\bbl@exp{%
\org@markright{\protect\foreignlanguage{\languagename}{\protect\bbl@restore@active\the\toks@}}}}%}
\fi\}
\{\iffbbl@single\else
\bbl@ifunset{markboth }\bbl@redefine\bbl@redefinerobust\markboth#1#2{%
\protected@edef\bbl@tempc%{
\let\@mkboth\markboth
\def\bbl@tempc{\let\@mkboth\markboth}
\else
\def\bbl@tempc{}
\fi
\bbl@ifunset{markboth }\bbl@redefine\bbl@redefinerobust\markboth#1#2{%
\protected@edef\bbl@tempb#1{%
\protect\foreignlanguage{\languagename}{\protect\bbl@restore@active#1}%
\bbl@ifblank{#1}%
{\toks@{}}%
{\toks@{\expandafter{\bbl@tempb{#1}}}}%
\bbl@ifblank{#2}%
{\@emptokena}%
{\@emptokena}{\expandafter{\bbl@tempb{#2}}}%
\bbl@exp{\\org@markboth{\the\toks@}{\the\@emptokena}}%
\bbl@tempc
\fi} % end ifbbl@single, end \IfBabelLayout
11.4 Preventing clashes with other packages

11.4.1 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:

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

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.

\begin{verbatim}
\AtBeginDocument{\ifin@ \AtBeginDocument{\@ifpackageloaded{ifthen}{%}
  \let\temp@pref\pageref
  \let\pageref\org@pageref
  \let\temp@ref\ref
  \let\ref\org@ref
  \@safe@activesfalse
  \org@ifthenelse{#1}\
  {\let\pageref\bbi@temp@pref
    \let\ref\bbi@temp@ref
    \@safe@activesfalse
    \@safe@activesfalse
    \@safe@activesfalse
    \@safe@activesfalse
    #2}
  {\let\pageref\bbi@temp@pref
    \let\ref\bbi@temp@ref
    \@safe@activesfalse
    #3}}%
}
\end{verbatim}

Then we can redefine \ifthenelse:

\begin{verbatim}
  \bbi@redefine@long\ifthenelse#1#2#3{%
  \AtBeginDocument{%
    \@ifpackageloaded{ifthen}{%}

  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.

\begin{verbatim}
\let\temp@pref\pageref
\let\pageref\org@pageref
\let\temp@ref\ref
\let\ref\org@ref
\end{verbatim}

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.

\begin{verbatim}
\@safe@activesfalse
\org@ifthenelse(#1)\{
  \let\pageref\bbi@temp@pref
  \let\ref\bbi@temp@ref
  \@safe@activesfalse
  #2\{
  \let\pageref\bbi@temp@pref
  \let\ref\bbi@temp@ref
  \@safe@activesfalse
  #3\}
  \}
\}
\end{verbatim}

11.4.2 varioref

\@@vpageref When the package varioref is in use we need to modify its internal command \@@vpageref in order to prevent problems when an active character ends up in the argument of \vref. The same needs to happen for \vrefpagemum.
The package `variorref` defines \Ref to be a robust command which uppercases the first character of the reference text. In order to be able to do that it needs to access the expandable form of \ref. So we employ a little trick here. We redefine the (internal) command \Ref, to call \orgref instead of \ref. The disadvantage of this solution is that whenever the definition of \Ref changes, this definition needs to be updated as well.

```latex
\expandafter\def\csname Ref \endcsname#1{\protect\edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
```

### 11.4.3 hhline

Crosslines in tables are created with the `hhline` package. This package is fairly simple but it has one major drawback: it uses the `:` character which, in the default `babel` configuration, is made active by the `frenchsupport` package. Therefore we need to `reload` the package when the `:` is an active character.

So at \begin{document} we check whether `hhline` is loaded.

```latex
\AtEndOfPackage{\AtBeginDocument{%\ifpackage{hhline}%

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

```latex
\expandafter\ifx\csname normal@char\endcsname:\relax
\else
\fi
```

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 temporarily to change it to letter again.

```latex
\makeatletter
\def@currname{hhline}\input{hhline.sty}\makeatother
```

### 11.4.4 hyperref

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

```latex
\AtBeginDocument{%\if\pdfstringdefDisableCommands\undefined\else\pdfstringdefDisableCommands{\languageshorthands{system}}%
```
11.4.5 fancyhdr

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

```latex
\DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
  \lowercase{\foreignlanguage{#1}}}
```

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

```latex
\def\substitutefontfamily#1#2#3{%
  \lowercase{\immediate\openout15=#1#2.fd\relax}\
  \immediate\write15{\string\ProvidesFile{#1#2.fd}[^J
  \string\DeclareFontFamily{#1}{#2}{}[m]{n}{<->ssub * #3/m/n}[^J
  \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}[^J
  \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}[^J
  \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}[^J
  \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}[^J
  \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}[^J
  \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}[^J
  \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}[^J
  ]%}
  \closeout15}
```

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

11.5 Encoding and fonts

Because documents may use non-ASCII font encodings, we make sure that the logos of Te\TeX{} and L\TeX{} always come out in the right encoding. There is a list of non-ASCII encodings.

Unfortunately, `fontenc` deletes its package options, so we must guess which encodings has been loaded by traversing `@filelist` to search for ⟨enc⟩enc.def. If a non-ASCII has been loaded, we define versions of \TeX{} and \LaTeX{} for them using `\ensureascii`. The default ASCII encoding is set, too (in reverse order): the “main” encoding (when the document begins), the last loaded, or OT1.

```latex
\ensureascii
```
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 font encoding 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 \atfplatformloaded is disabled for this package. Now we have to rework parsing the internal macro \filelist which contains all the filenames loaded.

\AtBeginDocument{%
\if@fplatformloaded(fontspec)%
 {\xdef\latinencoding{%
 \ifdef\UTFencname@undefined
 EU@ifcase\bbl@engine\or2\or1\fi
 \else
 \UTFencname
 \fi}%
 %}
}{\edef\latinencoding{OT1}%
 \ifx cf@encoding\bbl@t@one
 \xdef\latinencoding{\bbl@t@one}%
 \else
 \ifx@fontenc@load@list@undefined
 \ifx@aded{def}{t1enc}{\xdef\latinencoding{\bbl@t@one}}{{}\}{}% 
 \else
 \def\@elt{#1}{,1,}%
 \edef\bbl@tempa{\expandafter\@gobbletwo\@fontenc@load@list}
 \let\@elt\relax
 \bbl@xin{,T1,}\bbl@tempa
 \ifin@% 
 \xdef\latinencoding{\bbl@t@one}%
 %}
 \fi
 \fi\}%
\%}
\fi
\latintext Then we can define the command \textlatin which is a declarative switch to a latin font-encoding. Usage of this macro is deprecated.

\begin{verbatim}
def\encodingdefault{\latinencoding}
\end{verbatim}

\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}
def\textlatin[1]{\leavevmode{\latintext #1}}
endverbatim

\textlatinThis 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.

\textlatinWork in progress. This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.

This code is currently placed here for practical reasons.
The following command is executed only if there is a right-to-left script (once). It activates
the \everypar hack for xetex, to properly handle the par direction. Note text and par dirs
are decoupled to some extent (although not completely).

\def\bbl@mathboxdir{%
  \ifcase\bbl@thetextdir\relax
    \everyparbox{\textdir TLT\relax}%
  \else
    \everyparbox{\textdir TRT\relax}%
  \fi}
\else % pdftex=0, xetex=2
  \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
  \DisableBabelHook{babel-bidi}
  \newcount\bbl@dirlevel
  \chardef\bbl@thetextdir\z@
  \chardef\bbl@thepardir\z@
  \def\bbl@textdir#1{%
    \ifcase#1\relax
      \chardef\bbl@thetextdir\z@
      \bbl@textdir@i\beginL\endL
    \else
      \chardef\bbl@thetextdir\@ne
      \bbl@textdir@i\beginR\endR
    \fi}
  \def\bbl@textdir@i#1#2{%\ifhmode
    \ifnum\currentgrouplevel>\z@
      \ifnum\currentgrouplevel=\bbl@dirlevel
        \bbl@error{Multiple bidi settings inside a group}\%{I'll insert a new group, but expect wrong results.}\%
        \bgroup\aftergroup#2\aftergroup\egroup
      \else
        \chardef\currentgrouptype\or % 0 bottom
        \aftergroup#2 % 1 simple {}
      \or
        \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
      \or
        \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
      \or
        \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
      \or
        \bgroup\aftergroup#2\aftergroup\egroup % output math disc insert vcent mathchoice
      \else
        \aftergroup#2 % 14 \begingroup
      \fi
      \bbl@dirlevel\currentgrouplevel
    \fi
  \fi
  \def\bbl@pardir#1{\chardef\bbl@thepardir#1\relax}
\def\bbl@xeeverypar{
\ifcase\bbl@thepardir
\ifcase\bbl@thetextdir\else\begin{R}
\else
\setbox\z@\lastbox\begin{R}\box\z@
\fi}
\let\bbl@severypar\everypar
\newtoks\everypar
\everypar=\bbl@severypar
\bbl@severypar{\bbl@xeeverypar\the\everypar}
\def\bbl@tempb{
\let\bbl@textdir@i\@gobbletwo
\let\bbl@xebidipar\@empty
\AddBabelHook{bidi}{foreign}{
\def\bbl@tempa{\def\BabelText{####1}}
\ifcase\bbl@thetextdir
\expandafter\bbl@tempa\expandafter{\BabelText{\LR{####1}}}\else
\expandafter\bbl@tempa\expandafter{\BabelText{\RL{####1}}}\fi}
\def\bbl@pardir##1\ifcase##1\relax\setLR\else\setRL\fi}}
\@ifpackagewith{babel}{bidi=bidi}{\bbl@tempb}{}\@ifpackagewith{babel}{bidi=bidi-l}{\bbl@tempb}{}\@ifpackagewith{babel}{bidi=bidi-r}{\bbl@tempb}{}i

A tool for weak L (mainly digits). We also disable warnings with hyperref.
\DeclareRobustCommand\babelsublr[1]{\leavevmode{\bbl@textdir\z@#1}}
\AtBeginDocument{\ifx\pdfstringdefDisableCommands\@undefined\else
\ifx\pdfstringdefDisableCommands\relax\else
\pdfstringdefDisableCommands{\let\babelsublr\@firstofone}\else
\pdfstringdefDisableCommands{\let\babelsublr\@empty}}\fi}

11.7 Local Language Configuration

\loadlocalcfg At some sites it may be necessary to add site-specific actions to a language definition file. This can be done by creating a file with the same name as the language definition file, but with the extension .cfg. For instance the file norsk.cfg will be loaded when the language definition file norsk.1df is loaded. For plain-based formats we don’t want to override the definition of \loadlocalcfg from plain.def.
\bbl@trace{Local Language Configuration}
\@ifpackagewith{babel}{noconfigs}{\let\loadlocalcfg\@undefined}
{\let\loadlocalcfg\@gobble}
{\def\loadlocalcfg\@firstofone}
{\def\loadlocalcfg\@empty}

Just to be compatible with \LaTeX{} 2.09 we add a few more lines of code:
\@ifpackagewith{unexpandable@protect}{\undefined}
{\def\@unexpandable@protect\@undefined}
{\def\@unexpandable@protect\@empty}

12 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.

\ProvidesFile{switch.def}[\texttt{\date} \texttt{\version} Babel switching mechanism]

\ProvidesFile{switch.def}[\texttt{\date} \texttt{\version} Load macros for plain if not \LaTeX\]

\ProvidesFile{switch.def}[\texttt{\date} \texttt{\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.

\bl@version
\bl@date
\adddialect
\bbl@version
\bbl@date
\begin{group}
\count@1\relax
\def\bbl@elt##1##2##3##4{% 
\ifnum\count@=##2\relax \bbl@info{\string#1 = using hyphenrules for \string\language\the\count@} \def\bbl@elt####1####2####3####4{}% \fi}
\bbl@cs{languages}
\end{group}

\bl@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 MYLANG, but unfortunately mixed case names cannot be trapped). Note \texttt{l@} is encapsulated, so that its case does not change.

\bbl@fixname
\begin{group}
\def\bbl@temp{\noexpand\@ifundefined{\noexpand\bbl@tempe\texttt{l@}}{}}
\edef\bbl@temp{\lowercase\expandafter{\bbl@temp}
\uppercase\expandafter{\bbl@temp}}
\expandafter{\@empty}
\end{group}

\texttt{\bbl@iflanguage}
Users might want to test (in a private package for instance) which language is currently active. For this we provide a test macro, \iflanguage, that has three arguments. It checks whether the first argument is a known language. If so, it compares the first argument with the value of \language. Then, depending on the result of the comparison, it executes either the second or the third argument.

12.1 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.

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

As \TeX 2.09 writes to files expanded whereas \TeX 2ε takes care not to expand the arguments of \write statements we need to be a bit clever about the way we add information to .aux files. Therefore we introduce the macro \xstring which should expand to the right amount of \string's.
Since version 3.5 babel writes entries to the auxiliary files in order to typeset table of contents etc. in the correct language environment.

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

The previous solution works for one level of nesting groups, but as soon as more levels are used it is no longer adequate. For that case we need to keep track of the nested languages using a stack mechanism. This stack is called \bbl@language@stack and initially empty.

When using a stack we need a mechanism to push an element on the stack and to retrieve the information afterwards.

The stack is simply a list of languagenames, separated with a ‘+’ sign; the push function can be simple:

\bbl@pop@language
\bbl@language@stack
\bbl@push@language
\bbl@language@stack%
\def\bbl@push@language{%\def\bbl@language@stack{\languagename+\bbl@language@stack}}

Retrieving information from the stack is a little bit less simple, as we need to remove the element from the stack while storing it in the macro \languagename. For this we first define a helper function.

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

\bbl@language@stack
\bbl@pop@language
\languagename\%\bbl@language@stack{}
\def\bbl@language@stack{\languagename+\bbl@language@stack}
\def\bbl@pop@lang#1+#2-#3{\edef\languagename{#1}\xdef#3{#2}}

The reason for the somewhat weird arrangement of arguments to the helper function is the fact it is called in the following way. This means that before \bbl@pop@lang is executed Tex first expands the stack, stored in \bbl@language@stack. The result of that is that the argument string of \bbl@pop@lang contains one or more language names, each followed by a ‘+’-sign (zero language names won’t occur as this macro will only be called after something has been pushed on the stack) followed by the ‘-’-sign and finally the reference to the stack.

\let\bbl@ifrestoring\@secondoftwo
\def\bbl@pop@language{%\expandafter\bbl@pop@lang\bbl@language@stack-\bbl@language@stack
\let\bbl@ifrestoring\@firstoftwo
\expandafter\bbl@set@language\expandafter{\languagename}}%
\def\bbl@pop@language@stack{}
\let\bbl@ifrestoring\@secondoftwo

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

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

\chardef\localeid\z@
\def\bbl@id@assign{%
  \bbl@ifunset{bbl@id@@\languagename}\%
  \count@\bbl@id@last\relax
  \advance\count@\@ne
  \bbl@csarg\chardef{id@@\languagename}\count@
  \edef\bbl@id@last{\the\count@}\%
  \ifcase\bbl@engine\or
    \directlua{
      Babel = Babel or {}
      Babel.locale_props = Babel.locale_props or {}
      Babel.locale_props[\bbl@id@last] = {}
      Babel.locale_props[\bbl@id@last].name = '\languagename'
    }%
    \%
    \fi}
  \chardef\localeid\bbl@cl{id@}}

The unprotected part of \selectlanguage.

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

\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 messed up. This is a bug, but preserved for backwards compatibility. The list of auxiliary files can be extended by redefining \BabelContentsFiles, but make sure they are loaded inside a group (as aux, toc, lof, and lot do) or the last language of the document will remain active afterwards.

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

\edef\BabelContentsFiles{toc,lof,lot}
\edef\bbl@set@language#1{% from selectlanguage, pop@
  \edef\languagename{%
    \ifnum\escapechar=\expandafter`\string#1\@empty
      \else\string#1\@empty\fi
    \select@language{\languagename}%
    % write to auxs
    \expandafter\ifx\csname date\languagename\endcsname\relax\else
      \if@files\w
        \ifx\bbl@aux\gobbletwo\else % Set if single in the first, redundant
          \protected@write@auxout{%\string\bbl@aux{\languagename}}%
          \fi
        \bbl@usehooks{write}{}
      \fi
      \fi}
  \def\select@language#{% from set@, babel@aux
    % set hmap
    \ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi
    % set name
    \edef\languagename{#1}%
    \bbl@fixname\languagename

134
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{lang} command at definition time by expanding the \csname primitive.

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

The switching of the values of \lefthyphenmin and \righthyphenmin is somewhat different. First we save their current values, then we check if \lang hyphenmins is defined. If it is not, we set default values (2 and 3), otherwise the values in \lang hyphenmins will be used.
\bl@id@assign
\% switch captions, date
\ifcase\bl@select@type
\ifhmode
\hskip\z@skip % trick to ignore spaces
\csname captions#1\endcsname\relax
\csname date#1\endcsname\relax
\loop\ifdim\lastskip>\z@\unskip\repeat\unskip
\else
\csname captions#1\endcsname\relax
\csname date#1\endcsname\relax
\fi
\else
\if\bl@usedategroup % if \foreign... within \lang>date
\bl@usedategroupfalse
\ifhmode
\hskip\z@skip % trick to ignore spaces
\csname date#1\endcsname\relax
\loop\ifdim\lastskip>\z@\unskip\repeat\unskip
\else
\csname date#1\endcsname\relax
\fi
\fi
\fi
\fi
\% switch extras
\bl@usehooks{beforeextras}{}%
\csname extras#1\endcsname\relax
\bl@usehooks{afterextras}{}%
\% > babel-ensure
\% > babel-sh-<short>
\% > babel-bidi
\% > babel-fontspec
\% hyphenation - case mapping
\ifcase\bl@opt@hyphenmap\or
\def\BabelLower##1##2{\lccode##1=##2\relax}\
\ifnum\bl@hymapsel>4\else
\csname\languagename @bbl@hyphenmap\endcsname
\fi
\chardef\bbl@opt@hyphenmap\z@
\else
\ifnum\bl@hymapsel>\bbl@opt@hyphenmap\else
\csname\languagename @bbl@hyphenmap\endcsname
\fi
\global\let\bbl@hymapsel\@cclv
\% hyphenation - patterns
\global\let\bbl@hyphenmap@e@clv
\% hyphenation - mins
\babel@savevariable\lefthyphenmin
\babel@savevariable\righthyphenmin
\expandafter\ifx\csname #1hyphenmins\endcsname\relax
\set@hyphenmins\tw@\thr@@\relax
\else
\expandafter\expandafter\expandafter\set@hyphenmins\csname #1hyphenmins\endcsname\relax
\fi
\fi
\otherlanguage {The otherlanguage environment can be used as an alternative to using the \selectlanguage declarative command. When you are typesetting a document which

\end{otherlanguage}
mixes left-to-right and right-to-left typesetting you have to use this environment in order to let things work as you expect them to. The \ignorespaces command is necessary to hide the environment when it is entered in horizontal mode.

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

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

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

The \foreignlanguage command is another substitute for the \selectlanguage command. This command takes two arguments, the first argument is the name of the language to use for typesetting the text specified in the second argument. Unlike \selectlanguage this command doesn’t switch everything, it only switches the hyphenation rules and the extra definitions for the language specified. It does this within a group and assumes the \extras{lang} command doesn’t make any \global changes. The coding is very similar to part of \selectlanguage.

\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).

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

In other words, at the beginning of a paragraph \foreignlanguage enters into hmode with the surrounding lang, and with \foreignlanguage* with the new lang.

\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{\noexpand\leavevmode}
\foreignlanguage 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.

\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 \bbl@hyphenation, so do nothing with this value. If the exceptions for a language (by its number, not its name, so that :ENC is taken into account) has been set, then use \hyphenation with both global and language exceptions and empty the latter to mark they must not be set again.
The environment `hyphenrules` can be used to select *just* the hyphenation rules. This environment does *not* change `\language` name and when the hyphenation rules specified were not loaded it has no effect. Note however, `/lccode`'s and font encodings are not set at all, so in most cases you should use other language.*.

```
def\hyphenrules#1{%
edef\bbl@tempf(#1)%
\bbl@fixname\bbl@tempf
\bbl@iflanguage\bbl@tempf{%\expandafter\patterns\expandafter{\bbl@tempf}%
\languageshorthands{none}%
\expandafter\expandafter\expandafter\set@hyphenmins\csname\bbl@tempf hyphenmins\endcsname%
\@ifundefined{bbl@hyphenation@#1}%
\@empty
{\space csname bbl@hyphenation@#1 \endsname}%
\nxdef\bbl@hyphlist{\bbl@hyphlist number \language ,}%
\fi
\endgroup
}"

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

```
def\providehyphenmins#1#2{%
exdef\bbl@tempf{#1}{#2}%
\expandafter\expandafter\expandafter\set@hyphenmins\csname\bbl@tempf hyphenmins\endcsname%
\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
\xdef\bbl@hyphlist{\bbl@hyphlist number \language ,}%
\fi
\set@hyphenmins
This macro sets the values of `\lefthyphenmin` and `\righthyphenmin`. It expects two values as its argument.

```
def\set@hyphenmins#1#2{%
\let\endhyphenrules@empty
\providelanguage
The macro `\providelanguage` should be used in the language definition files to provide a *default* setting for the hyphenation parameters `\lefthyphenmin` and `\righthyphenmin`. If the macro `/⟨lang⟩/language` is already defined this command has no effect.

```
def\providelanguage#1{%
exdef\bbl@tempf{#1}{#2}%
\expandafter\expandafter\expandafter\set@hyphenmins\csname\bbl@tempf hyphenmins\endcsname%
\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
\xdef\bbl@hyphlist{\bbl@hyphlist number \language ,}%
\fi
\set@hyphenmins
This macro sets the values of `\lefthyphenmin` and `\righthyphenmin`. It expects two values as its argument.

```
def\set@hyphenmins#1#2{%
\eqdef\bbl@tempf{#1}{#2}%
\expandafter\expandafter\expandafter\set@hyphenmins\csname\bbl@tempf hyphenmins\endcsname%
\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
\xdef\bbl@hyphlist{\bbl@hyphlist number \language ,}%
\fi
\set@hyphenmins
This macro sets the values of `\lefthyphenmin` and `\righthyphenmin`. It expects two values as its argument.

```
def\set@hyphenmins#1#2{%
\let\endhyphenrules@empty
```
\ProvidesLanguage The identification code for each file is something that was introduced in \TeX\ 2e. When the command \ProvidesFile does not exist, a dummy definition is provided temporarily. For use in the language definition file the command \ProvidesLanguage is defined by babel. Depending on the format, ie, on if the former is defined, we use a similar definition or not.

\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.

\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.

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\localename\setlocale
\let\textlocale\setlocale
\let\languagetext\setlocale

12.2 Errors

\@nolerr 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.

\@nopatterns When the package was loaded without options not everything will work as expected. An error message is issued in that case. When the format knows about \PackageError it must be \LaTeX, so we can safely use its error handling interface. Otherwise we’ll have to ‘keep it simple’. Infos are not written to the console, but on the other hand many people think warnings are errors, so a further message type is defined: an important info which is sent to the console.

\@noopterr

\def\bbl@nulllanguage{\string\language=0}\ifx\PackageError\@undefined\def\bbl@error#1#2{%\begingroup\newline\char=`\^^J\def\{%\MessageBreak\PackageError{ babel}{#1}{#2}\errhelp{#2}\errmessage{\#1}\endgroup\}\def\bbl@warning#1{%\begingroup\newline\char=`\^^J\def\{%\MessageBreak\PackageWarning{ babel}{#1}\message{\#1}\endgroup\}\let\bbl@infowarn\bbl@warning\def\bbl@info#1{%\begingroup\newline\char=`\^^J\def\{%\MessageBreak\PackageInfo{ babel}{#1}\endgroup\}\\else\def\bbl@error#1#2{%\begingroup\def\{%\MessageBreak\PackageError{ babel}{#1}{#2}\endgroup\}\\def\bbl@warning#1{%\begingroup\def\{%\MessageBreak\PackageWarning{ babel}{#1}\endgroup\}\\def\bbl@infowarn#1{%\begingroup\def\{%\MessageBreak\GenericWarning{( babel) \@spaces\@spaces\@spaces}{Package babel Info: \#1}\endgroup\}\\def\bbl@info#1{%\begingroup\def\{%\MessageBreak\PackageInfo{ babel}{#1}\endgroup\}}}
13 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.

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

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

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

- When someone wants to use several hyphenation patterns with Si\LaTeXX\ the above scheme won't work. The reason is that Si\LaTeX\ overwrites the contents of the every job register with its own message.

- Plain \TeX\ does not use the every job register so the message would not be displayed.

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

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

\begin{verbatim}
\ProvidesFile{hyphen.cfg}[\langle\langle date\rangle\rangle \langle\langle version\rangle\rangle Babel hyphens]
xdef\bbl@format{\jobname}
\ifx\AtBeginDocument\@undefined
\def\@empty{}
\let\orig@dump\dump
\def\dump{\ifx\@ztryfc\@undefined
\else
\toks0=\expandafter{\@preamblecmds}\edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}\def\@begindocumenthook{}\fi
\let\dump\orig@dump\let\orig@dump\@undefined\dump}\fi
\@empty{}
\toks@{}
\def\bbl@languages{}
\def\process@line#1#2 #3 #4 {\%\ifx=#1\%\process@synonym{#2}\%\else\process@language{#1#2}{#3}{#4}\%\fi\ignorespaces}
\def\process@synonym#1\%\bbl@languages{\toks@{#1}\def\bbl@languages{\ifx\toks@{\relax\process@synonym{#1\%\bbl@languages{\else\process@language{#1\%\bbl@languages{#3}{#4}\%\fi\ignorespaces}
\end{verbatim}

\texttt{\process@line} Each line in the file \texttt{language.dat} is processed by \texttt{\process@line} after it is read. The first thing this macro does is to check whether the line starts with \texttt{=}.

When the first token of a line is an \texttt{=}, the macro \texttt{\process@synonym} is called; otherwise the macro \texttt{\process@language} will continue.

\texttt{\process@synonym} This macro takes care of the lines which start with an \texttt{=}. It needs an empty token register to begin with. \texttt{\bbl@languages} is also set to empty.

When no languages have been loaded yet, the name following the \texttt{=} will be a synonym for hyphenation register \texttt{0}. So, it is stored in a token register and executed when the first pattern file has been processed. (The \texttt{\relax} just helps to the \texttt{\if} below catching synonyms without a language.) Otherwise the name will be a synonym for the language loaded last.

143
We also need to copy the hyphenmin parameters for the synonym.

\def\process@synonym#1{\
  \ifnum\last@language=\m@ne\
    \toks@\expandafter{\the\toks@\relax}\process@synonym{#1}\
  \else\
    \expandafter\chardef\csname l@#1\endcsname\last@language\
    \wlog{\string\l@#1=\string\language\the\last@language}\
    \expandafter\let\csname #1hyphenmins\expandafter\endcsname\csname\languagename hyphenmins\endcsname\
    \let\bbl@elt\relax\
    \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}{\{}{\}}{}}\
  \fi}

\process@language

The macro \process@language is used to process a non-empty line from the ‘configuration file’. It has three arguments, each delimited by white space. The first argument is the ‘name’ of a language; the second is the name of the file that contains the patterns. The optional third argument is the name of a file containing hyphenation exceptions. The first thing to do is call \addlanguage to allocate a pattern register and to make that register ‘active’. Then the pattern file is read.

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

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

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

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

When the hyphenation patterns have been processed we need to see if a file with hyphenation exceptions needs to be read. This is the case when the third argument is not empty and when it does not contain a space token. (Note however there is no need to save hyphenation exceptions into the format.) \bbl@languages saves a snapshot of the loaded languages in the form \bbl@elt{⟨language-name⟩}{⟨number⟩} {⟨patterns-file⟩}{⟨exceptions-file⟩}. Note the last 2 arguments are empty in ‘dialects’ defined in language.dat with =. Note also the language name can have encoding info.

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

Now, hooks are defined. For efficiency reasons, they are dealt here in a special way. Besides luatex, format-specific configuration files are taken into account.
The configuration file can now be opened for reading.

```
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}%
def\languagename{english}%
\iffalse
\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
Pattern registers are allocated using count register \last@language. Its initial value is 0. The definition of the macro \newlanguage is such that it first increments the count register and then defines the language. In order to have the first patterns loaded in pattern register number 0 we initialize \last@language with the value $-1$.
```

\last@language\m@ne

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

```
\loop
While reading from the input, it is useful to switch off recognition of the end-of-line character. This saves us stripping off spaces from the contents of the control sequence.
```

```
\endlinechar\m@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 \ifeof\fi \relax
\ifx\bbl@line\@empty\else
\edef\bbl@line{\bbl@line \space \space \space}\relax
\expandafter\process@line\bbl@line\relax
\fi
\fi
```

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
\fim
```

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
```
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
\let\process@language\undefined
\let\bbl@get@enc\undefined
\let\bbl@hyph@enc\undefined
\let\bbl@tempa\undefined
\let\bbl@hook@loadkernel\undefined
\let\bbl@hook@everylanguage\undefined
\let\bbl@hook@loadpatterns\undefined
\let\bbl@hook@loadexceptions\undefined
⟨\patterns⟩
```

Here the code for iniT EX ends.

# 14 Font handling with fontspec

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

```
{%More package options%} \iffoodoo\bbl@engine
\DeclareOption{bidi=basic-r}{%
\ExecuteOptions{bidi=basic}}
\DeclareOption{bidi=basic}{%
{\let\bbl@beforeforeign\leavevmode
% TODO - to locale_props, not as separate attribute
\newattribute\bbl@attr@dir
% I don't like it, hackish:
\frozen@everymath\expandafter{%
  \expandafter\bbl@mathboxdir\the\frozen@everymath}%
\frozen@everydisplay\expandafter{%
  \expandafter\bbl@mathboxdir\the\frozen@everydisplay}%
\bbl@exp{\output{\bodydir\pagedir\the\output}}%
\AtEndOfPackage{\EnableBabelHook{babel-bidi}}}
\else
\DeclareOption{bidi=basic-r}{%
\ExecuteOptions{bidi=basic}}
\DeclareOption{bidi=basic}{%
{\bbl@error
% The bidi method `basic' is available only in\%
% lualatex. I'll continue with `bidi=default', so\%
% expect wrong results}%
{See the manual for further details.}%
\let\bbl@beforeforeign\leavevmode
\AtEndOfPackage{%
\EnableBabelHook{babel-bidi}%
\bbl@xebidipar}}
\def\bbl@loadxebidi#1{%
% \ifx\RTLfootnotetext\@undefined
% \AtEndOfPackage{%
% \EnableBabelHook{babel-bidi}%
% \bbl@xebidipar}}
\ifx\RTLfootnotetext\@undefined
\AtEndOfPackage{%
\EnableBabelHook{babel-bidi}%
% bidi needs fontspec
```
3591 \fi
3592 \usepackage#1{bidi}%
3593 \fi}
3594 \DeclareOption{bidi=bidi}{
3595 {\bbl@tentative{bidi=bidi}%
3596 \bbl@loadxebidi{}%}
3597 \DeclareOption{bidi=bidi-r}{
3598 {\bbl@tentative{bidi=bidi-r}%
3599 \bbl@loadxebidi{[rldocument]}%}
3600 \DeclareOption{bidi=bidi-l}{
3601 {\bbl@tentative{bidi=bidi-l}%
3602 \bbl@loadxebidi{}%}
3603 \fi
3604 \DeclareOption{bidi=default}{%
3605 {\let\bbl@beforeforeign\leavevmode
3606 {\ifodd\bbl@engine
3607 \newattribute\bbl@attr@dir
3608 \bbl@exp{\output{\bodydir\pagedir\the\output}}%
3609 \fi}
3610 \AtEndOfPackage{%
3611 \EnableBabelHook{babel-bidi}%
3612 {\ifodd\bbl@engine\else
3613 \bbl@xebidipar
3614 \fi}}
3615 ((//More package options))
3616 With explicit languages, we could define the font at once, but we don't. Just wait and see if
3617 the language is actually activated. bbl@font replaces hardcoded font names inside
3618 \..family by the corresponding macro \..default.
3619 ((/*Font selection*/) =
3620 \bbl@trace{Font handling with fontspec}
3621 @onlypreamble\babelfont
3622 \newcommand\babelfont[2][]{1=langs/scripts 2=fam
3623 \bbl@foreach{#1}{%
3624 \expandafter\ifx\csname date##1\endcsname\relax
3625 \IfFileExists{babel-##1.tex}\
3626 {\babelprovide{##1}}{%
3627 \edef\bbl@tempa{#1}%
3628 \def\bbl@tempb{#2}% Used by \bbl@babelfont
3629 \ifx\fontspec\@undefined
3630 \usepackage{fontspec}%
3631 \EnableBabelHook{babel-fontspec} Just calls \bbl@switchfont
3632 \bbl@babelfont;
3633 \newcommand\babelfont[2][]{1=features 2=fontname, @font=rm|sf|tt
3634 \bbl@ifunset{\bbl@tempb family}\
3635 {\bbl@providefam{\bbl@tempb}}%
3636 {\bbl@exp{%
3637 \bbl@sreplace\bbl@tempb family >%
3638 {\@nameuse{\bbl@tempb default}}}%}
3639 % For the default font, just in case:
3640 \bbl@ifunset{\bbl@lsys@languagename}{\bbl@provide@lsys{\languagename}{}%}
3641 \expandafter\ifblank\expandafter{\bbl@tempa}%
3642 {\bbl@csarg\edef{\bbl@tempb delayed}<>{#1}{#2}}% save bbl@mdflt@
3643 \bbl@exp{%
3644 {\let\bbl@bblfont delayed \bbl@bblfont delayed \languagename}<\bbl@bblfont delayed \languagename>%
3645 \bbl@font@set<\bbl@bblfont delayed \languagename>%
If the family in the previous command does not exist, it must be defined. Here is how:

```latex
\def\bbl@providefam#1{\newcommand\bbl@#1default{}\bbl@add\bbl@font@fams{#1}\DeclareRobustCommand\bbl@#1family{\not@math@alphabet\bbl@#1family\relax\fontfamily\bbl@#1default\selectfont}}
```

The following macro is activated when the hook babel-fontspec is enabled. But before we define a macro for a warning, which sets a flag to avoid duplicate them.

```latex
\def\bbl@nostdfont#1{\bbl@ifunset{bbl@WFF\f@family}{\bbl@csarg\gdef{WFF\f@family}{}\bbl@infowarn{The current font is not a babel standard family:\%\#1\fontname\font\% There is nothing intrinsically wrong with this warning, and\% you can ignore it altogether if you do not need these\% families. But if they are used in the document, you should be\% aware ‘babel’ will no set Script and Language for them, so\% you may consider defining a new family with \string\babelfont.\% See the manual for further details about \string\babelfont.\% Reported}}\gdef\bbl@switchfont{\bbl@ifunset{bbl@lsys\languagename}{\bbl@provide@lsys{\languagename}}{}\lowercase{\edef\bbl@tempa{\bbl@cl{sname}}}\bbl@foreach\bbl@font@fams{\bbl@ifunset{bbl@##1dflt@\languagename}{\bbl@ifunset{bbl@##1dflt@\bbl@tempa}{\bbl@ifunset{bbl@##1dflt@}{\bbl@exp{\global\let\bbl@##1dflt@\languagename}{}\bbl@foreach\bbl@font@fams{\bbl@ifunset{bbl@##1dflt@\languagename}{\bbl@cs{famrst@##1}{\global\let\bbl@##1dflt@\languagename}}}}}}}}}
```

149
The following is executed at the beginning of the aux file or the document to warn about fonts not defined with \babelfont.

```latex
\ifx\f@family@undefined\else % if latex
  \ifcase\bbl@engine % if pdftex
    \let\bbl@ckeckstdfonts\relax
  \else
    \def\bbl@ckeckstdfonts{%
      \begingroup
        \global\let\bbl@ckeckstdfonts\relax
        \let\bbl@tempa\@empty
        \bbl@foreach\bbl@font@fams{%
          \bbl@ifunset{bbl@##1dflt@}%
            \@nameuse{##1family}%
            \def\gdef{WFF@f@family}{% Flag
              \bbl@exp{\\bbl@add\\bbl@tempa{\space\space\fontname\font\\%}
              \space\space\fontname\font\\}}%
          \bbl@exp{\expandafter\xdef\csname ##1dflt\endcsname{\f@family}}%
        }%
        \ifx\bbl@tempa\@empty\else
          \bbl@infowarn{The following font families will use the default\%
            settings for all or some languages:\%
            \bbl@tempa \%
            There is nothing intrinsically wrong with it, but\%
            'babel' will no set Script and Language, which could\%
            be relevant in some languages. If your document uses\%
            these families, consider redefining them with \string\babelfont.\%
            Reported}%
        \fi
      \endgroup}
  \fi
\fi
\fi
\fi
```

Now the macros defining the font with fontspec.

When there are repeated keys in fontspec, the last value wins. So, we just place the ini settings at the beginning, and user settings will take precedence. We must deactivate temporarily \bbl@mapselect because \selectfont is called internally when a font is defined.

```
\def\bbl@font@set#1#2#3{% eg \bbl@rmdflt@lang \rmdefault \rmfamily
  \bbl@xin{<>}{#1}%
  \ifin@
    \bbl@exp{\\bbl@fontspec@set\#1\expandafter\@gobbletwo#1\#3}%
  \fi
  \bbl@exp{\let\#2{#1} % eg, \rmdefault{\bbl@rmdflt@lang}
    \bbl@exp{\\\bbl@ifsamestring(#2){\f@family}{}\#3\let\bbl@tempa\relax{}}}%
  \% TODO - next should be global?, but even local does its job. I'm
  \% still not sure -- must investigate:
  \def\bbl@fontspec@set##1##2##3##4{% eg \bbl@rmdflt@lang fnt-opt fnt-nme \xxfamily
    \let\bbl@temppe\bbl@mapselect
    \let\bbl@mapselect@relax
    \let\bbl@temp@fam##4% eg, '\rmfamily', to be restored below
    \let##4@empty % Make sure \renewfontfamily is valid
    \bbl@exp{%
      \let\bbl@temp@fam\bbl@stripslash\space% eg, '\\rmfamily' \
      \keys_if_exist:nnF{fontspec-opentype}{Script/\bbl@cl{##nname}}%
      {\\newfontscript{\bbl@cl{##nname}}{\bbl@cl{soft}}}%
      \keys_if_exist:nnF{fontspec-opentype}{Language/\bbl@cl{##lname}}%
```

```
\newfontlanguage{\bbl@cl{lname}}{\bbl@cl{lotf}}
\renewfontfamily{#4}
[\bbl@cs{lsys@{\languagename},#2}]#{3} ie \bbl@exp{..}{#3}
\begingroup
#4
\xdef#1{\f@family} eg, \bbl@rmdflt@lang{FreeSerif(0)}
\endgroup
\let#4\bbl@temp@fam
\bbl@exp{\let\bbl@stripslash#4\space}>\bbl@temp@pfam
\let\bbl@mapselect\bbl@tempfam

font@rst and famrst are only used when there is no global settings, to save and restore de
previous families. Not really necessary, but done for optimization.
\def\bbl@font@fams{rm,sf,tt}

The default font families. They are eurocentric, but the list can be expanded easily with
\babelfont.
\def\bbl@font@rss{#1)#2}{#3#4}

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

15.1 XeTeX

Unfortunately, the current encoding cannot be retrieved and therefore it is reset always to utf8, which seems a sensible default. 
\LaTeX{} sets many “codes” just before loading hyphen.cfg. That is not a problem in luatex, but in xetex they must be reset to the proper value. Most of the work is done in \texttt{xelatex.ini}, so here we just “undo” some of the changes done by \LaTeX{}. Anyway, for consistency Lua\TeX{} also resets the catcodes.

```
\begin{verbatim}
\% Reset chars "80-"C0 to category "other", no case mapping:
\catcode`@=11 \count@=128
\loop\ifnum\count@<192
\global\uccode\count@=0 \global\lccode\count@=0
\global\catcode\count@=12 \global\sfcode\count@=1000
\advance\count@ by 1 \repeat
% Other:
\def\O ##1 {\global\uccode"##1=0 \global\lccode"##1=0
\global\catcode"##1=12 \global\sfcode"##1=1000 }
% Letter:
\def\L ##1 ##2 ##3 {\global\catcode"##1=11
\global\uccode"##1="##2 \global\lccode"##1="##3
% Uppercase letters have sfcode=999:
\ifnum"##1="##3 \else \global\sfcode"##1=999 \fi }
% Letter without case mappings:
\l 00AA
\L 00B5 039C 00B5
\l 00BA
\O 00D7
\l 00DF
\O 00F7
\l 00FF 0178 00FF
\end{verbatim}
```

Some more common code.

```
\begin{verbatim}
\bbl@trace{Bidi footnotes}
\ifx\bbl@beforeforeign\leavevmode
\def\bbl@footnote#1#2#3{\@ifnextchar[\bl@footnote@o{#1}{#2}{#3}\bbl@footnote@x{#1}{#2}{#3}}
\def\bbl@footnote@x#1#2#3#4{\bgroup\select@language@x{\bbl@main@language}]
\end{verbatim}
```

```
\def\bbl@footnote@o#1#2#3{\bbl@trace{Footnote changes}}
\bbl@trace{Bidi footnotes}
\ifx\bbl@beforeforeign\leavevmode
\def\bbl@footnote#1#2#3{%}
\@ifnextchar[\{\bbl@footnote@o{#1}{#2}{#3}\bbl@footnote@x{#1}{#2}{#3}}
\def\bbl@footnote@x#1#2#3#4{%}
\bgroup
\select@language@x{\bbl@main@language}]
```
Now, the code.

\def\BabelStringsDefault{unicode}
\let\xebbl@stop\relax
\AddBabelHook{xetex}{encodedcommands}{\def\bbl@tempa{#1}{%\ifx\bbl@tempa\@empty\XeTeXinputencoding"bytes"%\else\XeTeXinputencoding"#1"%\fi\def\xebbl@stop\relax}}
\AddBabelHook{xetex}{stopcommands}{\xebbl@stop\let\xebbl@stop\relax}
\def\bbl@intraspace#1 #2 #3\@@{\bbl@csarg\gdef{xeisp\@languagename}{\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}}
\def\bbl@intrapenalty#1\@@{\bbl@csarg\gdef{xeipn\@languagename}{\XeTeXlinebreakpenalty #1\relax}}
}\endinput
15.2 Layout

In progress.

Note elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titleps, and geometry. \bbl@startskip and \bbl@endskip are available to package authors. Thanks to the TeX expansion mechanism the following constructs are valid: \adim\bbl@startskip, \advance\bbl@startskip\adim. Thanksto the TeX expansion mechanism the following constructs are valid: \adim\bbl@startskip, \advance\bbl@startskip\adim.
Consider `txt babel` as a shorthand for `tex--xet babel`, which is the bidi model in both `pdftex` and `xetex`.

```latex
\providecommand\bbl@provide@intraspace{}
\bbl@trace{Redefinitions for bidi layout}
\def\bbl@sspre@caption{\bbl@exp{every hbox\\bbl@textdir\bbl@cs{wdir\bbl@main@language}}}
\ifx\bbl@opt@layout\@nnil\endinput \% No layout
\def\bbl@startskip{\ifcase\bbl@thepardir\leftskip\else\rightskip\fi}
\def\bbl@endskip{\ifcase\bbl@thepardir\rightskip\else\leftskip\fi}
\fi
\IfBabelLayout{lists}{\bbl@sreplace\list{\@totalleftmargin\leftmargin}{\@totalleftmargin\bbl@listleftmargin}\% pdftex doesn't reverse ()}
\def\p@enumiii{\p@enumii\)\theenumii(} \% A poor test for bidi=
\def\@hangfrom#1{%\setbox\@tempboxa\hbox{{#1}}\hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi\noindent\box\@tempboxa}
\def\raggedright{%\let\\@centercr\bbl@startskip\z@skip\bbl@endskip\z@skip\parindent\z@\parfillskip\bbl@startskip}
\def\raggedleft{%\let\\@centercr\bbl@startskip\@flushglue\bbl@endskip\z@skip\parindent\z@\parfillskip\bbl@endskip}
\fi
\IfBabelLayout{contents}{\bbl@sreplace\@dottedtocline{\leftskip}{\bbl@startskip}{\bbl@sreplace\@dottedtocline{\rightskip}{\bbl@endskip}}}
\IfBabelLayout{columns}{\bbl@sreplace\@outputdblcol{\hb@xt@\textwidth}{\bbl@outputhbox}\% A poor test for bidi=
\def\@hangfrom#1{%\setbox\@tempboxa\hbox{{#1}}\hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi\noindent\box\@tempboxa}
\def\raggedright{%\let\\@centercr\bbl@startskip\z@skip\bbl@endskip\z@skip\parindent\z@\parfillskip\bbl@startskip}
\def\raggedleft{%\let\\@centercr\bbl@startskip\@flushglue\bbl@endskip\z@skip\parindent\z@\parfillskip\bbl@endskip}
\fi
\IfBabelLayout{lists}{\bbl@sreplace\list{\@totalleftmargin\leftmargin}{\@totalleftmargin\bbl@listleftmargin}\% pdftex doesn't reverse ()}
\def\p@enumiii{\p@enumii\)\theenumii(} \% A poor test for bidi=
\def\@hangfrom#1{%\setbox\@tempboxa\hbox{{#1}}\hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi\noindent\box\@tempboxa}
\def\raggedright{%\let\\@centercr\bbl@startskip\z@skip\bbl@endskip\z@skip\parindent\z@\parfillskip\bbl@startskip}
\def\raggedleft{%\let\\@centercr\bbl@startskip\@flushglue\bbl@endskip\z@skip\parindent\z@\parfillskip\bbl@endskip}
\fi
\IfBabelLayout{contents}{\bbl@sreplace\@dottedtocline{\leftskip}{\bbl@startskip}{\bbl@sreplace\@dottedtocline{\rightskip}{\bbl@endskip}}}
\IfBabelLayout{columns}{\bbl@sreplace\@outputdblcol{\hb@xt@\textwidth}{\bbl@outputhbox}\% A poor test for bidi=
\def\@hangfrom#1{%\setbox\@tempboxa\hbox{{#1}}\hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi\noindent\box\@tempboxa}
\def\raggedright{%\let\\@centercr\bbl@startskip\z@skip\bbl@endskip\z@skip\parindent\z@\parfillskip\bbl@startskip}
\def\raggedleft{%\let\\@centercr\bbl@startskip\@flushglue\bbl@endskip\z@skip\parindent\z@\parfillskip\bbl@endskip}
\fi
```

155
Implicitly reverses sectioning labels in bidi=basic, because the full stop is not in contact with L numbers any more. I think there must be a better way.

\section{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, lua(e)tex is taken into account. Formerly, loading of patterns on the fly didn’t work in this format, but with the new loader it does. Unfortunately, the format is not based on babel, and data could be duplicated, because languages are reassigned above those in the format (nothing serious, anyway). Note even with this format language.dat is used (under the principle of a single source), instead of language.def.

Of course, there is room for improvements, like tools to read and reallocate languages, which would require modifying the language list, and better error handling.

We need catcode tables, but no format (targeted by babel) provide a command to allocate them (although there are packages like ctablestack). For the moment, a dangerous approach is used – just allocate a high random number and cross the fingers. To complicate things, etex.sty changes the way languages are allocated.
\begin{verbatim}
4024 (luatex)
4025 \ifx\AddBabelHook\@undefined
4026 \bbl@trace{Read language.dat}
4027 \ifx\bbl@readstream\@undefined
4028 \csname newread\endcsname\bbl@readstream
4029 \fi
4030 \begingroup
4031 \toks@{}
4032 \count@\z@ % 0=start, 1=0th, 2=normal
4033 \def\bbl@process@line#1#2 #3 #4 {%
4034 \ifx=#1%
4035 \bbl@process@synonym(#2)%
4036 \else
4037 \bbl@process@language(#1#2){#3}{#4}%
4038 \fi
4039 \ignorespaces
4040 \def\bbl@manylang{%
4041 \ifnum\count@>\@ne
4042 \bbl@info{Non-standard hyphenation setup}%
4043 \fi
4044 \let\bbl@manylang\relax}
4045 \def\bbl@process@language#1#2#3{%
4046 \ifcase\count@
4047 \ifundefined{zth@#1}{\count@\tw@}{\count@\@ne}%
4048 \or
4049 \count@\tw@
4050 \fi
4051 \ifnum\count@=\tw@
4052 \expandafter\addlanguage\csname l@#1\endcsname
4053 \language\allocationnumber
4054 \chardef\bbl@last\allocationnumber
4055 \bbl@manylang
4056 \let\bbl@elt\relax
4057 \edef\bbl@languages{%
4058 \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
4059 \fi
4060 \the\toks@
4061 \toks@{}}
4062 \def\bbl@process@synonym@aux#1#2{%
4063 \global\expandafter\chardef\csname l@#1\endcsname \l@#1\endcsname
4064 \language\allocationnumber
4065 \chardef\bbl@last\allocationnumber
4066 \bbl@manylang
4067 \let\bbl@elt\relax
4068 \edef\bbl@languages{%
4069 \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}\{#3}\}%
4070 \fi
4071 \the\toks@
4072 \toks@{}}
4073 \def\bbl@process@synonym@aux#1#2{%
4074 \ifcase\count@
4075 \toks@\expandafter{\the\toks@\relax\bbl@process@synonym@aux{#1}{#2}}%
4076 \or
4077 \ifundefined{zth@#1}{\bbl@process@synonym@aux{#1}{#2}{}}{\bbl@process@synonym@aux{#1}{#2}{0}}%
4078 \else
4079 \bbl@process@synonym@aux{#1}{\the\bbl@last}\%
4080 \fi
4081 \fi
4082 \ifx\bbl@languages\@undefined % Just a (sensible?) guess
4083 \chardef\l@english\z@
4084 \chardef\l@USenglish\z@
4085 \chardef\bbl@last\z@
4086 \gdef\bbl@languages{%
4087 \bbl@elt{english}{0}{hyphen.tex}{0}\{0\}%
4088 \bbl@elt{USenglish}{0}{hyphen.tex}{0}\{0\}%
4089 \edef\bbl@hyphen\relax
4090 \edef\bbl@hyphen@aux{}%
\def\bbl@patterns@lua#1{% 
\language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax 
  \csname l@#1\endcsname 
  \edef\bbl@tempa{#1}% 
\else 
  \csname l@#1:\f@encoding\endcsname 
  \edef\bbl@tempa{#1:\f@encoding}% 
\fi \relax 
\@namedef{lu@texhyphen@loaded@\the\language}{}% Temp 
  \@ifundefined{bbl@hyphendata@\the\language}% 
  {\def\bbl@elt##1##2##3##4{% 
    \ifnum##2=\csname l@\bbl@tempa\endcsname % #2=spanish, dutch:OT1...
      \def\bbl@tempb{##3}% 
      \ifx\bbl@tempb\@empty% 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 '\bbl@tempa'. Reported}}% 
  {\expandafter\expandafter\expandafter\bbl@luapatterns 
    \csname bbl@hyphendata@\the\language\endcsname}%}
\endinput\fi 
\begin{group}
\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
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.callback)
  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

\begin{verbatim}
        ss = ss .. '%d?' .. i
end
        ss = ss:gsub('^[%d%.]', '%d?')
        ss = ss:gsub('%d%.$', '%d?')
        pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
if n == 0 then
    tex.sprint(        [[\string\csname\space bbl@info\endcsname{New pattern: }]        .. p .. \[]])
        pats = pats .. ' ' .. p
else
    tex.sprint(        [[\string\csname\space bbl@info\endcsname{Renew pattern: }]        .. p .. \[]])
end
end
lang.patterns(lg, pats)
\end{verbatim}

\ifx\newattribute@undefined\else
\newattribute\bbl@attr@locale
\AddBabelHook{luatex}{beforeextras}{%        \setattribute\bbl@attr@locale\localeid}
\fi
\def\BabelStringsDefault{unicode}
\let\luabbl@stop\relax
\AddBabelHook{luatex}{encodedcommands}{%        \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@languages%        \@ifundefined{bbl@hyphendata@\the\language}{}{%        \\bbl@ech\csname\string\space bbl@info\space\string\endcsname{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}%        \ifin@%        \ifx\bbl@patterns@\@empty\else%        \directlua{ Babel.addpatterns(160
\end{verbatim}
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.
15.4 Southeast Asian scripts

First, some general code for line breaking, used by \baposthyphenation. In progress. Replace regular (i.e., implicit) discretionaries by spaceskips, based on the previous glyph (which I think makes sense, because the hyphen and the previous char go always together). Other discretionaries are not touched. For the moment, only 3 SA languages are activated by default (see Unicode UAX 14).

\directlua{
Babel = Babel or {}
Babel.linebreaking = Babel.linebreaking or {}
Babel.linebreaking.before = {}
Babel.linebreaking.after = {}
Babel.locale = {} % Free to use, indexed with \localeid
function Babel.linebreaking.add_before(func)
tex.print([\noexpand\csname bbl@luahyphenate\endcsname])
table.insert(Babel.linebreaking.before, func)
end
function Babel.linebreaking.add_after(func)
tex.print([\noexpand\csname bbl@luahyphenate\endcsname])
table.insert(Babel.linebreaking.after, func)
end
}
\def\bbl@intraspace#1 #2 #3\@@{%
\directlua{
Babel = Babel or {}
Babel.intraspaces = Babel.intraspaces or {}
Babel.intraspaces[\csname bbl@sbcp@\languagename\endcsname] = %
{b = #1, p = #2, m = #3}
Babel.locale_props[\the\localeid].intraspace = %
{b = #1, p = #2, m = #3}
}
\def\bbl@intrapenalty#1\@@{%
\directlua{
Babel = Babel or {}
Babel.intrapenalties = Babel.intrapenalties or {}
Babel.intrapenalties[\csname bbl@sbcp@\languagename\endcsname] = #1
Babel.locale_props[\the\localeid].intrapenalty = #1
}
}\begin{group}
\catcode`\%=12
\catcode`\^=14
\catcode`\'=12
\catcode`\_=12
\gdef\bbl@seaintraspace{\^}
\let\bbl@seaintraspace\relax
\directlua{
Babel = Babel or {}
Babel.sea_enabled = true
Babel.sea_ranges = Babel.sea_ranges or {}
function Babel.set_chranges (script, chrng)
local c = 0
for s, e in string.gmatch(chrng..',', '(.%)%.%(.-)%(.)%s') do
    Babel.sea_ranges[script..c] = {tonumber(s,16), tonumber(e,16)}
    c = c + 1
end
end
function Babel.sea_disc_to_space (head)
local sea_ranges = Babel.sea_ranges
local last_char = nil
local quad = 655360 /* 10 pt = 655360 = 10 * 65536 */
for item in node.traverse(head) do
    local i = item.id
    if i == node.id('glyph') then
        last_char = item
    elseif i == 7 and item.subtype == 3 and last_char
        and last_char.char > 0xC99 then
        quad = font.getfont(last_char.font).size
        for lg, rg in pairs(sea_ranges) do
            if last_char.char > rg[1] and last_char.char < rg[2] then
                lg = lg:sub(1, 4) /* Remove trailing number of, eg, Cyrillic */
                local intraspace = Babel.intraspaces[lg]
                local intrapenalty = Babel.intrapenalties[lg]
                local n
                if intrapenalty ~= 0 then
                    n = node.new(14, 0) /* penalty */
                    n.penalty = intrapenalty
                    node.insert_before(head, item, n)
                end
                n = node.new(12, 13) /* (glue, spaceskip) */
                node.setglue(n, intraspace.b * quad,
                              intraspace.p * quad,
                              intraspace.m * quad)
                node.insert_before(head, item, n)
                node.remove(head, item)
            end
        end
    end
end
end
}
\bbl@luahyphenate
\catcode`%=14
\gdef\bbl@cjintraspace{%
\let\bbl@cjintraspace\relax
\directlua{
Babel = Babel or {
require 'babel-data-cjk.lua'
Babel.cjk_enabled = true
function Babel.cjk_linebreak(head)
    local GLYPH = node.id('glyph')
    local last_char = nil
    local quad = 655360 /* 10 pt = 655360 = 10 * 65536 */
    local last_class = nil
    local last_lang = nil

    for item in node.traverse(head) do
        if item.id == GLYPH then
            local lang = item.lang
            local LOCALE = node.get_attribute(item,
                                               luatexbase.registernumber '\bbl@attr@locale')
            local props = Babel.locale_props[LOCALE]
            local class = Babel.cjk_class[item.char].c
            if class == 'cp' then class = 'cl' end % ] as CL

163
if class == 'id' then class = 'I' end

local br = 0
if class and last_class and Babel.cjk_breaks[last_class][class] then
  br = Babel.cjk_breaks[last_class][class]
end

if br == 1 and props.linebreak == 'c' and
  lang ~= \the\l@nohyphenation\space and
last_lang ~= \the\l@nohyphenation then
  local intrapenalty = props.intrapenalty
  if intrapenalty ~= 0 then
    local n = node.new(14, 0) % penalty
    n.penalty = intrapenalty
    node.insert_before(head, item, n)
  end
  local intraspace = props.intraspace
  local n = node.new(12, 13) % (glue, spaceskip)
  node.setglue(n, intraspace.b * quad,
              intraspace.p * quad,
              intraspace.m * quad)
  node.insert_before(head, item, n)
end

quad = font.getfont(item.font).size
last_class = class
last_lang = lang
else % if penalty, glue or anything else
  last_class = nil
end

lang.hyphenate(head)
end
%
\bbl@luahyphenate}
\gdef\bbl@luahyphenate{%
\let\bbl@luahyphenate\relax
\directlua{
  luatexbase.add_to_callback('hyphenate',
  function (head, tail)
    if Babel.linebreaking.before then
      for k, func in ipairs(Babel.linebreaking.before) do
        func(head)
      end
    end
    if Babel.cjk_enabled then
      Babel.cjk_linebreak(head)
    end
    lang.hyphenate(head)
    if Babel.linebreaking.after then
      for k, func in ipairs(Babel.linebreaking.after) do
        func(head)
      end
    end
    if Babel.sea_enabled then
      Babel.sea_disc_to_space(head)
    end, 'Babel.hyphenate')

15.5 CJK line breaking

Minimal line breaking for CJK scripts, mainly intended for simple documents and short texts as a secondary language. Only line breaking, with a little stretching for justification, without any attempt to adjust the spacing. It is based on (but does not strictly follow) the Unicode algorithm.

We first need a little table with the corresponding line breaking properties. A few characters have an additional key for the width (fullwidth vs. halfwidth), not yet used. There is a separate file, defined below.

Work in progress.

Common stuff.

```
165 \AddBabelHook{luatex}{loadkernel}{%
166 ⟨⟨ Restore Unicode catcodes before loading patterns ⟩⟩}
167 \ifx\DisableBabelHook\undefined\edef\DisableBabelHook{}
168 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
169 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ckeckstdfonts}
170 \DisableBabelHook{babel-fontspec}
171 \AddBabelHook{babel-fontspec}{beforerestart}{\bbl@ccheckstdfonts}
172 \AddBabelHook{babel-fontspec}{beforeflush}{\bbl@ckeckstdfonts}
173 \AddBabelHook{luatex}{loadkernel}{%
174 ⟨⟨ Restore Unicode catcodes before loading patterns ⟩⟩}
175 }%
15.6 Automatic fonts and ids switching

After defining the blocks for a number of scripts (must be extended and very likely fine tuned), we define a short function which just traverses the node list to carry out the replacements. The table `loc_to_scr` gets the locale form a script range (note the locale is the key, and that there is an intermediate table built on the fly for optimization). This locale is then used to get the `\language` and the `\localeid` as stored in `locale_props`, as well as the font (as requested). In the latter table a key starting with `/` maps the font from the global one (the key) to the local one (the value). Maths are skipped and discretionary are handled in a special way.

```directlua
Babel.script_blocks = {
    'Arab' = {{0x0600, 0x06FF}, {0x08A0, 0x08FF}, {0x0750, 0x077F},
               {0xFE70, 0xFEFF}, {0xFB50, 0xFDFF}, {0x1EE00, 0x1EEFF}},
    'Armn' = {{0x0530, 0x058F}},
    'Beng' = {{0x0980, 0x09FF}},
    'Copt' = {{0x03E2, 0x03EF}, {0x2C80, 0x2CFF}, {0x102E0, 0x102FF}},
    'Cyril' = {{0x0400, 0x04FF}, {0x0500, 0x052F}, {0x1C80, 0x1C8F},
                {0x2DE0, 0x2DFF}, {0xA640, 0xA69F}},
    'Deva' = {{0x0900, 0x097F}, {0xA8E0, 0xA8FF}},
    'Ethi' = {{0x1200, 0x137F}, {0x1380, 0x139F}, {0x2D80, 0x2DDF},
                {0x0AB00, 0x0AB2F}},
    'Geor' = {{0x0A00, 0x0FF}, {0x2D00, 0x2D2F}},
    'Grek' = {{0x0370, 0x03E1}, {0x03F0, 0x03FF}, {0x1F00, 0x1FFF}},
    'Hans' = {{0x2E80, 0x2EFF}, {0x3000, 0x303F}, {0x31C0, 0x31EF},
                 {0x3300, 0x33FF}, {0x3400, 0x4DBF}, {0x4E00, 0x9FFF},
                 {0xF900, 0xFAFF}, {0x20000, 0x2A6DF}, {0x2A700, 0x2B73F},
                 {0x2B740, 0x2B81F}, {0x2B820, 0x2CEAF},
                 {0x2C80, 0x2EBEF}, {0x2F800, 0x2FA1F}},
    'Hebr' = {{0x0900, 0x0FF},
    'Jpan' = {{0x3000, 0x303F}, {0x30A0, 0x30FF},
                 {0x0100, 0x017F}, {0x0E00, 0x0E7F}, {0xF900, 0xFAFF}},
    'Knda' = {{0x0D80, 0x0EFF},
    'Kore' = {{0x1100, 0x11FF}, {0x3130, 0x31BF},
                 {0x4E00, 0x9AFF}, {0xA960, 0xA97F}, {0x0C00, 0x0D7F},
                 {0x0F00, 0x0FFF}},
    'Laoo' = {{0x0E00, 0x0FF}},
    'Latn' = {{0x0000, 0x007F}, {0x0080, 0x00FF}, {0x0100, 0x017F},
                 {0x01E0, 0x01FF}, {0x2C60, 0x2C7F},
                 {0x4E00, 0x9AFF}, {0xA960, 0xA97F}, {0x0C00, 0x0D7F},
                 {0x0F00, 0x0FFF}},
    'Mahj' = {{0x11150, 0x1117F}},
    'Mlym' = {{0x0D00, 0x0D7F}},
    'Mymr' = {{0x1000, 0x109F}, {0xAA60, 0xAA7F}, {0xA960, 0xA97F}},
    'Orya' = {{0x0800, 0x0B7F}},
    'Sinh' = {{0x0D80, 0x0EFF},
    'Syr' = {{0x0700, 0x086F}, {0x0B7F},
                 {0x0F00, 0x0FFF}},
    'Tam' = {{0x0800, 0x0B7F}},
    'Telu' = {{0x0C00, 0x0C7F}},
    'Tfng' = {{0x2D30, 0x2D7F}},
    'Thai' = {{0x0E00, 0x0E7F}},
    'Tibt' = {{0x0F00, 0x0FFF}},
    'Vaii' = {{0x0A50, 0x0A63F}},
    'Yi' = {{0x0A00, 0x0A48F}, {0xA490, 0xA4CF}}
```
Babel.script_blocks.Cyrs = Babel.script_blocks.Cyril

function Babel.locale_map(head)
if not Babel.locale_mapped then return head end

local LOCALE = luatexbase.registernumber'bbl@attr@locale'
local GLYPH = node.id('glyph')
local inmath = false
local toloc_save
for item in node.traverse(head) do
  local toloc
  if not inmath and item.id == GLYPH then
    % Optimization: build a table with the chars found
    if Babel.chr_to_loc[item.char] then
      toloc = Babel.chr_to_loc[item.char]
    else
      for lc, maps in pairs(Babel.loc_to_scr) do
        for _, rg in pairs(maps) do
          if item.char >= rg[1] and item.char <= rg[2] then
            Babel.chr_to_loc[item.char] = lc
            toloc = lc
            break
          end
        end
      end
    end
  end
  elseif not inmath and item.id == 7 then
    item.replace = item.replace and Babel.locale_map(item.replace)
    item.pre = item.pre and Babel.locale_map(item.pre)
    item.post = item.post and Babel.locale_map(item.post)
  elseif item.id == node.id'math' then
    inmath = (item.subtype == 0)
  end
  end
  return head
end

if not inmath and item.id == GLYPH then
  local toloc
  % Optimization: build a table with the chars found
  if Babel.chr_to_loc[item.char] then
    toloc = Babel.chr_to_loc[item.char]
  else
    for lc, maps in pairs(Babel.loc_to_scr) do
      for _, rg in pairs(maps) do
        if item.char >= rg[1] and item.char <= rg[2] then
          Babel.chr_to_loc[item.char] = lc
          toloc = lc
          break
        end
      end
    end
  end
  % Now, take action, but treat composite chars in a different
  % fashion, because they 'inherit' the previous locale. Not yet
  % optimized.
  if not toloc and
    (item.char >= 0x0300 and item.char <= 0x036F) or
    (item.char >= 0x1AB0 and item.char <= 0x1AFF) or
    (item.char >= 0x1DC0 and item.char <= 0x1DFF) then
    toloc = toloc_save
  end
  if toloc and toloc > -1 then
    if Babel.locale_props[toloc].lg then
      item.lang = Babel.locale_props[toloc].lg
      node.set_attribute(item, LOCALE, toloc)
    end
    if Babel.locale_props[toloc]['/'.item.font] then
      item.font = Babel.locale_props[toloc]['/'.item.font]
    end
    toloc_save = toloc
  end
  elseif not inmath and item.id == 7 then
    item.replace = item.replace and Babel.locale_map(item.replace)
    item.pre = item.pre and Babel.locale_map(item.pre)
    item.post = item.post and Babel.locale_map(item.post)
  elseif item.id == node.id'math' then
    inmath = (item.subtype == 0)
  end
  end
  return head
end
The code for \babelcharproperty is straightforward. Just note the modified lua table can be different.

\newcommand\babelcharproperty[1]{%
\count@=#1\relax
\ifvmode
\expandafter\bbl@chprop
\else
\bbl@error{\string\babelcharproperty\space can be used only in vertical mode (preamble or between paragraphs)}%
\{}% See the manual for further info%
\fi}
\newcommand\bbl@chprop[3]{%
\@tempcnta=#1\relax
\bbl@ifunset{bbl@chprop@#2}{}
{\bbl@error{No property named '#2'. Allowed values are direction (bc), mirror (bmg), and linebreak (lb)}%}
{See the manual for further info}%
{}%
\loop
\bbl@cs{chprop@#2}{#3}
\ifnum\count@<\@tempcnta
\advance\count@\@ne
\repeat}
\def\bbl@chprop@direction#1{%
\directlua{
Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
Babel.characters[\the\count@]['d'] = '#1'
}}
\let\bbl@chprop@bc=\bbl@chprop@direction
\def\bbl@chprop@mirror#1{%
\directlua{
Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
Babel.characters[\the\count@][‘m’] = ‘\number#1’
}}
\let\bbl@chprop@bmg=\bbl@chprop@mirror
\def\bbl@chprop@linebreak#1{%
\directlua{
Babel.Babel.cjk_characters[\the\count@] = Babel.Babel.cjk_characters[\the\count@] or {}
Babel.Babel.cjk_characters[\the\count@][‘c’] = ‘#1’
}}
\let\bbl@chprop@lb=\bbl@chprop@linebreak
\def\bbl@chprop@locale#1{%
\directlua{
Babel.chr_to_loc = Babel.chr_to_loc or {}
Babel.chr_to_loc[\the\count@] = \bbl@ifblank{#1}{-1000}{\the\bbl@cs{id@@#1}}\space
}}%

Post-handling hyphenation patterns for non-standard rules, like ff to ff-f. There are still some issues with speed (not very slow, but still slow).

After declaring the table containing the patterns with their replacements, we define some auxiliary functions: str_to_nodes converts the string returned by a function to a node list, taking the node at base as a model (font, language, etc.); fetch_word fetches a series of glyphs and discretionaries, which pattern is matched against (if there is a match, it is called again before trying other patterns, and this is very likely the main bottleneck).

post_hyphenate_replace is the callback applied after tex.hyphenate. This means the automatic hyphenation points are known. As empty captures return a byte position (as
explained in the luatex manual), we must convert it to a utf8 position. With first, the last byte can be the leading byte in a utf8 sequence, so we just remove it and add 1 to the resulting length. With last we must take into account the capture position points to the next character. Here word_head points to the starting node of the text to be matched.

```lua
\begingroup
\catcode`\#=12
\catcode`\%=12
\catcode`\&=14
\directlua{
Babel.linebreaking.replacements = {}\n
function Babel.str_to_nodes(fn, matches, base)
local n, head, last
if fn == nil then return nil end
for s in string.utfvalues(fn(matches)) do
  if base.id == 7 then
    base = base.replace
  end
  n = node.copy(base)
  n.char = s
  if not head then
    head = n
  else
    last.next = n
  end
  last = n
end
return head
end

function Babel.fetch_word(head, funct)
local word_string = ''
local word_nodes = {}
local lang
local item = head
while item do
  if item.id == 29
    and not(item.char == 124) &% ie, not |
    and not(item.char == 61) &% ie, not =
    and (item.lang == lang or lang == nil) then
    lang = lang or item.lang
    word_string = word_string .. unicode.utf8.char(item.char)
    word_nodes[#word_nodes+1] = item
  elseif item.id == 7 and item.subtype == 2 then
    word_string = word_string .. '=
    word_nodes[#word_nodes+1] = item
  elseif item.id == 7 and item.subtype == 3 then
    word_string = word_string .. '|
    word_nodes[#word_nodes+1] = item
  elseif word_string == '' then
    &% pass
  else
end
```
function Babel.post_hyphenate_replace(head)
  local u = unicode.utf8
  local lbkr = Babel.linebreaking.replacements
  local word_head = head

  while true do
    local w, wn, nw, lang = Babel.fetch_word(word_head)
    if not lang then return head end
    if not lbkr[lang] then break end

    for k=1, #lbkr[lang] do
      local p = lbkr[lang][k].pattern
      local r = lbkr[lang][k].replace

      while true do
        local matches = { u.match(w, p) }
        if #matches < 2 then break end

        local first = table.remove(matches, 1)
        local last = table.remove(matches, #matches)

        &% Fix offsets, from bytes to unicode.
        first = u.len(w:sub(1, first-1)) + 1
        last = u.len(w:sub(1, last-1))

        local new &% used when inserting and removing nodes
        local changed = 0

        &% This loop traverses the replace list and takes the
        &% corresponding actions
        for q = first, last do
          local crep = r[q-first+1]
          local char_node = wn[q]
          local char_base = char_node

          if crep and crep.data then
            char_base = wn[crep.data+first-1]
          end

          if crep == {} then break end
          elseif crep == nil then
            changed = changed + 1
            node.remove(head, char_node)
          elseif crep and (crep.pre or crep.no or crep.post) then
            changed = changed + 1
            d = node.new(7, 0) &% (disc, discretionary)
            d.pre = Babel.str_to_nodes(crep.pre, matches, char_base)
            d.post = Babel.str_to_nodes(crep.post, matches, char_base)
          end
        end
      end
    end
  end
end
d.replace = Babel.str_to_nodes(crep.no, matches, char_base)
d.attr = char_base.attr
if crep.pre == nil then  &% TeXbook p96
d.penalty = crep.penalty or tex.hyphenpenalty
else
d.penalty = crep.penalty or tex.exhyphenpenalty
end
head, new = node.insert_before(head, char_node, d)
node.remove(head, char_node)
if q == 1 then
  word_head = new
end
elseif crep and crep.string then
  changed = changed + 1
  local str = crep.string(matches)
  if str == '' then
    if q == 1 then
      word_head = char_node.next
    end
    head, new = node.remove(head, char_node)
  elseif char_node.id == 29 and u.len(str) == 1 then
    char_node.char = string.utfvalue(str)
  else
    local n
    for s in string.utfvalues(str) do
      if char_node.id == 7 then
        log('Automatic hyphens cannot be replaced, just removed.')
      else
        n = node.copy(char_base)
      end
      n.char = s
      if q == 1 then
        head, new = node.insert_before(head, char_node, n)
      word_head = new
      else
        node.insert_before(head, char_node, n)
      end
    end
    node.remove(head, char_node)
  end % string length
  end % if char and char.string
end % for char in match
if changed > 20 then
texio.write('Too many changes. Ignoring the rest.')
elseif changed > 0 then
  w, wn, nw = Babel.fetch_word(word_head)
  end
end % for match
end % for patterns
word_head = nw
end % for words
return head
end % The following functions belong to the next macro
&% This table stores capture maps, numbered consecutively
```lua
Babel.capture_maps = {}

function Babel.capture_func(key, cap)
    local ret = "[[ " .. cap:gsub('{([0-9])}', "]..m[%1].." ]]"

    ret = ret:gsub('{([0-9])|([^|]+)|(.-)}', Babel.capture_func_map)
    ret = ret:gsub('%\[%\]%..\.', '')
    ret = ret:gsub('%..\[%\]%..\.', '')

    return key .. [[function(m) return ]] .. ret .. [[ end]]
end

function Babel.capt_map(from, mapno)
    return Babel.capture_maps[mapno][from] or from
end

&% Handle the {n|abc|ABC} syntax in captures
function Babel.capture_func_map(capno, from, to)
    local froms = {}
    for s in string.utfcharacters(from) do
        table.insert(froms, s)
    end
    local cnt = 1
    table.insert(Babel.capture_maps, {})
    local mlen = table.getn(Babel.capture_maps)
    for s in string.utfcharacters(to) do
        Babel.capture_maps[mlen][froms[cnt]] = s
        cnt = cnt + 1
    end
    return "]]..Babel.capt_map(m[" .. capno .. "]", ..
    (mlen) .. "])..' end
end
)}}

Now the \TeX{} high level interface, which requires the function defined above for converting strings to functions returning a string. These functions handle the \{n\} syntax. For example, \texttt{pre={1}{1}-} becomes \texttt{function(m) return m[1]..m[1]..'-' end}, where \texttt{m} are the matches returned after applying the pattern. With a mapped capture the functions are similar to \texttt{function(m) return Babel.capt_map(m[1], 1) end}, where the last argument identifies the mapping to be applied to \texttt{m[1]}. The way it is carried out is somewhat tricky, but the effect in not dissimilar to \texttt{lua} load – save the code as string in a \TeX{} macro, and expand this macro at the appropriate place. As \texttt{\directlua} does not take into account the current catcode of \texttt{\@}, we just avoid this character in macro names (which explains the internal group, too).
```
15.7 Layout

Work in progress.

Unlike xetex, luatex requires only minimal changes for right-to-left layouts, particularly in monolingual documents (the engine itself reverses boxes – including column order or headings –, margins, etc.) with bidi=basic, without having to patch almost any macro where text direction is relevant.

\@hangfrom is useful in many contexts and it is redefined always with the layout option. There are, however, a number of issues when the text direction is not the same as the box direction (as set by \bodydir), and when \parbox and \hangindent are involved. Fortunately, latest releases of luatex simplify a lot the solution with \shapemode.

With the issue #15 I realized commands are best patched, instead of redefined. With a few lines, a modification could be applied to several classes and packages. Now, tabular seems to work (at least in simple cases) with array, tabularx, hhline, colortbl, longtable, booktabs, etc. However, dcolumn still fails.
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, 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.

15.8 Auto bidi with basic and basic-r

The file babel-data-bidi.lua currently only contains data. It is a large and boring file and it is not shown here (see the generated file), but here is a sample:

```plaintext
[0x25]={d='et'},
[0x26]={d='on'},
[0x27]={d='on'},
[0x28]={d='on'}, m=0x29),
[0x29]={d='on'}, m=0x28),
```
For the meaning of these codes, see the Unicode standard.

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.

```plaintext
local d = node.new(DIR)

function Babel.bidi(head, ispar)
```

5034 ⟨∗basic-r⟩
5035 Babel = Babel or {}
5036
5037 Babel.bidi_enabled = true
5038
5039 require('babel-data-bidi.lua')
5040
5041 local characters = Babel.characters
5042 local ranges = Babel.ranges
5043
5044 local DIR = node.id("dir")
5045
5046 local function dir_mark(head, from, to, outer)
5047   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
5048   local d = node.new(DIR)
5049   d.dir = '+' .. dir
5050   node.insert_before(head, from, d)
5051   d = node.new(DIR)
5052   d.dir = '-' .. dir
5053   node.insert_after(head, to, d)
5054 end
5055
5056 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 inmath = 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'
    if inmath then dir = ('TRT' == tex.mathdir) and 'r' or 'l' end
  end
  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
  end
end

Next is based on the assumption babel sets the language AND switches the script with its dir. We treat a language block as a separate Unicode sequence. The following piece of code is executed at the first glyph after a 'dir' node. We don't know the current language until then. This is not exactly true, as the math mode may insert explicit dirs in the node list, so, for the moment there is a hack by brute force (just above).
if attr_dir == 1 then
  strong = 'r'
else if 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.

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' and not inmath then
  new_dir = true
  dir = nil
elseif item.id == node.id'math' then
  inmath = (item.subtype == 0)
extelse
  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, ie, a non-char), the textdir is set. This means you cannot insert, say, a whatsit, but this is what I would expect (with luacolor you may colorize some digits). Anyway, this behavior could be changed with a switch in the future. Note in the first branch only <an> is relevant if <al>.

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
else if dir == 'es' and last_n then  -- W3+W6
  last_es = item
else if dir == 'cs' then  -- it's right - do nothing
else 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
5148  dir_mark(head, first_n, last_n, 'r')
5149  dir_mark(head, first_d, last_d, outer)
5150  first_d, last_d = nil, nil
5151  elseif strong_lr == 'l' and type_n == '' then
5152     last_d = last_n
5153     end
5154  type_n = ''
5155  first_n, last_n = nil, nil
5156  end

R text in L, or L text in R. Order of dir_ mark's are relevant: d goes outside n, and 
therefore it's emitted after. See dir_ mark to understand why (but is the nesting actually 
necessary or is a flat dir structure enough?). Only L, R (and AL) chars are taken into 
account – everything else, including spaces, what's it, etc., are ignored:

5157  if dir == 'l' or dir == 'r' then
5158     if dir ~= outer then
5159         first_d = first_d or item
5160         last_d = item
5161     elseif first_d and dir ~= strong_lr then
5162         dir_mark(head, first_d, last_d, outer)
5163         first_d, last_d = nil, nil
5164     end
5165  end

5166  Mirroring. Each chunk of text in a certain language is considered a “closed” sequence. If 
<r on r> and <l on l>, it's clearly <r> and <l>, resp'tly, but with other combinations depends 
on outer. From all these, we select only those resolving <on> → <r>. At the beginning 
when last_lr is nil of an R text, they are mirrored directly.

5167  TODO - numbers in R mode are processed. It doesn't hurt, but should not be done.

5168  if dir and not last_lr and dir ~= 'l' and outer == 'r' then
5169     item.char = characters[item.char] and
5170     characters[item.char].m or item.char
5171     elseif (dir or new_dir) and last_lr ~= item then
5172     local mir = outer .. strong_lr .. (dir or outer)
5173     if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
5174         for ch in node.traverse(node.next(last_lr)) do
5175         if ch == item then break end
5176         if ch.id == node.id'glyph' and characters[ch.char] then
5177             ch.char = characters[ch.char].m or ch.char
5178         end
5179         end
5180     end

5181  Save some values for the next iteration. If the current node is ‘dir’, open a new sequence.

5182  if dir == 'l' or dir == 'r' then
5183     last_lr = item
5184     strong = dir_real -- Don’t search back - best save now
5185     strong_lr = (strong == 'l') and 'l' or 'r'
5186  elseif new_dir then
5187     last_lr = nil
5188     end
5189  end

5190  Mirror the last chars if they are no directed. And make sure any open block is closed, too.

5191  if last_lr and outer == 'r' then
5192     for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
if characters[ch.char] then
    ch.char = characters[ch.char].m or ch.char
end
end
end
end
if first_n then
    dir_mark(head, first_n, last_n, outer)
end
if first_d then
    dir_mark(head, first_d, last_d, outer)
end
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 {}

-- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
Babel.fontmap = Babel.fontmap or {}
Babel.fontmap[0] = {} -- l
Babel.fontmap[1] = {} -- r
Babel.bidi_enabled = true
Babel.mirroring_enabled = true
require('babel-data-bidi.lua')

local characters = Babel.characters
local ranges = Babel.ranges

local DIR = node.id('dir')
local GLYPH = node.id('glyph')

local function insert_implicit(head, state, outer)
    local new_state = state
    if state.sim and state.eim and state.sim ~= state.eim then
        dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
        local d = node.new(DIR)
        d.dir = '+' .. dir
        node.insert_before(head, state.sim, d)
        local d = node.new(DIR)
        d.dir = '-' .. dir
        node.insert_after(head, state.eim, d)
        new_state.sim, new_state.eim = nil, nil
    end
    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 = '+'. .. dir
        node.insert_before(head, state.san, d)
        local d = node.new(DIR)
        d.dir = '-'. .. dir
        node.insert_after(head, state.ean, d)
        new_state.san, new_state.ean = nil, nil
    end
    return head, new_state
end

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

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

180
local d = node.new(DIR)
 d.dir = '+TLT'
 _, new = node.insert_before(head, state.san, d)
 if state.san == state.sim then state.sim = new end
 local d = node.new(DIR)
 d.dir = '-TLT'
 _, new = node.insert_after(head, state.ean, d)
 if state.ean == state.eim then state.eim = new end
end

new_state.san, new_state.ean = nil, nil
return head, new_state
end

-- TODO - \hbox with an explicit dir can lead to wrong results
-- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>. A small attempt
-- was s made to improve the situation, but the problem is the 3-dir
-- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
-- well.

function Babel.bidi(head, ispar, hdir)
local d -- d is used mainly for computations in a loop
local prev_d = ''
local new_d = false
local nodes = {}
local outer_first = nil
local inmath = false
local glue_d = nil
local glue_i = nil
local has_en = false
local first_et = nil
local ATDIR = luatexbase.registernumber'bbl@attr@dir'
local save_outer
local temp = node.get_attribute(head, ATDIR)
if temp then
 temp = temp % 3
 save_outer = (temp == 0 and 'l') or
 (temp == 1 and 'r') or
 (temp == 2 and 'al')
elseif ispar then -- Or error? Shouldn't happen
 save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
else -- Or error? Shouldn't happen
 save_outer = ('TRT' == hdir) and 'r' or 'l'
end
-- when the callback is called, we are just _after_ the box,
-- and the textdir is that of the surrounding text
-- if not ispar and hdir == tex.textdir then
-- save_outer = ('TRT' == hdir) and 'r' or 'l'
-- end
local outer = save_outer
local temp = node.get_attribute(head, ATDIR)
if temp then
 temp = temp % 3
 save_outer = (temp == 0 and 'l') or
 (temp == 1 and 'r') or
 (temp == 2 and 'al')
elseif ispar then -- Or error? Shouldn't happen
 save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
else -- Or error? Shouldn't happen
 save_outer = ('TRT' == hdir) and 'r' or 'l'
end
-- '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
    end
    d = d or 'l'
  end
  local d_font = d_font or d
  d_font = (d_font == 'l' and 0) or
    (d_font == 'nsm' and 0) or
    (d_font == 'r' and 1) or
    (d_font == 'al' and 2) or
    (d_font == 'an' and 2) or nil
  if d_font and fontmap and fontmap[d_font][item_r.font] then
    item_r.font = fontmap[d_font][item_r.font]
  end
  end
  if new_d then
    table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
    if inmath then
      attr_d = 0
    else
      attr_d = node.get_attribute(item, ATDIR)
      attr_d = attr_d % 3
    end
    if attr_d == 1 then
      outer_first = 'r'
      last = 'r'
    elseif attr_d == 2 then
      outer_first = 'r'
      last = 'al'
    else
outer_first = 'l'
last = 'l'
end
outer = last
has_en = false
first_et = nil
new_d = false
end

if glue_d then
  if (d == 'l' and 'l' or 'r') ~= glue_d then
    table.insert(nodes, {glue_i, 'on', nil})
  end
  glue_d = nil
  glue_i = nil
end

elseif item.id == DIR then
  d = nil
  new_d = true
elseif item.id == node.id 'glue' and item.subtype == 13 then
  glue_d = d
  glue_i = item
  d = nil
else
  d = nil
end

-- AL <= EN/ET/ES -- W2 + W3 + W6
if last == 'al' and d == 'en' then
  d = 'an' -- W3
elseif last == 'al' and (d == 'et' or d == 'es') then
  d = 'on' -- W6
end

-- EN + CS/ES + EN -- W4
if d == 'en' and #nodes >= 2 then
  if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
    and nodes[#nodes-1][2] == 'en' then
    nodes[#nodes][2] = 'en'
  end
end

-- AN + CS + AN -- W4 too, because uax9 mixes both cases
if d == 'an' and #nodes >= 2 then
  if (nodes[#nodes][2] == 'cs')
    and nodes[#nodes-1][2] == 'an' then
    nodes[#nodes][2] = 'an'
  end
end

-- ET/EN -- W5 + W7->l / W6->on
if d == 'et' then
  first_et = first_et or (#nodes + 1)
elseif d == 'en' then
    has_en = true
    first_et = first_et or (#nodes + 1)
elseif first_et then -- d may be nil here !
    if has_en then
        if last == 'l' then
            temp = 'l' -- W7
        else
            temp = 'en' -- W5
        end
    else
        temp = 'on' -- W6
    end
    for e = first_et, #nodes do
        if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
    end
    first_et = nil
    has_en = false
end
if d then
    if d == 'al' then
        d = 'r'
        last = 'al'
    elseif d == 'l' or d == 'r' then
        last = d
    end
    prev_d = d
    table.insert(nodes, {item, d, outer_first})
end
outer_first = nil
end
-- TODO -- repeated here in case EN/ET is the last node. Find a
-- better way of doing things:
if first_et then -- dir may be nil here !
    if has_en then
        if last == 'l' then
            temp = 'l' -- W7
        else
            temp = 'en' -- W5
        end
    else
        temp = 'on' -- W6
    end
    for e = first_et, #nodes do
        if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
    end
    end
-- dummy node, to close things
    table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
------------------------ NEUTRAL ----------------------
outer = save_outer
last = outer
local first_on = nil

for q = 1, #nodes do
  local item
  local outer_first = nodes[q][3]
  outer = outer_first or outer
  last = outer_first or last
  local d = nodes[q][2]
  if d == 'an' or d == 'en' then d = 'r' end
  if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end
  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 Babel.mirroring_enabled and item.id == GLYPH
        and temp == 'r' and characters[item.char] then
        local font_mode = font.fonts[item.font].properties.mode
        if font_mode ~= 'harf' and font_mode ~= 'plug' then
          item.char = characters[item.char].m or item.char
        end
      end
    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
  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
    else
        if d == 'an' or d == 'l' then
            if nodes[q][3] then -- nil except after an explicit dir
                state.sim = item -- so we move sim 'inside' the group
            else
                state.sim = state.sim or item
            end
            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
elseif d == 'on' and state.san then
    state.san = state.san or item
    state.ean = item
end

return node.prev(head) or head
end

16 Data for CJK

It is a boring file and it is not shown here (see the generated file), but here is a sample:

```
[0x0021]={c='ex'},
[0x0024]={c='pr'},
[0x0025]={c='po'},
[0x0028]={c='op'},
[0x0029]={c='cp'},
[0x002B]={c='pr'},
```

For the meaning of these codes, see the Unicode standard.
17 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.

\ProvideLanguage{nil}{\langle\langle date\rangle\rangle \langle\langle version\rangle\rangle 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@nil\@undefined
\newlanguage\l@nil
\namedef{bbl@hyphendata@\the\l@nil}{{}{}}% Remove warning
\let\bbl@elt\relax
\edef\bbl@languages{% Add it to the list of languages
\bbl@languages\bbl@elt{nil}{\the\l@nil}{}{}}
\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.

\captionnil
\datenil
\LdfInit{nil}{\l@nil\@undefined}
\NewLanguage{\l@nil}
\namedef{bbl@hyphendata@\the\l@nil}{{}{}}% Remove warning
\let\bbl@elt\relax
\edef\bbl@languages{% Add it to the list of languages
\bbl@languages\bbl@elt{nil}{\the\l@nil}{}{}}
\fi

The macro \LdfFinish 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.

\LdfFinish{nil}
\endinput

18 Support for Plain \TeX\ (plain.def)

18.1 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 localhyphen.tex or whatever they like, but they mustn’t fiddle 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 \input, 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 \input sees, we need to set some category codes just to be able to change the definition of \input.
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...\openin0 hyphen.cfg

If the file wasn't found the following test turns out true.\ifeof0\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\else\fi

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

\a plain.tex\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.

\def\fmtname{babel-plain}\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.

18.2 Emulating some \LaTeX features

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

\def\@empty{}
\def\loadlocalcfg#1{\openin0#1.cfg\ifeof0\closein0\else\closein0\{\immediate\write16{*************************************}%\immediate\write16{* Local config file #1.cfg used}%\fi}


18.3 General tools

A number of \LaTeX macro's that are needed later on.

\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\relax
\def\@car#1#2\@nil{#1}
\def\@cdr#1#2\@nil{#2}
\let\@typeset@protect\relax
\let\protected@edef\edef
\long\def\@gobble#1{ }
\edef\@backslashchar{\expandafter\@gobble\string\}\
\def\g@addto@macro#1#2{{\toks\expandafter{#1#2}\xdef#1{\the\toks}}}
\def\@namedef#1{\expandafter\def\csname #1\endcsname}
\def\@nameuse#1{\csname #1\endcsname}
\def\@ifundefined#1{\expandafter\ifx\csname#1\endcsname\relax\expandafter\@firstoftwo\else\expandafter\@secondoftwo\fi}
\def\@expandtwoargs#1#2#3{\edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
\def\zap@space#1 #2{#1\ifx#2\@empty\else\expandafter\zap@space\fi #2}

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

\def\begindocument{\if\@preamblecmds\@undefined\def\@preamblecmds{}\fi\def\@onlypreamble#1{\expandafter\gdef\expandafter\@preamblecmds\expandafter{\@preamblecmds\do#1}}
\@onlypreamble\@onlypreamble

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

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

\begin{verbatim}
\def\AtEndOfPackage#1{\g@addto@macro\@endofldf{#1}}
\end{verbatim}

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

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

Mimick \LaTeX{}'s commands to define control sequences.

\begin{verbatim}
\def\newcommand{\@star@or@long\new@command}
\def\new@command#1{\@testopt\@newcommand#10}
\def\@newcommand#1[#2]{\@ifnextchar [{{\@xargdef#1[#2]}}%\@argdef#1[#2]}}
\long\def\@argdef#1[#2]#3{\@yargdef#1@ne{#2}{#3}}
\long\def\@xargdef#1[#2][#3]#4{\expandafter\def\expandafter#1\expandafter{#3}
\expandafter\@yargdef \csname \string#1\endcsname \tw@{#2}{#4}}
\def\providecommand{\@star@or@long\provide@command}
\def\provide@command#1{\begingroup\escapechar\m@ne\xdef\@gtempa{{\string#1}}%
\end{verbatim}
The following little macro \texttt{in@} is taken from \texttt{latex.ltx}; it checks whether its first argument is part of its second argument. It uses the boolean \texttt{in@}; allocating a new boolean inside conditionally executed code is not possible, hence the construct with the temporary definition of \texttt{bbl@tempa}.

\begin{verbatim}
\def\bbl@tempa{\csname newif\endcsname\ifin@}
\ifx\in@\@undefined
  \def\in@#1#2{\def\in@@##1#1##2##3\in@@{\ifx\in@##2\in@false\else\in@true\fi}#2#1\in@\in@@}
\else
  \let\bbl@tempa\@empty
\fi
\bbl@tempa
\end{verbatim}

\LaTeX{} has a macro to check whether a certain package was loaded with specific options. The command has two extra arguments which are code to be executed in either the true or false case. This is used to detect whether the document needs one of the accents to be activated (activegrave and activeacute). For plain \LaTeX{} we assume that the user wants them to be active by default. Therefore the only thing we do is execute the third argument (the code for the true case).

\begin{verbatim}
\def\@ifpackagewith#1#2#3#4(#3)

The \LaTeX{} macro \texttt{@ifl@aded} checks whether a file was loaded. This functionality is not needed for plain \LaTeX{} but we need the macro to be defined as a no-op.
For the following code we need to make sure that the commands \newcommand and \providecommand exist with some sensible definition. They are not fully equivalent to their \LaTeX{} versions; just enough to make things work in plain \TeX{} environments.

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

18.4 Encoding related macros

Code from \texttt{ltoutenc.dtx}, adapted for use in the plain \TeX{} environment.
Currently we only use the \LATEX2ε method for accents for those that are known to be made active in some language definition file.

\begin{verbatim}
\DeclareTextAccent{"}{OT1}{127}
\DeclareTextAccent{'}{OT1}{19}
\DeclareTextAccent{^}{OT1}{94}
\DeclareTextAccent{`}{OT1}{18}
\DeclareTextAccent{~}{OT1}{126}
\end{verbatim}

The following control sequences are used in babel.def but are not defined for \texttt{PLAIN} \LaTeX.

\begin{verbatim}
\DeclareTextSymbol{\textquotedblleft}{OT1}{92}
\DeclareTextSymbol{\textquotedblright}{OT1}{`"}
\DeclareTextSymbol{\textquoteleft}{OT1}{`\}`
\DeclareTextSymbol{\textquoteright}{OT1}{`\'}
\DeclareTextSymbol{i}{OT1}{16}
\DeclareTextSymbol{ss}{OT1}{25}
\end{verbatim}

For a couple of languages we need the \LaTeX-control sequence \texttt{\scriptsize} to be available. Because \texttt{PLAIN} \LaTeX doesn't have such a sophisticated font mechanism as \LaTeX has, we just \texttt{\let} it to \texttt{\sevenrm}.

\begin{verbatim}
\ifx\scriptsize\@undefined
  \let\scriptsize\sevenrm
\fi
\end{verbatim}

\section{Acknowledgements}

I would like to thank all who volunteered as β-testers for their time. Michel Goossens supplied contributions for most of the other languages. Nico Poppelier helped polish the text of the documentation and supplied parts of the macros for the Dutch language. Paul Wackers and Werenfried Spit helped find and repair bugs. During the further development of the babel system I received much help from Bernd Raichle, for which I am grateful.

\section*{References}


