Babel

Localization and internationalization

\TeX
pdf\TeX
\Lua\TeX
\LuaHB\TeX
\Xe\TeX
# Contents

## User guide

1. **The user interface**
   - 1.1 Monolingual documents
   - 1.2 Multilingual documents
   - 1.3 Mostly monolingual documents
   - 1.4 Modifiers
   - 1.5 Troubleshooting
   - 1.6 Plain
   - 1.7 Basic language selectors
   - 1.8 Auxiliary language selectors
   - 1.9 More on selection
   - 1.10 Shorthands
   - 1.11 Package options
   - 1.12 The base option
   - 1.13 ini files
   - 1.14 Selecting fonts
   - 1.15 Modifying a language
   - 1.16 Creating a language
   - 1.17 Digits
   - 1.18 Accessing language info
   - 1.19 Hyphenation and line breaking
   - 1.20 Selecting scripts
   - 1.21 Selecting directions
   - 1.22 Language attributes
   - 1.23 Hooks
   - 1.24 Languages supported by babel with ldf files
   - 1.25 Unicode character properties in lualtx
   - 1.26 Tweaking some features
   - 1.27 Tips, workarounds, known issues and notes
   - 1.28 Current and future work
   - 1.29 Tentative and experimental code

2. **Loading languages with language.dat**
   - 2.1 Format

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

4. **Changes**
   - 4.1 Changes in babel version 3.9

## Source code

5. **Identification and loading of required files**
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 locale directory</td>
<td>55</td>
</tr>
<tr>
<td>7 Tools</td>
<td>56</td>
</tr>
<tr>
<td>7.1 Multiple languages</td>
<td>60</td>
</tr>
<tr>
<td>8 The Package File (\LaTeX, babel.sty)</td>
<td>61</td>
</tr>
<tr>
<td>8.1 base</td>
<td>61</td>
</tr>
<tr>
<td>8.2 key=value options and other general option</td>
<td>63</td>
</tr>
<tr>
<td>8.3 Conditional loading of shorthands</td>
<td>65</td>
</tr>
<tr>
<td>8.4 Language options</td>
<td>66</td>
</tr>
<tr>
<td>9 The kernel of Babel (babel.def, common)</td>
<td>69</td>
</tr>
<tr>
<td>9.1 Tools</td>
<td>69</td>
</tr>
<tr>
<td>9.2 Hooks</td>
<td>71</td>
</tr>
<tr>
<td>9.3 Setting up language files</td>
<td>73</td>
</tr>
<tr>
<td>9.4 Shorthands</td>
<td>75</td>
</tr>
<tr>
<td>9.5 Language attributes</td>
<td>85</td>
</tr>
<tr>
<td>9.6 Support for saving macro definitions</td>
<td>87</td>
</tr>
<tr>
<td>9.7 Short tags</td>
<td>88</td>
</tr>
<tr>
<td>9.8 Hyphens</td>
<td>88</td>
</tr>
<tr>
<td>9.9 Multiencoding strings</td>
<td>90</td>
</tr>
<tr>
<td>9.10 Macros common to a number of languages</td>
<td>96</td>
</tr>
<tr>
<td>9.11 Making glyphs available</td>
<td>96</td>
</tr>
<tr>
<td>9.11.1 Quotation marks</td>
<td>96</td>
</tr>
<tr>
<td>9.11.2 Letters</td>
<td>97</td>
</tr>
<tr>
<td>9.11.3 Shorthands for quotation marks</td>
<td>98</td>
</tr>
<tr>
<td>9.11.4 Umlauts and tremas</td>
<td>99</td>
</tr>
<tr>
<td>9.12 Layout</td>
<td>100</td>
</tr>
<tr>
<td>9.13 Load engine specific macros</td>
<td>101</td>
</tr>
<tr>
<td>9.14 Creating languages</td>
<td>101</td>
</tr>
<tr>
<td>10 Adjusting the Babel behavior</td>
<td>113</td>
</tr>
<tr>
<td>11 The kernel of Babel (babel.def for \LaTeX{} only)</td>
<td>114</td>
</tr>
<tr>
<td>11.1 The redefinition of the style commands</td>
<td>114</td>
</tr>
<tr>
<td>11.2 Cross referencing macros</td>
<td>115</td>
</tr>
<tr>
<td>11.3 Marks</td>
<td>118</td>
</tr>
<tr>
<td>11.4 Preventing clashes with other packages</td>
<td>119</td>
</tr>
<tr>
<td>11.4.1 ifthen</td>
<td>119</td>
</tr>
<tr>
<td>11.4.2 varioref</td>
<td>120</td>
</tr>
<tr>
<td>11.4.3 hline</td>
<td>120</td>
</tr>
<tr>
<td>11.4.4 hyperref</td>
<td>121</td>
</tr>
<tr>
<td>11.4.5 fancyhdr</td>
<td>121</td>
</tr>
<tr>
<td>11.5 Encoding and fonts</td>
<td>121</td>
</tr>
<tr>
<td>11.6 Basic bidi support</td>
<td>123</td>
</tr>
<tr>
<td>11.7 Local Language Configuration</td>
<td>126</td>
</tr>
<tr>
<td>12 Multiple languages (switch.def)</td>
<td>127</td>
</tr>
<tr>
<td>12.1 Selecting the language</td>
<td>128</td>
</tr>
<tr>
<td>12.2 Errors</td>
<td>137</td>
</tr>
<tr>
<td>13 Loading hyphenation patterns</td>
<td>139</td>
</tr>
<tr>
<td>14 Font handling with fontspec</td>
<td>143</td>
</tr>
</tbody>
</table>
15 Hooks for XeTeX and LuaTeX

15.1 XeTeX ......................................................... 148
15.2 Layout ....................................................... 151
15.3 LuaTeX ...................................................... 152
15.4 Southeast Asian scripts ................................. 158
15.5 CJK line breaking .......................................... 161
15.6 Automatic fonts and its switching .................... 162
15.7 Layout ....................................................... 169
15.8 Auto bidi with basic and basic-r ........................ 171

16 Data for CJK .................................................. 182

17 The ‘nil’ language ........................................... 182

18 Support for Plain TeX (plain.def) ....................... 183

18.1 Not renaming hyphen.tex .................................. 183
18.2 Emulating some MiX features ............................ 184
18.3 General tools ............................................... 184
18.4 Encoding related macros ................................. 188

19 Acknowledgements ........................................... 191

Troubleshooting

Paragraph ended before \UTFviii@three@octets was complete .............. 5
No hyphenation patterns were preloaded for (babel) the language ‘LANG’ into the format .......................... 5
You are loading directly a language style ....................... 8
Unknown language ‘LANG’ ..................................... 8
Argument of \language@active@arg has an extra } .................. 12
Package babel Info: The following fonts are not babel standard families ........................... 27
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 kadingira mail 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 1df files). The alternative way based on 1in1 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).

\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:

\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:
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 (eg, 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[ngerman,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{} -- \alsiname{} -- \today
\selectlanguage{vietnamese}
\prefacename{} -- \alsiname{} -- \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):

```
\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:

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

- Another typical error when using babel is the following:

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

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

1.6 Plain

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

1. No predefined “axis” for modifiers are provided because languages and their scripts have quite different needs.
2. In old versions the error read “You have used an old interface to call babel”, not very helpful.
3. In old versions the error read “You haven't loaded the language LANG yet.”
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
\begin{otherlanguage}{language} ... \end{otherlanguage}

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

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

\begin{group}
\selectlanguage{inner-language}
\selectlanguage{outer-language}
\end{group}

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

Spaces after the environment are ignored.

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

Same as \foreignlanguage{} but as environment. Spaces after the environment are not ignored.

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

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

The environment \hyphenrules{} can be used to select only the hyphenation rules to be used (it can be used as command, too). This can for instance be used to select ‘nohyphenation’, provided that in \texttt{language.dat} the ‘language’ nohyphenation is defined by loading \texttt{zerohyph.tex}. It deactivates language shorthands, too (but not user shorthands).

Except for these simple uses, \hyphenrules{} is discouraged and \otherlanguage{} (the starred version) is preferred, as the former does not take into account possible changes in encodings of characters like, say, ‘ done by some languages (eg, italian, french, ukraineb).

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

1.9 More on selection

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

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

It defines \texttt{\text{\{tag1\}}\{\text{\{text\}\}} to be \foreignlanguage{}\{\text{\{language1\}\}}\{\text{\{text\}\}}, and \texttt{\begin{otherlanguage*}{\{tag1\}}} to be \begin{otherlanguage*}{}\{\text{\{language1\}\}}, and so on. Note \texttt{\{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.}
\textde{German text}

and

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

\text

\begin{finnish}
  Finnish text
\end{finnish}

\text

NOTE Something like \text{\foreignlanguage{french}{French text}} is legitimate – it defines \text{\french} and \french (and, of course, \begin{french}).

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

\text

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

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

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

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

\babelensure{polish}

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

\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).

5With it, encoded strings may not work as expected.
1.10 Shorthands

A shorthand is a sequence of one or two characters that expands to arbitrary \TeX code. Shorthands can be used for different kinds of things, for example: (1) in some languages shorthands such as "a are defined to be able to hyphenate the word if the encoding is OT1; (2) in some languages shorthands such as \ are used to insert the right amount of white space; (3) several kinds of discretionary and breaks can be inserted easily with "-, "=, etc. The package inputenc as well as xetex and luatex have alleviated entering non-ASCII characters, but minority languages and some kinds of text can still require characters not directly available on the keyboards (and sometimes not even as separated or precomposed Unicode characters). As to the point 2, now pdfTeX provides \nbccode, 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, "{\}).

\shorthandoff \shorthandon
\shorthandon \{(shorthands-list)\}
\shorthandoff \{(shorthands-list)\}

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

**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\*\{(char)\} initiates the definition of user-defined shorthand sequences. It has one argument, the character that starts these personal shorthands.

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

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

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

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

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

```
\useshorthands\*\{
\defineshorthand\*\{\{\}\}\{\babelhyphen\{soft\}\}
\defineshorthand\{-\}\{\babelhyphen\{hard\}\}
```

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

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

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

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

The command \languageshorthands\{\{language\}\} 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).\textsuperscript{6} Note that for this to work the language should have been specified as an option when loading the babel package. For example, you can use in english the shorthands defined by ngerman with

\textsuperscript{6} Actually, any name not corresponding to a language group does the same as none. However, follow this convention because it might be enforced in future releases of babel to catch possible errors.
\addto\extrasenglish{\languageshorthands{ngerman}}

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

**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

\(\langle\text{shorthand}\rangle\)

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

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

\title{Documento científico\babelshorthand{\textasciitilde}t́cnico}

For your records, here is a list of shorthands, but you must double check them, as they may change\footnote{Thanks to Enrico Gregorio}

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

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

**Basque**  " ‘ ~

**Breton**  : ; ? !

**Catalan**  " ‘

**Czech**  " -

**Esperanto**  ^

**Estonian**  " ~

**French**  (all varieties) : ; ? !

**Galician**  . ’ ~ < >

**Greek**  ~

**Hungarian**  :

**Kurmanji**  ^

**Latin**  " ^ =

**Slovak**  " ^ ’ ~

**Spanish**  . < > ' ~

**Turkish**  : ! =

In addition, the babel core declares ~ as a one-char shorthand which is let, like the standard ~, to a non breaking space\footnote{This declaration serves to nothing, but it is preserved for backward compatibility.}
\ifbeshorthand \{character\}\{true\}\{false\}

\textbf{New 3.23} Tests if a character has been made a shorthand.

\aliashorthand \{\{original\}\{\{alias\}\}

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.

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

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

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

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

1.11 Package options

\textbf{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.

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

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

\textbf{activegrave} Same for `. 

\textbf{shorthands=} \langle char\rangle\langle 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.
safe= none | ref | bib
Some \LaTeX macros are redefined so that using shorthands is safe. With safe=bib only \nocite, \bibcite and \bibitem are redefined. With safe=ref only \newlabel, \ref and \pageref are redefined (as well as a few macros from varioref and ifthen). With safe=none no macro is redefined. This option is strongly recommended, because a good deal of incompatibilities and errors are related to these redefinitions. As of 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).

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

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

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

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

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

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

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

silent New 3.9I No warnings and no \emph {infos} are written to the log file.\footnote {You can use alternatively the package silence. }

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

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

\footnotetext {You can use alternatively the package silence.}
New 3.9g  Sets the behavior of case mapping for hyphenation, provided the language defines it. It can take the following values:

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>off</td>
<td>Deactivates this feature and no case mapping is applied;</td>
</tr>
<tr>
<td>first</td>
<td>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.</td>
</tr>
<tr>
<td>select</td>
<td>Sets it only at \selectlanguage;</td>
</tr>
<tr>
<td>other</td>
<td>Also sets it at otherlanguage;</td>
</tr>
<tr>
<td>other*</td>
<td>Also sets it at otherlanguage* as well as in heads and foots (if the option headfoot is used) and in auxiliary files (ie, at \select@language), and it's the default if several language options have been stated. The option first can be regarded as an optimized version of other* for monolingual documents.</td>
</tr>
</tbody>
</table>

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  \textsf{\textbackslash{option-name}}\{\textsf{\textbackslash{code}}\}

This command is currently the only provided by base. Executes \textsf{\textbackslash{code}} when the file loaded by the corresponding package option is finished (at \textsf{\textbackslash{ldef\atfinish}}). The setting is global. So

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

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

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

10Turned off in plain.
11Duplicated options count as several ones.
12Providing foreign is pointless, because the case mapping applied is that at the end of the paragraph, but if either xetex or luatex change this behavior it might be added. On the other hand, other is provided even if I [JBL] think it isn't really useful, but who knows.
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 \...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 \texttt{\textbackslash babelprovide}), but a higher interface, based on package options, in under study. In other words, \texttt{\textbackslash babelprovide} is mainly meant for auxiliary tasks.

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

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

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

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

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

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

\newfontscript{Devanagari}{deva}

Other Indic scripts are still under development in luatex. On the other hand, xetex is better. The upcoming lualatex will be based on luahbtex, so Indic scripts will be rendered correctly with the option Renderert=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:

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

**East Asia scripts**  Settings for either Simplified of 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 shorts 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:

\begin{verbatim}
\documentclass{ltjbook}
\usepackage[japanese]{babel}
\end{verbatim}

**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>af</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-Cyril</td>
<td>Bosnian</td>
</tr>
<tr>
<td>as</td>
<td>Assamese</td>
<td>bs-Latn</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>cs</td>
<td>Czechul</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>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 Chiini</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>kjk</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>Kashmiri</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>Makuwa-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>ml</td>
<td>Malayalam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mn</td>
<td>Mongolian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mr</td>
<td>Marathi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ms-BN</td>
<td>Malay</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>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-Latn-BA</td>
<td>Serbian</td>
</tr>
<tr>
<td>ne</td>
<td>Nepali</td>
<td>sr-Latn-ME</td>
<td>Serbian</td>
</tr>
<tr>
<td>nl</td>
<td>Dutch</td>
<td>sr-Latn-XK</td>
<td>Serbian</td>
</tr>
<tr>
<td>nmg</td>
<td>Kwasio</td>
<td>sr-Latn</td>
<td>Serbian</td>
</tr>
<tr>
<td>nn</td>
<td>Norwegian Nynorsk</td>
<td>sr</td>
<td>Serbian</td>
</tr>
<tr>
<td>nnh</td>
<td>Ngiemboon</td>
<td>sv</td>
<td>Swedish</td>
</tr>
<tr>
<td>nus</td>
<td>Nuer</td>
<td>sw</td>
<td>Swahili</td>
</tr>
<tr>
<td>nyn</td>
<td>Nyankole</td>
<td>ta</td>
<td>Tamil</td>
</tr>
<tr>
<td>om</td>
<td>Oromo</td>
<td>te</td>
<td>Telugu</td>
</tr>
<tr>
<td>or</td>
<td>Odia</td>
<td>teo</td>
<td>Teso</td>
</tr>
<tr>
<td>os</td>
<td>Ossetic</td>
<td>th</td>
<td>Thai</td>
</tr>
<tr>
<td>pa-Arab</td>
<td>Punjabi</td>
<td>ti</td>
<td>Tigrinya</td>
</tr>
<tr>
<td>pa-Guru</td>
<td>Punjabi</td>
<td>tk</td>
<td>Turkmen</td>
</tr>
<tr>
<td>pa</td>
<td>Punjabi</td>
<td>to</td>
<td>Tongan</td>
</tr>
<tr>
<td>pl</td>
<td>Polish</td>
<td>tr</td>
<td>Turkish</td>
</tr>
<tr>
<td>pms</td>
<td>Piedmontese</td>
<td>twq</td>
<td>Tasawaq</td>
</tr>
<tr>
<td>ps</td>
<td>Pashto</td>
<td>tzm</td>
<td>Central Atlas Tamazight</td>
</tr>
<tr>
<td>pt-BR</td>
<td>Portuguese</td>
<td>ug</td>
<td>Uyghur</td>
</tr>
<tr>
<td>pt-PT</td>
<td>Portuguese</td>
<td>uk</td>
<td>Ukrainian</td>
</tr>
<tr>
<td>pt</td>
<td>Portuguese</td>
<td>ur</td>
<td>Urdu</td>
</tr>
<tr>
<td>qu</td>
<td>Quechua</td>
<td>uz-Arab</td>
<td>Uzbek</td>
</tr>
<tr>
<td>rm</td>
<td>Romangsh</td>
<td>uz-Cyrl</td>
<td>Uzbek</td>
</tr>
<tr>
<td>rn</td>
<td>Rundi</td>
<td>uz-Latn</td>
<td>Uzbek</td>
</tr>
<tr>
<td>ro</td>
<td>Romanian</td>
<td>uz</td>
<td>Uzbek</td>
</tr>
<tr>
<td>rof</td>
<td>Rombo</td>
<td>ru</td>
<td>Russian</td>
</tr>
<tr>
<td>ru</td>
<td>Russian</td>
<td>rwa</td>
<td>Vai</td>
</tr>
<tr>
<td>rwk</td>
<td>Rwa</td>
<td>rwb</td>
<td>Vai</td>
</tr>
<tr>
<td>sa-Beng</td>
<td>Sanskrit</td>
<td>sai</td>
<td>Vai</td>
</tr>
<tr>
<td>sa-Dev</td>
<td>Sanskrit</td>
<td>sai</td>
<td>Vai</td>
</tr>
<tr>
<td>sa-Gujr</td>
<td>Sanskrit</td>
<td>sah</td>
<td>Vai</td>
</tr>
<tr>
<td>sa-Knda</td>
<td>Sanskrit</td>
<td>saq</td>
<td>Cantonese</td>
</tr>
<tr>
<td>sa-Mlym</td>
<td>Sanskrit</td>
<td>sbp</td>
<td>Tamazight</td>
</tr>
<tr>
<td>sa-Telu</td>
<td>Sanskrit</td>
<td>se</td>
<td>Chinese</td>
</tr>
<tr>
<td>se</td>
<td>Northern Sami</td>
<td>seh</td>
<td>Chinese</td>
</tr>
<tr>
<td>ses</td>
<td>Koyraboro Senni</td>
<td>ses</td>
<td>Chinese</td>
</tr>
<tr>
<td>sg</td>
<td>Sango</td>
<td>shi</td>
<td>Chinese</td>
</tr>
<tr>
<td>shi-Latn</td>
<td>Tachelhit</td>
<td>shi</td>
<td>Chinese</td>
</tr>
<tr>
<td>shi-Tfng</td>
<td>Tachelhit</td>
<td>shi</td>
<td>Chinese</td>
</tr>
<tr>
<td>si</td>
<td>Sinhala</td>
<td>si</td>
<td>Chinese</td>
</tr>
<tr>
<td>sk</td>
<td>Slovak</td>
<td>si</td>
<td>Chinese</td>
</tr>
<tr>
<td>sl</td>
<td>Slovenian</td>
<td>su</td>
<td>Chinese</td>
</tr>
<tr>
<td>sv</td>
<td>Swedish</td>
<td>svu</td>
<td>Chinese</td>
</tr>
<tr>
<td>twq</td>
<td>Tasawaq</td>
<td>twq</td>
<td>Chinese</td>
</tr>
<tr>
<td>u</td>
<td>Vai</td>
<td>uvi</td>
<td>Chinese</td>
</tr>
<tr>
<td>vai</td>
<td>Vai</td>
<td>vai</td>
<td>Chinese</td>
</tr>
<tr>
<td>vae</td>
<td>Walser</td>
<td>vay</td>
<td>Chinese</td>
</tr>
<tr>
<td>vi</td>
<td>Viëtnamese</td>
<td>vi</td>
<td>Chinese</td>
</tr>
<tr>
<td>vun</td>
<td>Vunjo</td>
<td>vun</td>
<td>Chinese</td>
</tr>
<tr>
<td>vun</td>
<td>Vunjo</td>
<td>wae</td>
<td>Chinese</td>
</tr>
<tr>
<td>wae</td>
<td>Walser</td>
<td>xog</td>
<td>Chinese</td>
</tr>
<tr>
<td>yav</td>
<td>Yangben</td>
<td>yav</td>
<td>Chinese</td>
</tr>
<tr>
<td>yi</td>
<td>Yiddish</td>
<td>yi</td>
<td>Chinese</td>
</tr>
<tr>
<td>yo</td>
<td>Yoruba</td>
<td>yo</td>
<td>Chinese</td>
</tr>
<tr>
<td>yue</td>
<td>Cantonese</td>
<td>yue</td>
<td>Chinese</td>
</tr>
<tr>
<td>zgh</td>
<td>Standard Moroccan Tamazight</td>
<td>zgh</td>
<td>Chinese</td>
</tr>
<tr>
<td>zh-Hant-HK</td>
<td>Chinese</td>
<td>zh-Hant-MO</td>
<td>Chinese</td>
</tr>
<tr>
<td>zh-Hant</td>
<td>Chinese</td>
<td>zh-Hant</td>
<td>Chinese</td>
</tr>
<tr>
<td>zh</td>
<td>Chinese</td>
<td>zh</td>
<td>Chinese</td>
</tr>
<tr>
<td>zu</td>
<td>Zulu</td>
<td>zu</td>
<td>Chinese</td>
</tr>
</tbody>
</table>
In some contexts (currently \texttt{\textbackslash babel} font) an ini file may be loaded by its name. Here is the list of the names currently supported. With these languages, \texttt{\textbackslash babel} 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 \texttt{\textbackslash babelprovide} with a valueless import.

+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+-----------------------+
| aghem                 | centralatlastamazight |
| akan                  | centralkurdish        |
| albanian              | chechen               |
| american              | cherokee              |
| amharic               | chiga                 |
| arabic                | chinese-hans-hk       |
| arabic-algeria        | chinese-hans-mo       |
| arabic-DZ             | chinese-hans-sg       |
| arabic-morocco        | chinese-hans          |
| arabic-MA             | chinese-hant-hk       |
| arabic-syria          | chinese-hant-mo       |
| arabic-SY             | chinese-hant          |
| armenian              | chinese-simplified-hongkongsarchina |
| assamese              | chinese-simplified-macausarchina |
| asturian              | chinese-simplified-singapore |
| asu                   | chinese-simplified    |
| australian            | chinese-traditional-hongkongsarchina |
| austrian              | chinese-traditional-macausarchina |
| azerbaijani-cyrillic  | chinese-traditional   |
| azerbaijani-cyrl      | chinese              |
| azerbaijani-latin     | colognian             |
| azerbaijani-latn      | cornish               |
| azerbaijani           | croatian              |
| bafia                 | czech                 |
| bambara               | danish                |
| basaa                 | duala                 |
| basque                | dutch                 |
| belarusan             | dzongkha              |
| bemba                 | embu                  |
| bena                  | english-au            |
| bengali               | english-australia     |
| bodo                  | english-ca            |
| bosnian-cyrillic      | english-canada        |
| bosnian-cyrl          | english-gb            |
| bosnian-latin         | english-newzealand    |
| bosnian-latn          | english-nz            |
| bosnian               | english-unitedkingdom |
| brazilian             | english-unitedstates  |
| breton                | english-us            |
| british               | english               |
| bulgarian             | esperanto             |
| burmese               | estonian              |
| canadian              | ewe                   |
| cantonese             | ewondo                |
| catalan               | faroese               |
filipino
finnish
french-be
french-belgium
french-ca
french-canada
french-ch
french-lu
french-luxembourg
french-switzerland
french
friulian
fulah
galician
ganda
georgian
german-at
german-austria
german-ch
german-switzerland
german
greek
gujarati
gusii
hausa-gh
hausa-ghana
hausa-ne
hausa-niger
hausa
hawaiian
hebrew
hindi
hungarian
icelandic
igbo
inarisami
indonesian
interlingua
irish
italian
japanese
jolafonyi
kabuverdianu
kabyle
kako
kalaallisut
kalenjin
kamba
kannada
kashmiri
kazakh
khmer
kikuyu
kinyarwanda
konkani
korean
koyraborosenni
koyrachiini
kwasio
kyrgyz
lakota
langi
lao
latvian
lingala
lithuanian
lowersorbian
lsorbian
lubakatanga
luo
luxembourgish
luyia
macedonian
machame
makhuwameetto
makonde
malagasy
malay-bn
malay-brunei
malay-sg
malay-singapore
malay
malayalam
maltese
manx
marathi
masai
mazanderani
meru
meta
mexican
mongolian
morisyen
mundang
nama
nepali
newzealand
ngiemboon
ngomba
norsk
northernluri
northernsami
northernndebele
norwegianbokmal
norwegiannynorsk
nswissgerman
nuer
nyankole
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.[13]

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

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. Alternatively, you may set a font for a script – just precede its name (lowercase) with a star (eg, *devanagari). Babel takes care of the font language and the font script when languages are selected (as well as the writing direction); see the recognized languages above. In most cases, you will not need font-options, which is the same as in fontspec, but you may add further key/value pairs if necessary.

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

\documentclass{article}
\usepackage[swedish, bidi=default]{babel}
\babelprovide[import]{hebrew}
\babelfont{rm}{FreeSerif}

[13]See also the package combofont for a complementary approach.
If on the other hand you have to resort to different fonts, you could replace the red line above with, say:

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

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

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

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

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

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

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

This makes sure the OpenType script for Devanagari is deva and not dev2, 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).

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

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

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

**WARNING** Using \texttt{setxxx}font and \babelfont at the same time is discouraged, but very often works as expected. However, be aware with \texttt{setxxx}font the language system will not be set by babel and should be set with fontspec if necessary.
1.15 Modifying a language

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

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

\addto\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:

\renewcommand\spanishchaptername{Foo}

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

\addto\extras{russian}{\mymacro}

There is a counterpart for code to be run when a language is unselected: \noextras{lang}.

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

\AtBeginDocument{\renewcommand\contentsname{Foo}}

The changes may be discarded with a language selector, and the original value restored.
NOTE These macros (\captions{lang}, \extras{lang}) may be redefined, but must not be used as such – they just pass information to babel, which executes them in the proper context.

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

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

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

1.16 Creating a language

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

\babelprovide \langle options \rangle \langle language-name \rangle

If the language \langle language-name \rangle has not been loaded as class or package option and there are no \langle 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 language-name \rangle 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
(babel) it in the preamble with something like:
(babel) \renewcommand\mylangchaptername{..}
(babel) Reported on input line 18.

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

EXAMPLE If you need a language named arhinish:

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

EXAMPLE 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.

**import** = `{language-tag}`

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

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

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

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

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

There are about 200 ini files, with data taken from the 1df 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: \<language>date, which takes three arguments, namely, year, month and day numbers. In fact, \today calls \<language>today, which in turn calls \<language>date{\the\year}{\the\month}{\the\day}.

**captions** = `{language-tag}`

Loads only the strings. For example:

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

**hyphenrules** = `{language-list}`

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

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

If none of the listed hyphenrules exist, the default behavior applies. Note in this example we set chavacano as first option – without it, it would select spanish even if chavacano exists. A special value is +, which allocates a new language (in the \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:
In other engines it just suppresses hyphenation (because the pattern list is empty).

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

**script**

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

**language**

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

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

**onchar** ids | fonts

New 3.38 This option is much like an ‘event’ called with 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 the this locale (as set with \babelfont). This option is not compatible with mapfont. Characters can be added with \babelfontcharproperty.

**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 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.

**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.

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

1.17 Digits

New 3.20 About thirty ini files define a field named digits.native. When it is present, two macros are created: \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 maparabic in \babelprovide, \arabic is redefined to produce the native digits (this is done globally, to avoid inconsistencies in, for example, page numbering, and note as well dates do not rely on \arabic.)

For example:

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

Languages providing native digits in all or some variants are Arabic, Assamese, Bangla, Tibetan, Bodo, Central Kurdish, Dzongkha, Persian, Gujarati, Hindi, Khmer, Kannada, Konkani, Kashmiri, Lao, Northern Luri, Malayalam, Marathi, Burmese, Mazanderani, Nepali, Odia, Punjabi, Pashto, Tamil, Telugu, Thai, Uyghur, Urdu, Uzbek, Vai, Cantonese, Chinese.

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

1.18 Accessing language info

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

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

\textbackslash{iflanguage} \{\textbackslash{language}\}\{\textbackslash{true}\}\{\textbackslash{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 \textbackslash{iflanguage}, but note here “language” is used in the \TeX sense, as a set of hyphenation patterns, and not as its babel name. This macro takes three arguments. The first argument is the name of a language; the second and third arguments are the actions to take if the result of the test is true or false respectively.
If an ini file has been loaded for the current language, you may access the information stored in it. This macro 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).

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`.

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

With all of them, hyphenation in the rest of the word is enabled. If you don’t want to enable it, there is a starred counterpart: `\babelhyphen{*{soft}}` (which in most cases is equivalent to the original `\-`), `\babelhyphen{*{hard}}`, etc.
Note hard is also good for isolated prefixes (eg, anti-) and nobreak for isolated suffixes (eg, -ism), but in both cases `\babelhyphen*{nobreak}` is usually better. There are also some differences with \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).

\texttt{\textbackslash babelhyphenation} \{\langle language\rangle, \langle language\rangle, ...\}\{\langle exceptions\rangle\}

**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\langle lang\rangle` as well as the language-specific encoding (not set in the preamble by default). Multiple `\babelhyphenation`'s are allowed. For example:

\begin{Verbatim}
\texttt{\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 `\babelexceptions` (below) you may fine-tune line breaking (only \texttt{luatex}). Even if there are no patterns for the language, you can add at least some typical cases.

\texttt{\textbackslash babelpatterns} \{\langle language\rangle, \langle language\rangle, ...\}\{\langle patterns\rangle\}

**New 3.9m** In \texttt{luatex only}\footnote{With \texttt{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.} adds or replaces patterns for the languages given or, without the optional argument, for all languages. If a pattern for a certain combination already exists, it gets replaced by the new one.

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

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

**New 3.31** (Only \texttt{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 (\texttt{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 `\babelprovide`. See the sample on the babel repository. With both Unicode engines, spacing is based on the “current” em unit (the size of the previous char in \texttt{luatex}, and the font size set by the last `\selectfont` in \texttt{xetex}).

\texttt{\textbackslash babelposthyphenation} \{\langle hyphenrules-name\rangle\}\{\langle lua-pattern\rangle\}\{\langle replacement\rangle\}
With \texttt{luatex} it is now possible to define non-standard hyphenation rules, like \texttt{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):

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

In the replacements, a captured char may be mapped to another, too. For example, if the first capture reads ([\texttt{i\textumlaut{u}}]), the replacement could be \{1|\texttt{i}|\textumlaut{u}\}, which maps \texttt{i} to \texttt{i}, and \texttt{\textumlaut{u}} to \texttt{\textumlaut{u}}, so that the diaeresis is removed.

This feature is activated with the first \texttt{\babelposthyphenation}. See the \texttt{babel wik} for a more detailed description and some examples. It also describes an additional replacement type with the key \texttt{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 \texttt{string} replacement to replace a character (or series of them) by another character (or series of them). Thus, to enter \texttt{ž} as \texttt{zh} and \texttt{š} as \texttt{sh} in a newly created locale for transliterated Russian:

```
\babelprovide[hyphenrules=+]{russian-latin} % Create locale  
\babelposthyphenation{russian-latin}{([sz])h} % Create rule  
{  
  \texttt{string = \{1|sz|\$z\}},  
  remove  
}
```

In other words, it is a quite general tool. (A counterpart \texttt{\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 \texttt{\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. Some languages sharing the same script define macros to switch it (eg, \texttt{\textcyrillic}), but be aware they may also set the language to a certain default. Even the babel core defined \texttt{\textlatin}, but is was somewhat buggy because in some cases it messed up encodings and fonts (for example, if the main Latin encoding was \texttt{LY1}), and therefore it has been deprecated.

\texttt{\ensureasci} \{(text)\}

\footnote{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.}

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

If non-ASCII encodings are not loaded (or no encoding at all), it is no-op (also \TeX and \LaTeX are not redefined); otherwise, \ensureascii switches to the encoding at the beginning of the document if ASCII-savvy, or else the last ASCII-savvy encoding loaded. For example, if you load LY1, LGR, then it is set to LY1, but if you load LY1, T2A it is set to T2A.

The symbol encodings T51, T3, and T53 are not taken into account, since they are not used for “ordinary” text (they are stored in \BabelNonText, used in some special cases when no Latin encoding is explicitly set).

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

1.21 Selecting directions

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

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

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

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

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

There are some package options controlling bidi writing.

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

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

In luatex, basic-r provides a simple and fast method for R text, which handles numbers and unmarked L text within an R context many in typical cases. **New 3.19** Finally, basic supports both L and R text, and it is the preferred method (support for basic-r is currently limited). (They are named basic mainly because they only consider the intrinsic direction of scripts and weak directionality.)
New 3.29 In xetex, bidi-r and bidi-l resort to the package bidi (by Vafa Khalighi). Integration is still somewhat tentative, but it mostly works. For RL documents use the former, and for LR ones use the latter. There are samples on GitHub, under /required/babel/samples. See particularly lua-bidibasic.tex and lua-secenum.tex.

EXAMPLE The following text comes from the Arabic Wikipedia (article about Arabia). Copy-pasting some text from the Wikipedia is a good way to test this feature. Remember basic is available in luatex only.

```latex
\documentclass{article}
\usepackage[bidi=basic]{babel}
\babelprovide[import, main]{arabic}
\babelfont{rm}{FreeSerif}
\begin{document}

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

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

In this example, and thanks to 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).
**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` must be defined to select the main language):

```latex
\newcommand\refrange[2]{{\textsubscript{LR}{\text{\ref{#1}}-\text{\ref{#2}}}}}
```

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

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

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

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

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

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

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

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

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

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

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

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

---

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

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

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

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

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

\texttt{\textbackslash babelsublr} \texttt{\{(lr-text)\}}

Digits in pdftex must be marked up explicitly (unlike \textit{luatex} with \texttt{bidi=basic} or \texttt{bidi=basic-r} and, usually, \textit{xetex}). This command is provided to set \texttt{\{(lr-text)\}} in L mode if necessary. It's intended for what Unicode calls weak characters, because words are best set with the corresponding language. For this reason, there is no \texttt{rl} counterpart. Any \texttt{\babelsublr} in 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:

```
RTL A ltr text \thechapter{} and still ltr RTL B
```

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

```
RTL A \texttt{\foreignlanguage{english}\{ltr text \thechapter{} and still ltr\}} RTL B
```

\texttt{\textbackslash BabelPatchSection} \texttt{\{(section-name)\}}

Mainly for bidi text, but it could be useful in other cases. \texttt{\textbackslash BabelPatchSection} and the corresponding option \texttt{layout=sectioning} takes a more logical approach (at least in many cases) because it applies the global language to the section format (including the \texttt{\chaptername} in \texttt{\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 \texttt{layout=sectioning} all the standard sectioning commands are redefined (it also “isolates” the page number in heads, for a proper bidi behavior), but with this command you can set them individually if necessary (but note then tocs and marks are not touched).

\texttt{\textbackslash BabelFootnote} \texttt{\{(cmd)\}\{\texttt{\{(local-language)\}\{\{before\}\}}\{\{after\}\}}

\texttt{New 3.17} Something like:

```
\texttt{\textbackslash BabelFootnote\{\textbackslash parsfootnote\}\{\texttt{\textbackslash languagename}\}\{}\}{}
```
defines \parsfootnote so that \parsfootnote{note} is equivalent to:

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

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

\BabelFootnote{\enfootnote}{english}

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

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

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

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

The same name can be applied to several events. Hooks may be enabled and disabled for all defined events with \EnableBabelHook{\name}, \DisableBabelHook{\name}. Names containing the string babel are reserved (they are used, for example, by \useshortands* to add a hook for the event afterextras). \AddBabelHook They may be also applied to a specific language with the optional argument; language-specific settings are executed after global ones.
Current events are the following; in some of them you can use one to three \TeX parameters (#1, #2, #3), with the meaning given:
addialect (language name, dialect name) Used by luababel.def to load the patterns if not preloaded.

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

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

defaultcommands Used (locally) in \StartBabelCommands.

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

stopcommands Used to reset the above, if necessary.

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

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

afterextras Just after executing \extras{/language}. For example, the following deactivates shorthands in all languages:

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

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

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

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

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

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

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

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

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

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

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

1.24 Languages supported by babel with ldf files

In the following table most of the languages supported by babel with and .ldf file are listed, together with the names of the option which you can load babel with for each language. Note this list is open and the current options may be different. It does not include ini files.
There are more languages not listed above, including hindi, thai, thaicjk, latvian, turkmen, magyar, mongolian, romansh, lithuanian, spanglish, vietnamese, japanese, pinyin, arabic, farsi, ibygreek, bgreek, serbianc, frenchle and friulan. Most of them work out of the box, but some may require extra fonts, encoding files, a preprocessor or even a complete framework (like CJK or luatexja). For example, if you have got the velthuis/devnag package, you can create a file with extension .dn:

```
\documentclass{article}
\usepackage[hindi]{babel}
\begin{document}
\end{document}
```
Then you preprocess it with devnag `<file>` , which creates `<file>.tex`; you can then typeset the latter with \TeX.

### 1.25 Unicode character properties in \texttt{luatex}

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

\begin{verbatim}
\texttt{\textbackslash babelcharproperty \{\texttt{char-code}\}\{\texttt{to-char-code}\}\{\texttt{property}\}\{\texttt{value}\}}
\end{verbatim}

New 3.32 Here, \{\texttt{char-code}\} 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:

\begin{verbatim}
\texttt{\textbackslash babelcharproperty\{`¿\}\{mirror\}\{`?\}\{\texttt{value}\}}
\texttt{\textbackslash babelcharproperty\{`-\}\{direction\}\{l\}\{\texttt{value}\}}
\texttt{\textbackslash babelcharproperty\{`\}\{linebreak\}\{cl\}\{\texttt{value}\}}
\end{verbatim}

New 3.39 Another property is locale, which adds characters to the list used by onchar in \texttt{\textbackslash babelprovide}, or, if the last argument is empty, removes them. The last argument is the locale name:

\begin{verbatim}
\texttt{\textbackslash babelcharproperty\{`\}\{locale\}\{english\}}
\end{verbatim}

### 1.26 Tweaking some features

\begin{verbatim}
\texttt{\textbackslash babeladjust \{\texttt{key-value-list}\}}
\end{verbatim}

New 3.36 Sometimes you might need to disable some babel features. Currently this macro understands the following keys (and only for \texttt{luatex}), 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{\textbackslash babeladjust\{bidi.text=off\}} if you are using an alternative algorithm or with large sections not requiring it. With \texttt{luahbtex} 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 book \texttt{and} you use \texttt{\ref} inside the argument of \texttt{\chapter} (or just use \texttt{\ref} inside \texttt{\MakeUppercase}), \TeX will keep complaining about an undefined label. To prevent such problems, you could revert to using uppercase labels, you can use \texttt{\textbackslash lowercase\{\texttt{\ref\{foo\}\}} inside the argument of \texttt{\chapter}, or, if you will not use shorthands in labels, set the \texttt{safe} option to none or \texttt{bib}.
• Both ltxdoc and babel use \AtBeginDocument to change some catcodes, and babel reloads \texttt{hhline} to make sure that has the right one, so if you want to change the catcode of $|$ it has to be done using the same method at the proper place, with

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

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 $|$ and $:$).

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

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

(A recent version of inputenc is required.)

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

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

• Babel does not take into account \verb|\lccodes| 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 (i.e., 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 \textit{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 \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.}

But that is the easy part, because they don’t require modifying the \texttt{\LaTeX} internals. Calendars (Arabic, Persian, Indic, etc.) are under study.

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

An option to manage bidirectional document layout in \textit{luatex} (lists, footnotes, etc.) is almost finished, but \texttt{xetex} required more work. Unfortunately, proper support for \texttt{xetex} requires patching somehow lots of macros and packages (and some issues related to \texttt{\special}s 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{\textbackslash foreignlanguage*} (a new starred version of \texttt{\textbackslash foreignlanguage}).

\noindent \textbf{Old and deprecated stuff}

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

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

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

So, for example:

\begin{verbatim}
\setmainfont[Language=Turkish]{Minion Pro}
\babelfSstore{turkish}
\setmainfont{Minion Pro}
\babelfSfeatures{turkish}{Language=Turkish}
\end{verbatim}

2 Loading languages with language.dat

\texttt{\LaTeX} and most engines based on it (pdf\TeX, \texttt{xetex}, \texttt{\epsilon\LaTeX}, the main exception being \texttt{luatex}) require hyphenation patterns to be preloaded when a format is created (eg. \texttt{\LaTeX}, \texttt{Xe\LaTeX}, \texttt{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).

\textbf{New 3.9q} With \texttt{luatex}, however, patterns are loaded on the fly when requested by the language (except the “0th” language, typically english, which is preloaded always) \footnote{This feature was added to 3.9o, but it was buggy. Both 3.9o and 3.9p are deprecated.} Until
3.9n, this task was delegated to the package \texttt{luatex-hyphen}, by Khaled Hosny, Élie Roux, and Manuel Pégourié-Gonnard, and required an extra file named \texttt{language.dat.lua}, but now a new mechanism has been devised based solely on \texttt{language.dat}. \textbf{You must rebuild the formats} if upgrading from a previous version. You may want to have a local \texttt{language.dat} for a particular project (for example, a book on Chemistry).\footnote{The loader for lua(\texttt{e})\TeX{} is slightly different as it's not based on babel but on etex. \texttt{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 \texttt{language.dat}.}

### 2.1 Format

In that file the person who maintains a \LaTeX{} environment has to record for which languages he has hyphenation patterns \textit{and} in which files these are stored\footnote{This is because different operating systems sometimes use very different file-naming conventions.}. When hyphenation exceptions are stored in a separate file this can be indicated by naming that file \textit{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 \LaTeX{} that the hyphenation patterns just processed have to be known under an alternative name. Here is an example:

```plaintext
% File     : language.dat
% Purpose  : tell initex what files with patterns to load.
english english.hyphenations
=british

dutch hyphen.dutch exceptions.dutch % Nederlands
german hyphen.ger
```

You may also set the font encoding the patterns are intended for by following the language name by a colon and the encoding code.\footnote{This is not a new feature, but in former versions it didn't work correctly.} For example:

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

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

A typical error when using babel is the following:

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

It simply means you must reconfigure \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 \textit{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 \texttt{babel.def}, i.e., the definitions of
the macros that produce texts. Also the language-switching possibility which has been built into the babel system has its implications.

The following assumptions are made:

- **Some of the language-specific definitions might be used by plain \TeX users, so the files have to be coded so that they can be read by both \LaTeX and plain \TeX. The current format can be checked by looking at the value of the macro \texttt{\textbackslash 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{\textbackslash (lang)hyphenmins}, \texttt{\textbackslash captions(lang)}, \texttt{\textbackslash date(lang)}, \texttt{\textbackslash extras(lang)} and \texttt{\textbackslash noextras(lang)} (the last two may be left empty); where \texttt{(lang)} 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{\textbackslash date(lang)} but not \texttt{\textbackslash captions(lang)} does not raise an error but can lead to unexpected results.**

- **When a language definition file is loaded, it can define \texttt{\textbackslash l@lang} to be a dialect of \texttt{\textbackslash language0} when \texttt{\textbackslash l@lang} is undefined.**

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

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

Some recommendations:

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

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

- **Avoid adding things to \texttt{\textbackslash noextras(lang)} except for umlauthigh and friends, \texttt{\textbackslash bb1@deactivate}, \texttt{\textbackslash bb1@nonfrenchspacing}, and language-specific macros. Use always, if possible, \texttt{\textbackslash bb1@save} and \texttt{\textbackslash bb1@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 \texttt{\textbackslash 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 \texttt{\textbackslash latin\texttt{\textbackslash text is deprecated}}.**

- **Please, for "private" internal macros do not use the \texttt{\textbackslash bb1@ prefix. It is used by babel and it can lead to incompatibilities.**

\textsuperscript{24}But not removed, for backward compatibility.
There are no special requirements for documenting your language files. Now they are not included in the base babel manual, so provide a standalone document suited for your needs, as well as other files you think can be useful. A PDF and a “readme” are strongly recommended.

3.1 Guidelines for contributed languages

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

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

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

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

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

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

3.2 Basic macros

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

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

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

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

\renewcommand\spanishhyphenmins(34)

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

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

\captions{lang}
The macro \captions{lang} defines the macros that hold the texts to replace the original hard-wired texts.

\date{lang}
The macro \date{lang} defines \today.

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

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

\bbl@declare@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 \ProvidesPackage command. \ProvidesLanguage syntax is similar to the syntax of the \ProvidesPackage command.

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

\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, \TeX{} can be instructed to load a local configuration file. This file can, for instance, be used to add strings to \captions{lang} to support local document classes. The user will be informed that this configuration file has been loaded. This macro is called by \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 (babel 3.9 and later).

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

\ifx\undefined\l@<language>
  \@nopatterns{<language>}
  \adddialect\l@<language>0
\fi
\adddialect\l@<dialect>\l@<language>
NOTE If for some reason you want to load a package in your style, you should be aware it cannot be done directly in the ldf file, but it can be delayed with \AtEndOfPackage. Macros from external packages can be used inside definitions in the ldf itself (for example, \extras{<language>}), but if executed directly, the code must be placed inside \AtEndOfPackage. A trivial example illustrating these points is:

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

3.4 Support for active characters

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

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

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

\bbl@add@special \bbl@remove@special The \TeX{}book states: “Plain \TeX{} includes a macro called \dospecials that is essentially a set macro, representing the set of all characters that have a special category code.” \cite[p. 380]{\TeXbook} It is used to set text ‘verbatim’. To make this work if more characters get a special category code, you have to add this character to the macro \dospecial. \LaTeX{} adds another macro called \@sanitize representing the same character set, but without the curly braces. The macros \bbl@add@special⟨char⟩ and \bbl@remove@special⟨char⟩ add and remove the character ⟨char⟩ to these two sets.

3.5 Support for saving macro definitions

Language definition files may want to redefine macros that already exist. Therefore a mechanism for saving (and restoring) the original definition of those macros is provided. We provide two macros for this\footnote{This mechanism was introduced by Bernd Raichle.}.

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

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

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

3.6 Support for extending macros

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

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

3.7 Macros common to a number of languages

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

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

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

\setlow@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. 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.

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

3.8 Encoding-dependent strings

New 3.9a Babel 3.9 provides a way of defining strings in several encodings, intended mainly for luatex and xetex. This is the only new feature requiring changes in language files if you want to make use of it. Furthermore, it must be activated explicitly, with the package option strings. If there is no strings, these blocks are ignored, except \SetCases (and except if forced as described below). In other words, the old way of defining/switching strings still works and it's used by default.

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

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

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

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

A “selector” is a name to be used as value in package option strings, optionally followed by extra info about the encodings to be used. The name unicode must be used for xetex and luatex (the key strings has also other two special values: generic and encoded).

If a string is set several times (because several blocks are read), the first one takes precedence (ie, it works much like \providecommand).

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

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

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

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

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

A real example is:

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

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

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

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

\StartBabelCommands{german,austrian}{date}
  \SetString\monthiname{Februar}
  \SetString\monthiiiname{M"{a}rz}
  \SetString\monthivname{April}
  \SetString\monthvname{Mai}
  \SetString\monthviiiname{Juni}
  \SetString\monthviiiname{Juli}
  \SetString\monthviiiname{August}
  \SetString\monthixname{September}
  \SetString\monthxname{Oktober}
  \SetString\monthxiiiname{November}
  \SetString\monthxviiname{Dezenber}
  \SetString\today{\number\day. \csname month\romannumeral\month name\endcsname \space \number\year}
\EndBabelCommands

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

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

\footnotesize\textsuperscript{26} In future releases further categories may be added.
\StartBabelCommands  *{⟨language-list⟩}{{category}⟩[⟨selector⟩]

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

\EndBabelCommands  Marks the end of the series of blocks.

\AfterBabelCommands  {⟨code⟩}

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

\SetString  ⟨{macro-name}⟩{⟨string⟩}

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

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

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

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

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

#1 is replaced by the roman numeral.

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

Sets globally code to be executed at \MakeUpperCase and \MakeLowerCase. The code would typically be things like \let\BB\bb and \uccode or \lccode (although for the reasons explained above, changes in l/c codes may not work). A ⟨map-list⟩ 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 Mi\TeX, we could set for Turkish:

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

\footnote{This replaces in 3.9g a short-lived \UseStrings which has been removed because it did not work.}
\SetCase
{\uccode`i="9D\relax
 \uccode"19=`I\relax}
{\lccode"9D=`i\relax
 \lccode`I="19\relax}
\EndBabelCommands

(Note the mapping for OT1 is not complete.)

\SetHyphenMap
{(to-lower-macros)}

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

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

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

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

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

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

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

4 Changes

4.1 Changes in babel version 3.9

Most of 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 \b 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 and error with a conditional.

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

• \l@english was defined incorrectly (using \let instead of \chardef).

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

Part II
Source code

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

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

We define some basic macros which just make the code cleaner. \bbl@add is now used internally instead of \addto because of the unpredictable behavior of the latter. Used in babel.def and in babel.sty, which means in \texttt{B}X 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.

```
\bbl@add@list\bbl@add@list\bbl@add@list
```

This internal macro adds its second argument to a comma separated list in its first argument. When the list is not defined yet (or empty), it will be initiated. It presumes

```
Because the code that is used in the handling of active characters may need to look ahead, we take extra care to ‘throw’ it over the \else and \fi parts of an if-statement.\footnote{This code is based on code presented in TUGboat vol. 12, no2, June 1991 in “An expansion Power Lemma” by Sonja Maus.} These macros will break if another if...fi statement appears in one of the arguments and it is not enclosed in braces.

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.

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.

To check if a macro is defined, we create a new macro, which does the same as \@ifundefined. However, in an \-tex engine, it is based on \ifcsname, which is more efficient, and do not waste memory.
\def\bbl@ifblank#1{\bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo}
\long\def\bbl@ifblank@i#1#2\@nil#3#4#5\@nil{#4}

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

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

\def\bbl@replace#1#2#3{% in #1 -> repl #2 by #3
\toks@{}%
\def\bbl@replace@aux##1#2##2#2{%
\ifx\@nil##2%
\toks@\expandafter{\the\toks@##1}%
\else
\toks@\expandafter{\the\toks@##1#3}%
\bbl@afterfi
\bbl@replace@aux##2#2%
\fi}%
\def\bbl@replace@aux\@nil\@nil#1#2#3#4#5{
\bbl@replace\@secondoftwo\@firstoftwo
\bbl@replace\@afterfi\expandafter\@firstoftwo
\bbl@afterfi
}\bbl@replace
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).

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

\begin{verbatim}
\edef\bbl@samestring{% #1 and #2 are #1\@empty % Unused macros if old Plain TeX
\edef\bbl@exp{\bl@parsedef\detokenize{macro:}}#2->#3\relax{%\begin{group}
\edef\bbl@tempa{#1}\edef\bbl@tempb{#2}\edef\bbl@tempe{#3}\bbl@xin{\bbl@tempc}{\bbl@tempe}% If not in macro, do nothing\ifin@
\bbl@exp{\\bbl@replace\\bbl@tempe{\bbl@tempc}{\bbl@tempd}}%\edef\bbl@tempc{% Expanded an executed below as 'uplevel'
\\\makeatletter % "internal" macros with @ are assumed
\\\scantokens{%\bbl@tempa\\@namedef{\bbl@stripslash#1}\bbl@tempb{\bbl@tempe}}%\catcode64=\the\catcode64\relax}% Restore @\else\let\bbl@tempc\@empty % Not \relax\fi\bbl@exp{% For the 'uplevel' assignments\end{group}\bbl@tempc}{% empty or expand to set #1 with changes\fi}
\end{verbatim}

\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\bbl@exp{% #1 and #2 are % For the 'uplevel' assignments\end{verbatim}

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

\begin{verbatim}
\chardef\bbl@engine=%\ifx\directlua\@undefined\ifx\XeTeXinputencoding\@undefined\z@\else59\fi\fi\end{verbatim}

59
Some files identify themselves with a \LaTeX macro. The following code is placed before
them to define (and then undefine) if not in \LaTeX.

```latex
\ifx\ProvidesFile\@undefined
  \def\ProvidesFile#1[#2 #3 #4]{%
  \wlog{File: #1 #4 #3 <#2>}%
  \let\ProvidesFile\@undefined}
\fi
```

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

```latex
\ifx\directlua\@undefined\else
  \ifx\bbl@luapatterns\@undefined
    \input luababel.def
  \fi
\fi
```

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

```latex
\ifx\AtBeginDocument\@undefined
  \input plain.def\relax
\fi
```

### 7.1 Multiple languages

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

```latex
\ifx\language\@undefined
  \csname newcount\endcsname\language
\fi
```

\lastlanguage 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.

\begin{verbatim}
176 ⟨⟨Define core switching macros⟩⟩ ≡
177 \ifx\newlanguage\@undefined
178 \csname newcount\endcsname\last@language
179 \def\addlanguage#1{%
180 \global\advance\last@language\@ne
181 \ifnum\last@language<\@cclvi
182 \else
183 \errmessage{No room for a new \string\language!}%
184 \fi
185 \global\chardef#1\last@language
186 \wlog{\string#1 = \string\language\the\last@language}
187 \else
188 \countdef\last@language=19
189 \def\addlanguage{\alloc@9\language\chardef@cclvi}
190 \fi
191 ⟨⟨/Define core switching macros⟩⟩
\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.

\section{The Package File (\LaTeX, babel.sty)}

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

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

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

\subsection{base}

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

\begin{verbatim}
192 ⟨⟨package⟩⟩
193 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
194 \ProvidesPackage{babel}{⟨⟨date⟩⟩ ⟨⟨version⟩⟩ The Babel package}
195 \@ifpackagewith{babel}{debug}{
196 {\providecommand\bbl@trace[1]{\message{\message[^J[ #1 ]}={()=>
197 \let\bbl@debug\@firstofone}
198 {\providecommand\bbl@trace[1]{}%
199 \let\bbl@debug\@gobble}
\end{verbatim}

61
\ifx\bbl@switchflag\undefined % Prevent double input
\let\bbl@switchflag\relax
\input switch.def\relax
\fi

\langle\langle
Load patterns in luatex\rangle\rangle
\langle\langle
Basic macros\rangle\rangle
\def\AfterBabelLanguage{\%
If the format created a list of loaded languages (in \bbl@languages), get the name of the 0-th to show the actual language used.
\ifx\bbl@languages\undefined\else
\begingroup
\catcode`\^^I=12
@ifpackagewith{babel}{showlanguages}{%
\def\bbl@elt{\wlog{\*\<languages>\%}
\bbl@languages
\wlog{</languages>\%}
\endgroup}{}
\def\bbl@elt{\ifnum#2=\z@
\gdef\bbl@nulllanguage{#1}\
\def\bbl@elt##1##2##3##4{}%\fi}
\bbl@languages
\fi
\ifodd\bbl@engine
\def\bbl@activate@preotf{\%
\let\bbl@activate@preotf\relax % only once
\directlua{
Babel = Babel or {}
function Babel.pre_otfload_v(head)
if Babel.numbers and Babel.digits_mapped then
  head = Babel.numbers(head)
end
if Babel.bidi_enabled then
  head = Babel.bidi(head, false, dir)
end
return head
end

function Babel.pre_otfload_h(head, gc, sz, pt, dir)
if Babel.numbers and Babel.digits_mapped then
  head = Babel.numbers(head)
end
if Babel.bidi_enabled then
  head = Babel.bidi(head, false, dir)
end
return head
end

luatexbase.add_to_callback('pre_linebreak_filter',
  Babel.pre_otfload_v, 'Babel.pre_otfload_v',
  luatexbase.priority_in_callback('pre_linebreak_filter',

Now the base option. With it we can define (and load, with \texttt{luatex}) hyphenation patterns, even if we are not interested in the rest of \texttt{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 \texttt{\BabelModifiers} at \texttt{\bbl@load@language}; when no modifiers have been given, the former is \texttt{\relax}. How modifiers are handled are left to language styles; they can use \texttt{\in@}, loop them with \texttt{@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.

```latex
\DeclareOption{KeepShorthandsActive}{}
\DeclareOption{activeacute}{}
\DeclareOption{activegrave}{}
\DeclareOption{debug}{}
\DeclareOption{noconfigs}{}
\DeclareOption{showlanguages}{}
\DeclareOption{silent}{}
\DeclareOption{mono}{}
\DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
% Don't use. Experimental:
\newif\ifbbl@single
\DeclareOption{selectors=off}{\bbl@singletrue}
```

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@tempa1=#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@nil
```
Now we finish the first pass (and start over).

8.3 Conditional loading of shorthands

If there is no shorthands=chars, the original babel macros are left untouched, but if there is, these macros are wrapped (in babel.def) to define only those given.

A bit of optimization: if there is no shorthands=, then \bbl@ifshorthand is always true, and it is always false if shorthands is empty. Also, some code makes sense only with shorthands=.

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

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

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

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.
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.

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

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

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

Now, we make sure an option is explicitly declared for any language set as global option, by checking if an .ldf exists. The previous step was, in fact, somewhat redundant, but that way we minimize accessing the file system just to see if the option could be a language.
If a main language has been set, store it for the third pass.
\ifx\bbl@opt@main\@nnil\else
\expandafter
\let\expandafter\bbl@loadmain\csname ds@\bbl@opt@main\endcsname
\DeclareOption{\bbl@opt@main}{\bbl@loadmain}\fi

And we are done, because all options for this pass have 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):
\def\AfterBabelLanguage#1{\bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang}{}%}
\DeclareOption*{}
\ProcessOptions*

This finished the second pass. Now the third one begins, which loads the main language set with the key main. A warning is raised if the main language is not the same as the last named one, or if the value of the key main is not a language. Then execute directly the option (because it could be used only in main). After loading all languages, we deactivate \AfterBabelLanguage.
\ifx\bbl@opt@main\@nnil
\edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
\let\bbl@tempc\@empty
\bbl@for\bbl@tempb\bbl@tempa{%
\bbl@xin\{,\bbl@tempb,\{,\bbl@loaded,\}\%
\fin\edef\bbl@tempc{\bbl@tempb}\fi}
\def\bbl@tempa!#1\@nnil{\def\bbl@tempb{#1}}
\expandafter\bbl@tempa\bbl@loaded,\@nnil
\ifx\bbl@tempb\bbl@tempc\else
\bbl@warning{\%
Last declared language option is `\bbl@tempc',\%
but the last processed one was `\bbl@tempb'.\%
The main language cannot be set as both a global\%
and a package option. Use `main=\bbl@tempc' as\%
option. Reported}%
\fi\else
\DeclareOption{\bbl@opt@main}{\bbl@loadmain}\ExecuteOptions{\bbl@opt@main}\fi
\def\AfterBabelLanguage{\bbl@error
{Too late for \string\AfterBabelLanguage}%
{Languages have been loaded, so I can do nothing})

In order to catch the case where the user forgot to specify a language we check whether \bbl@main@language, has become defined. If not, no language has been loaded and an error message is displayed.
The kernel of Babel (babel.def, common)

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

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

9.1 Tools

The file babel.def expects some definitions made in the \LaTeX\2ε style file. So, In \LaTeX\2.09 and Plain we must provide at least some predefined values as well some tools to set them (even if not all options are available). There are no package options, and therefore and alternative mechanism is provided. For the moment, only babeloptionstrings and babeloptionmath are provided, which can be defined before loading babel. \BabelModifiers can be set too (but not sure it works).

\BabelStringsDefault can be set too (but not sure it works).
\def\AfterBabelLanguage#1#2{}
\ifx\BabelModifiers@undefined\let\BabelModifiers\relax\fi
\let\bbl@afterlang\relax
\def\bbl@opt@safe{BR}
\ifx\@uclclist@undefined\let\@uclclist\@empty\fi
\ifx\bbl@trace\@undefined\def\bbl@trace#1{}\fi
\expandafter\newif\csname ifbbl@single\endcsname
\ifx\bbl@switchflag\@undefined % Prevent double input
\let\bbl@switchflag\relax
\input switch.def\relax
\fi
\bbl@trace{Compatibility with language.def}
\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
\chardef\l@english\z@
\fi
\langle\langle
Load patterns in luatex\rangle\rangle
\langle\langle
Basic macros\rangle\rangle
\def\addto#1#2{\%
\ifx#1\@undefined
\def#1{#2}%%
\else
\ifx#1\relax
\def#1{#2}%%
\else
{\toks@\expandafter{#1#2}%%
\xdef#1{\the\toks@}}%%
\fi
\fi}
\addto For each language four control sequences have to be defined that control the
language-specific definitions. To be able to add something to these macro once they have
been defined the macro \addto is introduced. It takes two arguments, a \langle control sequence \rangle
and \TeX-code to be added to the \langle control sequence \rangle. If the \langle control sequence \rangle has not been defined before it is defined now. The control
sequence could also expand to \relax, in which case a circular definition results. The net
result is a stack overflow. Otherwise the replacement text for the \langle control sequence \rangle is
expanded and stored in a token register, together with the \TeX-code to be added. Finally
the \langle control sequence \rangle is redefined, using the contents of the token register.
\addto
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.

```
def\bbl@withactive#1\#2{% 
\begingroup
  \lcodo{\=\#2}\relax
  \lowercase{\endgroup\#1}}
```

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

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

```
def\bbl@redefine#1{% 
  \edef\bbl@tempa{\bbl@stripslash#1}%
  \expandafter\let\csname org@\bbl@tempa\endcsname#1%
  \expandafter\def\csname\bbl@tempa\endcsname}
```

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

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

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

```
def\bbl@redefine@long#1{%
  \edef\bbl@tempa{\bbl@stripslash#1}%
  \expandafter\let\csname org@\bbl@tempa\endcsname#1%
  \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
```

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

```
@onlypreamble\bbl@redefine@long
```

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

```
def\bbl@redefinerobust#1{%
  \edef\bbl@tempa{\bbl@stripslash#1}%
  \bbl@ifunset{\bbl@tempa\space}{}
  \edef\bbl@tempa##1,#3=##2,##3\@empty{
    \def\bbl@tempb{##2}}% 
  \expandafter\let\csname org@\bbl@tempa\endcsname#1%
  \\expandafter\def\csname\bbl@tempa\endcsname
}
```

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

```
@onlypreamble\bbl@redefinerobust
```

### 9.2 Hooks

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

```
bbl@trace{Hooks}
```

```
cnewcommand\AddBabelHook[3][]{%
  \bbl@ifunset{\bbl@hk@#2}{\EnableBabelHook{#2}}{}
  \def\bbl@tempa##1,#3=##2,##3\@empty{\def\bbl@tempb{##2}}%
}
```

71
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 \texttt{hyphen.cfg} are also loaded (just in case you need them for some reason).

\begin{verbatim}
def\bbl@usehooks#1#2{% 
\def\bbl@elt##1{\@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1@}#2}}% 
\@nameuse{bbl@ev@#1@}% 
\ifx\languagename\@undefined\else % Test required for Plain (?) 
\def\bbl@elt##1{\@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1@\languagename}#2}}% 
\@nameuse{bbl@ev@#1@\languagename}% 
\fi}
\end{verbatim}

\newcommand\EnableBabelHook[1]{\bbl@csarg\let{hk@#1}\@firstofone}
\newcommand\DisableBabelHook[1]{\bbl@csarg\let{hk@#1}\@gobble}
\def\bbl@usehooks#1#2{% 
\bbl@foreach\bbl@tempa\expandafter{\bbl@evargs,% <- don't delete this comma 
\expandafter{\bbl@tempb\@@}}% 
\edef\bbl@tempa{\bbl@tempb\@@}% 
\edef\bbl@tempb##1{\ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}% 
\edef\bbl@tempa{\bbl@tempb\@@}%}

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

\begin{verbatim}
\edef\bbl@evargs{,% \texttt{\expandafter{\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}
\end{verbatim}

\begin{verbatim}
def\bbl@usehooks#1#2{% 
\def\bbl@elt##1{\@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1@\languagename}#2}}% 
\@nameuse{bbl@ev@#1@\languagename}% 
\ifx\languagename\@undefined\else % Test required for Plain (?) 
\def\bbl@elt##1{\@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1@\languagename}#2}}% 
\@nameuse{bbl@ev@#1@\languagename}% 
\fi}
\end{verbatim}

\newcommand\EnableBabelHook[1]{\bbl@csarg\let{hk@#1}\@firstofone}
\newcommand\DisableBabelHook[1]{\bbl@csarg\let{hk@#1}\@gobble}
\def\bbl@usehooks#1#2{% 
\bbl@foreach\bbl@tempa\expandafter{\bbl@evargs,% <- don't delete this comma 
\expandafter{\bbl@tempb\@@}}% 
\edef\bbl@tempa{\bbl@tempb\@@}% 
\edef\bbl@tempb##1{\ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}% 
\edef\bbl@tempa{\bbl@tempb\@@}%}

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

\begin{verbatim}
\edef\bbl@evargs{,% \texttt{\expandafter{\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}
\end{verbatim}

\begin{verbatim}
def\bbl@usehooks#1#2{% 
\def\bbl@elt##1{\@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1@\languagename}#2}}% 
\@nameuse{bbl@ev@#1@}% 
\ifx\languagename\@undefined\else % Test required for Plain (?) 
\def\bbl@elt##1{\@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1@\languagename}#2}}% 
\@nameuse{bbl@ev@#1@\languagename}% 
\fi}
\end{verbatim}

\newcommand\EnableBabelHook[1]{\bbl@csarg\let{hk@#1}\@firstofone}
\newcommand\DisableBabelHook[1]{\bbl@csarg\let{hk@#1}\@gobble}
\def\bbl@usehooks#1#2{% 
\bbl@foreach\bbl@tempa\expandafter{\bbl@evargs,% <- don't delete this comma 
\expandafter{\bbl@tempb\@@}}% 
\edef\bbl@tempa{\bbl@tempb\@@}% 
\edef\bbl@tempb##1{\ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}% 
\edef\bbl@tempa{\bbl@tempb\@@}%}

\texttt{babelensure} The user command just parses the optional argument and creates a new macro named \texttt{\bbl@e@[language]}. We register a hook at the \texttt{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{\bbl@e@[language]} contains \texttt{\bbl@ensure{\langle include\rangle}}\{\langle exclude\rangle\}\{\langle fontenc\rangle\}, which in turn loops over the macros names in \texttt{\bbl@captionslist}, excluding (with the help of \texttt{in@}) those in the exclude list. If the \texttt{fontenc} is given (and not \texttt{\relax}), the \texttt{\fontencoding} is also added. Then we loop over the include list, but if the macro already contains \texttt{\foreignlanguage}, nothing is done. Note this macro (1) is not restricted to the preamble, and (2) changes are local.
9.3 Setting up language files

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

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

Another character that needs to have the correct category code during processing of language definition files is the equals sign, ‘=’, because it is sometimes used in constructions
with the \let primitive. Therefore we store its current catcode and restore it later on. Now we check whether we should perhaps stop the processing of this file. To do this we first need to check whether the second argument that is passed to \LdfInit is a control sequence. We do that by looking at the first token after passing \#2 through string. When it is equal to \@backslashchar we are dealing with a control sequence which we can compare with \@undefined. If so, we call \ldf@quit to set the main language, restore the category code of the @-sign and call \endinput. When \#2 was not a control sequence we construct one and compare it with \relax. Finally we check \originalTeX.

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

\ldf@quit This macro interrupts the processing of a language definition file.
\def\ldf@quit{#1}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}1{%
\bbl@afterlang
\let\bbl@afterlang\relax
\let\BabelModifiers\relax
\let\bbl@screset\relax
\def\ldf@finish{#1}1%
\loadlocalcfg{#1}%
After the preamble of the document the commands \LdfInit, \ldf@quit and \ldf@finish are no longer needed. Therefore they are turned into warning messages in \LaTeX.

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

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

\def\main@language#1{% 
\def\bbl@main@language{#1} 
\let\languagename\bbl@main@language 
\bbl@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{% 
\@nameuse{bbl@beforestart} 
\if@filesw 
\immediate\write\@mainaux{\string\@nameuse{bbl@beforestart}} 
\fi 
\expandafter\selectlanguage\expandafter{\bbl@main@language} 
\if\bbl@single % must go after the line above 
\renewcommand\selectlanguage[1]{} 
\renewcommand\foreignlanguage[2]{#2} 
\global\let\babel@aux@gobbletwo \@gobbletwo % Also as flag 
\fi 
\ifcase\bbl@engine\or\pagedir\bodydir\fi} % TODO - a better place

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

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

\subsection{9.4 Shorthands}

\bbl@add@special

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

Items are added to the lists without checking its existence or the original catcode. It does not hurt, but should be fixed. It's already done with \nfss@catcodes, added in 3.10.
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.

\bbl@active@def, \bbl@active@def, \bbl@active@def

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

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.

\begin{verbatim}
\long@namedef{#3@arg#1}##1{\expandafter\ifx\csname#2@sh@#1\string##1\endcsname\relax
\bbl@afterelse\csname#4#1\endcsname##1%
\else\bbl@afterfi\csname#2@sh@#1\string##1\endcsname\fi}}%
\initiate@active@char

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

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

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 treatement to avoid making them \relax).

\begin{verbatim}
def@\initiate@active@char#1#2#3{%\bbl@csargedef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%\ifx#1\@undefined\bbl@csargedef{oridef@@#2}{\let\noexpand#1\noexpand\@undefined}%\else\bbl@csargdef{oridef@@#2}#1%\bbl@csargedef{oridef@#2}{\let\noexpand#1%\expandafter\noexpand\csname bbl@oridef@@#2\endcsname}%\fi\end{verbatim}

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\langle char\rangle to expand to the character in its default state. If the character is mathematically active when babel is loaded (for example \texttt{')}) the normal expansion is somewhat different to avoid an infinite loop (but it does not prevent the loop if the mathcode is set to *8000 a posteriori*).

\begin{verbatim}
\ifx#1#3\relax\expandafter\let\csname normal@char#2\endcsname#3%
\else\bbl@info{Making #2 an active character}%\ifnum\mathcode`#2=\ifodd\bbl@engine"1000000 \else"8000 \fi
\namedef{normal@char#2}{\textormath{#3}{\csname bbl@oridef@@#2\endcsname}}%
\else\namedef{normal@char#2}{#3}%\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 \texttt{KeepShorthandsActive}). It is re-activate again at \texttt{\begin{document}}. We also need to make sure that the shorthands are active during the processing of the \texttt{.aux} file. Otherwise some citations may give unexpected results in the printout when a shorthand was used in
the optional argument of \bibitem for example. Then we make it active (not strictly necessary, but done for backward compatibility).

Now we have set \normal@char, we must define \active@char, to be executed when the character is activated. We define the first level expansion of \active@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 to start the search of a definition in the user, language and system levels (or eventually normal@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

\active@prefix \normal@char

(where \active@char is one control sequence!).

The next level of the code checks whether a user has defined a shorthand for himself with this character. First we check for a single character shorthand. If that doesn’t exist we check for a shorthand with an argument.
In order to do the right thing when a shorthand with an argument is used by itself at the end of the line we provide a definition for the case of an empty argument. For that case we let the shorthand character expand to its non-active self. Also, when a shorthand combination such as ‘’ ends up in a heading TeX would see \protect \protect . To prevent this from happening a couple of shorthand needs to be defined at user level.

\begin{verbatim}
\edef\csname user@group @sh@#2@@\endcsname\expandafter\noexpand\csname normal@char#2\endcsname\expandafter\edef\csname user@group @sh@#2@\protect@\endcsname\expandafter\noexpand\csname user@active#2\endcsname
\end{verbatim}

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

\begin{verbatim}
\if\string'#2%
\let\prim@s\bbl@prim@s\let\active@math@prime#1%
\fi
\bbl@usehooks{initiateactive}{#1}{#2}{#3}}
\end{verbatim}

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

\begin{verbatim}
\DeclareOption{math=active}{\def\bbl@mathnormal\relax}
\DeclareOption{math=normal}{\def\bbl@mathnormal\textormath}
\end{verbatim}

Initiating a shorthand makes active the char. That is not strictly necessary but it is still done for backward compatibility. So we need to restore the original catcode at the end of package and the end of the ldf.

\begin{verbatim}
\@ifpackagewith{babel}{KeepShorthandActive}\let\bbl@restoreactive\@gobble\def\bbl@restoreactive#1{%\AfterBabelLanguage\CurrentOption{\catcode`#1=\the\catcode`#1\relax}\AtEndOfPackage{\catcode`#1=\the\catcode`#1\relax}}\AtEndOfPackage{\let\bbl@restoreactive\@gobble}
\end{verbatim}

\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.

\begin{verbatim}
\def\bbl@sh@select#1#2{%\expandafter\ifx\csname#1@sh@#2@sel\endcsname\relax\bbl@afterelse\bbl@scndcs\else\bbl@afterfi\csname#1@sh@#2@sel\endcsname\fi}
\end{verbatim}

\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.

\begin{group}
\{\gdef\active@prefix\#1{\protect\@typeset@protect}\}
\else
\{\gdef\active@prefix\#1{\protect\@unexpandable@protect}
\noexpand\#1\}
\else
\{\gdef\active@prefix\#1{\protect\#1\}
\expandafter\@gobble\}
\fi\}
\{\gdef\active@prefix\#1{\ifincsname\string\#1\expandafter\@gobble\}
\else\ifincsname\protect\#1\expandafter\expandafter\expandafter\@gobble\fi\fi\}
\endgroup

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

\newif\if@safe@actives
\@safe@activesfalse
\bbl@restore@actives

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

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

\bbl@activate \bbl@deactivate

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

\def\bbl@activate\#1{\bbl@withactive{\expandafter\let\expandafter}\#1}\}
\def\bbl@deactivate\#1{\bbl@withactive{\expandafter\let\expandafter}\#1}\}

\bbl@firstcs \bbl@scndcs

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

\def\bbl@firstcs\#1\#2{\csname#1\endcsname}\}
\def\bbl@scndcs\#1\#2{\csname#2\endcsname}\}
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.

\textormath 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.

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

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

\begin{verbatim}
\end{verbatim}

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

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

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

\bbl@switch@sh The macro \bbl@switch@sh takes the list of characters apart one by one and subsequently switches the category code of the shorthand character according to the first argument of \bbl@switch@sh.

But before any of this switching takes place we make sure that the character we are dealing with is known as a shorthand character. If it is, a macro such as `\active@char" should exist.

Switching off and on is easy – we just set the category code to ‘other’ (12) and `\active. With the starred version, the original catcode and the original definition, saved in \@initiate@active@char, are restored.

\def\bbl@switch@sh#1#2{% 
\ifx\@nil#2\else 
\bbl@ifunset{bbl@active@\string#2}{\bbl@error{I cannot switch `\string#2' on or off--not a shorthand} \% 
{This character is not a shorthand. Maybe you made\%
  a typing mistake? I will ignore your instruction} \% 
{You may proceed, but expect unexpected results}} \% 
{I cannot switch `\string#2' on or off--not a shorthand} \% 
{This character is not a shorthand. Maybe you made\%
  a typing mistake? I will ignore your instruction} \% 
{You may proceed, but expect unexpected results}} \% 
\catcode`#2\relax
\or 
\catcode`#2\active
\or 
\csname bbl@oricat@\string#2\endcsname
\csname bbl@oridef@\string#2\endcsname
\fi \% 
\bbl@afterfi\bbl@switch@sh#1\% 
\fi}
Note the value is that at the expansion time, eg, in the preample shorthands are usually deactivated.

1055 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
1056 \def\bbl@putsh#1{% 
1057 \bbl@ifunset{\bbl@active@#1}{\empty}{\csname\string#1@endcsname}}
1059 \csname language@name @sh@#1@endcsname}
1060 \ifx\bbl@opt@shorthands\@nnil\else
1064 \let\bbl@switch@sh\bbl@switch@sh
1065 \ifx\@empty#2\else\string#2@i\endcsname}
1069 \ifx#2\@nnil\else
1072 \fi
1073 \let\bbl@s@activate\bbl@activate
1074 \def\bbl@activate#1{% 
1075 \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{} }
1077 \let\bbl@s@deactivate\bbl@deactivate
1078 \def\bbl@deactivate#1{% 
1079 \fi
1080 \newcommand\ifbabelshorthand{\bbl@ifunset{\bbl@active@#1}{#3}{#2}}
1081 \def\bbl@prim@s{\prime\futurelet\@let@token\bbl@pr@m@s}
1082 \def\bbl@if@primes#1#2{\ifx#1\@let@token \else\ifx#2\@let@token \bbl@afterelse \else \bbl@afterfi \fi\fi}
1086 \let\bbl@s@activate\bbl@activate
1087 \def\bbl@activate#1{% 
1088 \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{} }
1091 \begingroup
1092 \catcode`\^=7 \catcode`*=\active \lccode`*=`\^\lowercase{% 
1094 \gdef\bbl@pr@m@s{% 
1096 \bbl@if@primes"% 
1097 \pr@@s
1099 \endgroup

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

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

1061 \def\bbl@pr@m@s{% 
1062 \pr@me\futurelet\@let@token\bbl@pr@m@s}
1063 \def\bbl@if@primes#1#2{% 
1067 \lowercase{% 
1069 \gdef\bbl@pr@m@s{% 
1071 \bbl@if@primes"% 
1072 \pr@@s
1074 \lowercase{% 
1078 \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{} }
1080 \newcommand\ifbabelshorthand[3]{\bbl@ifunset{\bbl@active@#1}{#3}{#2}}
1084 \ifx\@let@token
1086 \else\ifx\@let@token \bbl@afterelse \else \bbl@afterfi \fi\fi}
1090 \bbl@afterfi} 84

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

\bbl@activate{~}

\OT1dqpos \T1dqpos The position of the double quote character is different for the OT1 and T1 encodings. It will later be selected using the \f@encoding macro. Therefore we define two macros here to store the position of the character in these encodings.

\expandafter\def\csname OT1dqpos\endcsname{127}
\expandafter\def\csname T1dqpos\endcsname{4}

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

9.5 Language attributes

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

\languageattribute The macro \languageattribute checks whether its arguments are valid and then activates the selected language attribute. First check whether the language is known, and then process each attribute in the list.

\bbl@trace{Language attributes}
\newcommand\languageattribute[2]{%
\def\bbl@tempc{#1}%
\bbl@fixname\bbl@tempc
\bbl@iflanguage\bbl@tempc{%
\bbl@vforeach{#2}{%
Wewant to make sure that each attribute is selected only once; therefore we store the already selected attributes in \bbl@known@attrs. When that control sequence is not yet defined this attribute is certainly not selected before.

\ifx\bbl@known@attrs\@undefined
\in@false
\else
\bbl@xin@{,\bbl@tempc-##1,}{,\bbl@known@attrs,}%
\fi
\ifin@
\bbl@warning{%
You have more than once selected the attribute '##1'\%
for language #1. Reported}%
\else
\fi

\bbl@exp{%
\\bbl@add@list\\bbl@known@attrs{\bbl@tempc-##1}}%
\edef\bbl@tempa{\bbl@tempc-##1}%
\expandafter\bbl@ifknown@ttrib\expandafter{\bbl@tempa}\bbl@attributes%
{\csname\bbl@tempc @attr@##1\endcsname}%
{\@attrerr{\bbl@tempc}{##1}}%
\fi}}

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.

\bbl@exp{%
\\bbl@add@list\\bbl@known@attrs{\bbl@tempc-##1}}%
\edef\bbl@tempa{\bbl@tempc-##1}%
\expandafter\bbl@ifknown@ttrib\expandafter{\bbl@tempa}\bbl@attributes%
{\csname\bbl@tempc @attr@##1\endcsname}%
{\@attrerr{\bbl@tempc}{##1}}%
\fi}
This command should only be used in the preamble of a document.
\@onlypreamble\languageattribute

The error text to be issued when an unknown attribute is selected.
\newcommand*{\@attrerr}[2]{%
  {The attribute #2 is unknown for language #1.}%
  {Your command will be ignored, type <return> to proceed}}%

\bbl@declare@ttribute
This command adds the new language/attribute combination to the list of known attributes.
Then it defines a control sequence to be executed when the attribute is used in a document. The result of this should be that the macro \extras... for the current language is extended, otherwise the attribute will not work as its code is removed from memory at \begin{document}.
\def\bbl@declare@ttribute#1#2#3{%
  \bbl@xin@{,#2,}{,\BabelModifiers,}%
  \ifin@
    \AfterBabelLanguage(#1){\languageattribute(#1){#2}}%
  \fi
  \bbl@add@list\bbl@attributes{#1-#2}%
  \expandafter\def\csname#1@attr@#2\endcsname{#3}}%

\bbl@ifattributeset
This internal macro has 4 arguments. It can be used to interpret \TeX code based on whether a certain attribute was set. This command should appear inside the argument to \AtBeginDocument because the attributes are set in the document preamble, after babel is loaded.
The first argument is the language, the second argument the attribute being checked, and the third and fourth arguments are the true and false clauses.
\def\bbl@ifattributeset#1#2#3#4{%
  First we need to find out if any attributes were set; if not we're done.
  \ifx\bbl@known@attrs\@undefined\in@false\else\fi
  The we need to check the list of known attributes.
  \bbl@xin@{,#1-#2,}{,\bbl@known@attrs,}%
  \ifin@
    \bbl@afterelse#3%
  \else
    \bbl@afterfi#4%
  \fi
  \}

\bbl@ifknown@trib
An internal macro to check whether a given language/attribute is known. The macro takes 4 arguments, the language/attribute, the attribute list, the \TeX-code to be executed when the attribute is known and the \TeX-code to be executed otherwise.
\def\bbl@ifknown@trib#1#2{%
  We first assume the attribute is unknown.
  \let\bbl@tempa@secondoftwo
Then we loop over the list of known attributes, trying to find a match.

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

Finally we execute \bbl@tempa.

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

9.6 Support for saving macro definitions

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

\bbl@save The macro \bbl@save\langle csname\rangle saves the current meaning of the control sequence \langle csname\rangle to \originalTeX. To do this, we let the current meaning to a temporary control sequence, the restore commands are appended to \originalTeX and the counter is incremented.

\bbl@beginsave The initialization of a new save cycle: reset the counter to zero.

Before it’s forgotten, allocate the counter and initialize all.
The macro \b babel@savevariable\{variable\} saves the value of the variable. \{variable\} can be anything allowed after the \the primitive.

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

\textbf{9.7 Short tags}

\textbf{9.8 Hyphens}
be taken into account. Reported{\%}
\fi
\ifx\@empty#1{\%}
def\bbl@tempa{\%}
\bbl@fixname\bbl@tempa\%}
\bbl@iflanguage\bbl@tempa{\%}
de\bbl@tempa{\%}
\bbl@ifunset{bbl@hyphenation@\bbl@tempa}{\%}
def\bbl@allowhyphens{\%}
def\bbl@t@one{T1}\%}
def\allowhyphens{\%}
newcommand\babelnullhyphen{\%}
def\babelhyphen{\%}
def\bbl@hyphenchar{\%}
\ifnum\hyphenchar\font=\m@ne\%}
\babelnullhyphen\%}
def\bbl@usehyphen{\%}

\bbl@allowhyphens This macro makes hyphenation possible. Basically its definition is nothing more than \nobreak\hskip Opt plus Opt\%}
\def\bbl@allowhyphens{\%}
def\bbl@t@one{T1}\%}
def\allowhyphens{\%}
newcommand\babelnullhyphen{\%}
def\babelhyphen{\%}
def\bbl@hyphenchar{\%}
\ifnum\hyphenchar\font=\m@ne\%}
\babelnullhyphen\%}
def\bbl@usehyphen{\%}

\babelhyphen Macros to insert common hyphens. Note the space before @ in \babelhyphen. Instead of protecting it with \DeclareRobustCommand, which could insert a \relax, we use the same procedure as shorthands, with \active@prefix.\%}
def\babelnullhyphen{\%}
def\babelhyphen{\%}
def\bbl@hyphenchar{\%}
\ifnum\hyphenchar\font=\m@ne\%}
\babelnullhyphen\%}
def\bbl@usehyphen{\%}

The following two commands are used to wrap the “hyphen” and set the behavior of the rest of the word – the version with a single @ is used when further hyphenation is allowed, while that with @@ if no more 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.\%}
def\bbl@usehyphen{\%}
\leavevmode\%}
def\babelnullhyphen{\%}
def\babelhyphen{\%}
def\bbl@hyphenchar{\%}
\ifnum\hyphenchar\font=\m@ne\%}
\babelnullhyphen\%}
def\bbl@usehyphen{\%}

The following macro inserts the hyphen char.\%}
def\bbl@hyphenchar{\%}
def\babelnullhyphen{\%}
def\babelhyphen{\%}
def\bbl@hyphenchar{\%}

\TeX\% begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.
Finally, we define the hyphen “types”. Their names will not change, so you may use them in \dfl’s. After a space, the `\mbox` in `\bbl@hy@nobreak` is redundant.

```
1260 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}}
1261 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}}
1262 \def\bbl@hy@hard{\bbl@usehyphen{\bbl@hyphenchar}}
1263 \def\bbl@hy@hard{\bbl@usehyphen{\bbl@hyphenchar}}
1264 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}  \bbl@disc
1265 \def\bbl@hy@nobreak{\mbox{\bbl@hyphenchar}}
1266 \def\bbl@hy@repeat{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{}}}
1267 \def\bbl@hy@repeat{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{}}}
1268 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}  \bbl@disc
1269 \def\bbl@hy@repeat{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{}}}
```

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

```
1274 \def\bbl@disc#1#2{\nobreak\discretionary{#2-}{}{#1}\bbl@allowhyphens}
```

### 9.9 Multiencoding strings

The aim following commands is to provide a common interface for strings in several encodings. They also contains several hooks which can be used by \textsc{luatex} and \textsc{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.

```
1275 \bbl@trace{Multiencoding strings}
1276 \def\bbl@toglobal#1{\global#1\let#1#1}
1277 \def\bbl@recatcode#1{\@tempcnta="7F
1278 \def\bbl@tempa{\ifnum\@tempcnta="FF\else
1279 \catcode\@tempcnta=#1\relax
1280 \advance\@tempcnta\@ne
1281 \bbl@tempa
1282 \expandafter\bbl@tempa
1283 \fi}%
1284 \let\bbl@tolower\@empty\bbl@toupper\@empty
1285 \bbl@tempa}
```

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@uclc`. The parser is restarted inside `\langle\lang\rangle@bbl@uclc` because we do not know how many expansions are necessary (depends on whether strings are encoded). The last part is tricky – when uppercasing, we have:

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

and starts over (and similarly when lowercasing).

```
1286 \ifpagewith{babel}{nocase}%=\def\bbl@patchuclc{\relax}%
1287 {\let\bbl@patchuclc\relax}%
1288 {\def\bbl@patchuclc{%
```

90
The following package options control the behavior of \SetString.

\let\bbl@opt@strings@nil % accept strings=value
\DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
\DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
\def\BabelStringsDefault{generic}

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.
Parse the encoding info to get the label, input, and font parts. Select the behavior of \SetString. There are two main cases, depending of if there is an optional argument: without 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.
There are two versions of \bbl@scswitch. The first version is used when \ldfs are read, and it makes sure \(\langle\text{group}\rangle\langle\text{language}\rangle\) is reset, but only once (\bbl@screset is used to keep track of this). The second version is used in the preamble and packages loaded after babel and does nothing. The macro \bbl@forlang loops \bbl@L but its body is executed only if the value is in \BabelLanguages (inside babel) or \(\langle\text{language}\rangle\) is defined (after babel has been loaded). There are also two version of \bbl@forlang. The first one skips the current iteration if the language is not in \BabelLanguages (used in \ldfs), and the second one skips undefined languages (after babel has been loaded).

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 \bbl@encoded for string to be expanded in case transformations. It is \relax by default, but in \MakeUppercase and \MakeLowercase its value is a modified expandable \@changed@cmd.

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

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

\def\SetStringLoop##1##2{\def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}\count@\z@\bbl@loop\bbl@tempa{\#2}{% empty items and spaces are ok\advance\count@\@ne\toks@\expandafter{\bbl@tempa}\bbl@exp{\\SetString\bbl@temp1{\romannumeral\count@}{\the\toks@}}\count@{\the\count@{\relax}}}\count@\z@}

\def\SetStringLoop##1##2{\def#1{#2}}

\Delaying code Now the definition of \AfterBabelCommands when it is activated.
Case mapping  The command \SetCase provides a way to change the behavior of \MakeUppercase and \MakeLowercase. \bbl@tempa is set by the patched \@ucilclist to the parsing command.

```
\newcommand\SetCase[3][]{% 
\bbl@patchuclc 
\bbl@forlang\bbl@tempa{% 
\expandafter\bbl@encstring 
\csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}% 
\expandafter\bbl@encstring 
\csname\bbl@tempa @bbl@uc\endcsname{##2}% 
\expandafter\bbl@encstring 
\csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
```

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.

```
\newcommand\SetHyphenMap[1]{% 
\bbl@forlang\bbl@tempa{% 
\expandafter\stringdef 
\csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
```

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

```
\newcommand\BabelLower[2]{% one to one.
\ifnum\lccode#1=#2\else 
\babel@savevariable{\lccode#1}% 
\lccode#1=#2\relax
\fi}%
\newcommand\BabelLowerMM[4]{% many-to-many 
@tempcnta=#1\relax 
@tempcntb=#4\relax 
\def\bbl@tempa{% 
\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.

```
\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@@}
```
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

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

\def\set@low@box#1{\setbox\tw@hbox{,}\setbox\z@hbox{#1}\
  \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@\n  \setbox\z@hbox{\lower\dimen\z@ \box\z@}\ht\z@\ht\tw@ \dp\z@\dp\tw@}

The macro \save@sf@q is used to save and reset the current space factor.

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

9.11 Making glyphs available

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

9.11.1 Quotation marks

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

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

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

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

We also need the single quote character at the baseline.

\ProvideTextCommand{%\quotesinglebase}{OT1}{\save@sf@q{\set@low@box{\textquateright/}\n  \box\z@\kern-.04em\bbl@allowhyphens}}

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

\ProvideTextCommandDefault{%\quotesinglebase}{OT1}{\UseTextSymbol{OT1}{\quotesinglebase}}
The guillemet characters are not available in OT1 encoding. They are faked.

\g@llb{\provideTextCommand{\guillemotleft}{OT1}{%\ifmmode\ll\else\save@sf@q{\nobreak\raise.2ex\hbox{$\scriptscriptstyle\ll$}}\bbl@allowhyphens\%\fi}}
\g@llb{\provideTextCommand{\guillemotright}{OT1}{%\ifmmode\gg\else\save@sf@q{\nobreak\raise.2ex\hbox{$\scriptscriptstyle\gg$}}\bbl@allowhyphens\%\fi}}

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

\g@llb{\provideTextCommandDefault{\guillemotleft}{%\useTextSymbol{OT1}{\guillemotleft}}}\g@llb{\provideTextCommandDefault{\guillemotright}{%\useTextSymbol{OT1}{\guillemotright}}}

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

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

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

\g@llb{\provideTextCommandDefault{\guilsinglleft}{%\useTextSymbol{OT1}{\guilsinglleft}}}\g@llb{\provideTextCommandDefault{\guilsinglright}{%\useTextSymbol{OT1}{\guilsinglright}}}

9.11.2 Letters

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

\g@llb{\DeclareTextCommand{\ij}{OT1}{%i\kern-0.02em\bbl@allowhyphens j}}\g@llb{\DeclareTextCommand{\IJ}{OT1}{%I\kern-0.02em\bbl@allowhyphens J}}
\g@llb{\DeclareTextCommand{\ij}{T1}{\char188}}\g@llb{\DeclareTextCommand{\IJ}{T1}{\char156}}
Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.

\dj\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).

\SS  For the T1 encoding \SS is defined and selects a specific glyph from the font, but for other encodings it is not available. Therefore we make it available here.

\glq\grq  The ‘german’ single quotes.

9.11.3 Shorthands for quotation marks

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

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

The 'french' single guillemets.

The 'french' double guillemets.

9.11.4 Umlauts and tremas

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

`\umlaut{\umlauthigh}`

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

```latex
\umlaut{\umlauthigh} \umlaut{\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 ⟨dimen⟩ register.

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

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

\def\lower@umlaut#1{%
   \leavevmode\bgroup
   \U@D 1ex{\setbox\z@\hbox{\expandafter\char\csname f@encoding dqpos\endcsname}\
   \dimen@ -.45ex\advance\dimen@\ht\z@
   \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}\
   \expandafter\accent\csname f@encoding dqpos\endcsname\fontdimen5\font\U@D #1\%\
   \egroup}

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

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

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

\ifx\l@english\@undefined\chardef\l@english\z@\fi
\main@language{english}

9.12 Layout

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

9.13 Load engine specific macros

9.14 Creating languages

\babelprovide is a general purpose tool for creating and modifying languages. It creates the language infrastructure, and loads, if requested, an ini file. It may be used in conjunction to previously loaded ldf files.
% 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@chargroups\@nil
\bbl@forkv{#1}{% TODO - error handling
 \in@{/}{##1}%
 \ifin@
 \bbl@renewinikey##1@@{##2}%
 \else
 \bbl@csarg\def{KVP@##1}{##2}%
 \fi}%
% == import, captions ==
\ifx\bbl@KVP@import\@nil\else
 \bbl@exp{\bbl@ifblank{\bbl@KVP@import}}%
 \begingroup
 \def\BabelBeforeIni##1##2{\gdef\bbl@KVP@import{##1}\endinput}%
 \InputIfFileExists{babel-#2.tex}{}{}%
 \endgroup}%
\fi
\ifx\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@ifunset{\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}%
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. 

\def\bbl@setdigits#1#2#3#4#5{% 
\bbl@exp{% 
\def<\languagename digits>####1{% % ie, \langdigits 
\bbl@digits@\languagename####1\@nil} %
\def<\languagename counter>####1{% % ie, \langcounter 
$$$\\expandafter<\bbl@counter@\languagename>$$ $$\csname c@####1\endcsname$$} %
\def<\bbl@counter@\languagename>####1{% % ie, \bbl@counter@lang 
$$\expandafter<\bbl@digits@\languagename>$$ $$\number####1$$ \@nil}}% 
\def\bbl@tempa##1##2##3##4##5{% 
\bbl@exp{% Wow, quite a lot of hashes! :-( 
\def<\bbl@digits@\languagename>########1{% % ie, \bbl@digits@lang 
$$\ifx########1\@nil % ie, \bbl@digits@lang$$ 
$$\else$$ $$\ifx0########1#1%$$ $$\else$$ $$\ifx1########1#2%$$ $$\else$$ $$\ifx2########1#3%$$ $$\else$$ $$\ifx3########1#4%$$ $$\else$$ $$\ifx4########1#5%$$ $$\else$$ $$\ifx5########1##1%$$ $$\else$$ $$\ifx6########1##2%$$ $$\else$$ $$\ifx7########1##3%$$ $$\else$$ $$\ifx8########1##4%$$ $$\else$$ $$\ifx9########1##5%$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\else$$ $$\el...
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 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.libr`.

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@inisec), while \bbl@after@ini must be called explicitly after \bbl@read@ini if necessary.

\begin{verbatim}
def\bbl@iniwarning#1{% 
  \bbl@ifunset{bbl@@kv@identification.warning#1}{}% 
  {\bbl@warning{From babel-\@nameuse{bbl@lini@\languagename}.ini:\% 
               \@nameuse{bbl@@kv@identification.warning#1}\% 
               Reported }}}
\let\bbl@inikv@identification\bbl@inikv
\def\bbl@secpost@identification{% 
  \bbl@iniwarning{}% 
  \ifcase\bbl@engine 
  \bbl@iniwarning{.pdflatex}% 
  \or 
  \bbl@iniwarning{.lualatex}% 
  \or 
  \bbl@iniwarning{.xelatex}% 
  \fi 
  \bbl@exportkey{elname}{identification.name.english}{\{} 
  \bbl@exportkey{lname}{identification.name.opentype}{\{} 
  \bbl@exportkey{lbcp}{identification.tag.bcp47}{\{} 
  \bbl@exportkey{lotf}{identification.tag.opentype}{DFLT}% 
  \bbl@exportkey{esname}{identification.script.name}{\{} 
  \bbl@exportkey{sname}{identification.script.name.opentype}{\{} 
  \bbl@exportkey{sbcp}{identification.script.tag.bcp47}{\{} 
  \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}} 
\let\bbl@inikv@typography\bbl@inikv 
\let\bbl@inikv@characters\bbl@inikv 
\let\bbl@inikv@numbers\bbl@inikv 
\def\bbl@after@ini{% 
  \bbl@exportkey{lfthm}{typography.lefthyphenmin}{2}% 
  \bbl@exportkey{rgthm}{typography.righthyphenmin}{3}% 
  \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}{\{} 
  \bbl@exportkey{jstfy}{typography.justify}{w}% 
  \bbl@exportkey{chrng}{characters.ranges}{\{} 
  \bbl@exportkey{dgnat}{numbers.digits.native}{\} 
  \bbl@exportkey{rqtex}{identification.require.babel}{\} 
  \bbl@toglobal\bbl@savetoday 
  \bbl@toglobal\bbl@savedate}
\end{verbatim}

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.

\begin{verbatim}
\ifcase\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
\end{verbatim}

110
The auxiliary macro for captions define `<caption>name</caption>.

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

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

TODO. Remove copypaste pattern.

\ifcase\bibli@engine
  \bibli@csarg\def{inikv@date.gregorian}#1=#2\@@{% for defaults
  \bibli@inidate#1...\relax{#2}{}
  \bibli@csarg\def{inikv@date.islamic}#1=#2\@@{%
  \bibli@inidate#1...\relax{#2}{islamic}
  \bibli@csarg\def{inikv@date.hebrew}#1=#2\@@{%
  \bibli@inidate#1...\relax{#2}{hebrew}
  \bibli@csarg\def{inikv@date.persian}#1=#2\@@{%
  \bibli@inidate#1...\relax{#2}{persian}
  \bibli@csarg\def{inikv@date.indian}#1=#2\@@{%
  \bibli@inidate#1...\relax{#2}{indian}

  % eg: 1=months, 2=wide, 3=1, 4=dummy
  \def\bibli@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
    \bibli@trim@def\bibli@tempa{#1.#2}%
    \bibli@ifsamestring{\bibli@tempa}{months.wide}% to savedate
    {\bibli@trim\def\bibli@tempa{#3}%
      \bibli@trim\toks@{#5}%
    }{\bibli@exp{% 
      \bibli@add\bibli@savestrings{% 
      \SetString\langle\bibli@tempa name>{\the\toks@}}} }}
  \ifcase\bibli@engine\let\bibli@savedate\@empty\fi
\fi

Dates will require some macros for the basic formatting. They may be redefined by language, so “semi-public” names (camel case) are used. Oddly enough, the CLDR places particles like “de” inconsistently in either in the date or in the month name.
Language and Script values to be used when defining a font or setting the direction are set with the following macros.

```
\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]{{\ifnum#1<10 0\fi\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\number#1 %\else\ifnum#1<100 \number#1 %\else\ifnum#1<1000 \expandafter\@gobble\number#1 %\else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %\else\bbl@error{Currently two-digit years are restricted to the\range 0-9999.} % There is little you can do. Sorry.}%\fi\fi\fi\fi}}%
\newcommand\BabelDateyyyy[1]{{\number#1}} % FIXME - add leading 0
\def\bbl@replace@finish@iii#1{\bbl@exp{\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}}% Note after \bbl@replace \toks@ contains the resulting string.
% TODO - Using this implicit behavior doesn't seem a good idea.
\bbl@replace@finish@iii\bbl@toreplace
```

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

```
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}\
    {See the manual for details.}}\
  {\@nameuse{bbl@\csname bbl@info@#1\endcsname \languagename}}}
```

10 Adjusting the Babel bahavior

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

```
\newcommand\babeladjust[1]{% TODO. Error handling.
  \bbl@forkv{#1}{\@nameuse{bbl@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.}}}
```

```bash
\edef\bbl@adjust@lua@#1\#2{%
  \ifvmode
    \ifnum\currentgrouplevel=\z@
      \directlua{ Babel.#2 }%
      \expandafter\expandafter\expandafter\@gobble
    \fi
  \fi
  \f{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.}}}
```
2380 \bbl@adjust@lua{bidi\{bidi\_enabled=true\}}
2381 \@namedef{bbl\@ADJ@bidi\_text\off}{%}
2382 \bbl@adjust@lua{bidi\{bidi\_enabled=false\}}
2383 \@namedef{bbl\@ADJ@bidi\_mapdigits\on}{%}
2384 \bbl@adjust@lua{bidi\{digits\_mapped=true\}}
2385 \@namedef{bbl\@ADJ@bidi\_mapdigits\off}{%}
2386 \bbl@adjust@lua{bidi\{digits\_mapped=false\}}
2387 %
2388 \@namedef{bbl\@ADJ@linebreak\_sea\on}{%}
2389 \bbl@adjust@lua{linebreak\{sea\_enabled=true\}}
2390 \@namedef{bbl\@ADJ@linebreak\_sea\off}{%}
2391 \bbl@adjust@lua{linebreak\{sea\_enabled=false\}}
2392 \@namedef{bbl\@ADJ@linebreak\_cjk\on}{%}
2393 \bbl@adjust@lua{linebreak\{cjk\_enabled=true\}}
2394 \@namedef{bbl\@ADJ@linebreak\_cjk\off}{%}
2395 \bbl@adjust@lua{linebreak\{cjk\_enabled=false\}}
2396 %
2397 \def\bbl@adjust@layout\#1{%
2398 \ifvmode
2399   #1%
2400 \expandafter\@gobble
2401 \fi
2402 {\bbl@error % The error is gobbled if everything went ok.}
2403 {\text{Currently, layout related features can be adjusted only in vertical mode.}}%
2404 {\text{Maybe things change in the future, but this is what it is.}}}
2405 \@namedef{bbl\@ADJ@layout\_tabular\on}{%}
2406 \bbl@adjust@layout{\let@tabular\bbl@NL@\tabular}
2407 \@namedef{bbl\@ADJ@layout\_tabular\off}{%}
2408 \bbl@adjust@layout{\let@tabular\bbl@OL@\tabular}
2409 \@namedef{bbl\@ADJ@layout\_lists\on}{%}
2410 \bbl@adjust@layout{\let@list\bbl@NL@\list}
2411 \@namedef{bbl\@ADJ@layout\_lists\off}{%}
2412 \@namedef{bbl\@ADJ@\hyphenation\_extra\on}{%}
2413 \bbl@activateposthyphen
2414 \bbl@adjust@layout{\let@list\bbl@OL@\list}
2415 \@namedef{bbl\@ADJ@\hyphenation\_extra\off}{%}

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.

2416 {\def\format{lplain}
2417 \ifx\fmtname\format
2418 else
2419 \def\format{LaTeX2e}
2420 \ifx\fmtname\format
2421 else
2422 \aftergroup\endinput
2423 \fi
2424 \fi}
11.2 Cross referencing macros

The \LaTeX\ book states:

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

When the above quote should still be true when a document is typeset in a language that has active characters, special care has to be taken of the category codes of these characters when they appear in an argument of the cross referencing macros. When a cross referencing command processes its argument, all tokens in this argument should be character tokens with category ‘letter’ or ‘other’.

The only way to accomplish this in most cases is to use the trick described in the \TeX\book [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#1{\B} \expandafter\meaning\A \expandafter\meaning\A \expandafter\meaning\A

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.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

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

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

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

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

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

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.

\newlabel\ ATEX-internal macro \newlabel. This is needed because we need to make sure that shorthand characters expand to their non-active version.
Now that we made sure that \@testdef still has the same definition we can rewrite it. First we make the shorthands ‘safe’.

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

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

\begin{verbatim}
  \expandafter\let\expandafter\bbl@tempa\csname #1@#2\endcsname

  Then we define \bbl@tempb just as \@newl@bel does it.
\end{verbatim}

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

\begin{verbatim}
  \ifx\bbl@tempa\relax
  \else
    \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}\fi

  We do the same for \bbl@tempb.
\end{verbatim}

\begin{verbatim}
  \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}\fi

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

\begin{verbatim}
  \ifx\bbl@tempa\bbl@tempb
  \else
    \@tempswatrue
  \fi
\end{verbatim}

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

\begin{verbatim}
\begin{verbatim}
  \edef\@tempa{#2}\org@@citex[\@tempa]
\end{verbatim}

  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.
\end{verbatim}
Notice that we use \texttt{\def} here instead of \texttt{\bbl@redefine} because \texttt{\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 \texttt{\@citex}, so PR4087 doesn't seem fixable in a simple way. Just load natbib before.)

\begin{verbatim}
def\@citex[#1][#2]{\@safe@activestrue\edef\@tempa{#3}\@safe@activesfalse\org@@citex[#1][#2]{\@tempa}}
\end{verbatim}

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

\begin{verbatim}
def\@citex[#1]{\@safe@activestrue\org@@citex[#1]{\@tempa}\@safe@activesfalse}
\end{verbatim}

\texttt{\nocite} The macro \texttt{\nocite} which is used to instruct Bi\TeX to extract uncited references from the database.

\begin{verbatim}\bbll@redefine\nocite#1{% \@safe@activestrue\org@nocite{#1}\@safe@activesfalse}\end{verbatim}

\texttt{\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 \texttt{\@safe@activesfalse} is in effect. This switch needs to be reset inside the \texttt{\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 \bbll@cite@choice to select the proper definition for \bibcite. This new definition is then activated.

\begin{verbatim}\bbll@redefine\bibcite{% \bbll@cite@choice\bibcite\end{verbatim}

\texttt{\bbll@bibcite} The macro \texttt{\bbll@bibcite} holds the definition of \bibcite needed when neither natbib nor cite is loaded.

\begin{verbatim}\def\bbll@bibcite#1#2{% \org@bibcite{#1}{\@safe@activesfalse#2}}\end{verbatim}

\texttt{\bbll@cite@choice} The macro \texttt{\bbll@cite@choice} determines which definition of \bibcite is needed. First we give \bibcite its default definition.

\begin{verbatim}\def\bbll@cite@choice{% \global\let\bibcite\bbll@bibcite\end{verbatim}

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

\begin{verbatim}@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}\% \ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}\%\end{verbatim}

Make sure this only happens once.

\begin{verbatim}\global\let\bbll@cite@choice\relax\end{verbatim}
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.

\begin{verbatim}
\AtBeginDocument{\bbl@cite@choice}
\@bibitem
One of the two internal \LaTeX macros called by \bibitem that write the citation label on the .aux file.
\end{verbatim}

\begin{verbatim}
\bbl@redefine\@bibitem#1{\
\@safe@activestrue\org@@bibitem{#1}\@safe@activesfalse}
\end{verbatim}

\begin{verbatim}
\let\org@nocite\nocite
\let\org@@citex\@citex
\let\org@bibcite\bibcite
\let\org@@bibitem\@bibitem
\fi
\end{verbatim}

\subsection{Marks}
\begin{verbatim}
\markright
\def\bbl@trace{Marks}
\IfBabelLayout{sectioning}{
{\ifx\bbl@opt@headfoot\@nnil
\g@addto@macro\@resetactivechars{%
\protect\selectlanguage\expandafter{\bbl@main@language}\
\let\protect\noexpand\edef\thepage{\
\noexpand\babelsublr\unexpanded{\the\thepage}}}%
\fi}
{\if\bbl@single\else
\bbl@ifunset{markright }\bbl@redefine\bbl@redefinerobust
\markright#1{\
\bbl@ifblank{#1}{\org@markright{}}{\toks@{#1}\
\bbl@exp{\
\org@markright{\protect\\bbl@restore@actives\the	oks@}}}}%}
\fi}
{\if\bbl@single\else
\bbl@ifunset{markright }\bbl@redefine\bbl@redefinerobust
\markright#1{%
\bbl@ifblank{#1}{\org@markright{}}{\toks@{#1}\
\bbl@exp{%\org@markright{\protect{\language}\language{\ languagename}{\protect{\language}}}\
\org@markright{\foreignlanguage{\languagename}{\protect{\language}}}}}%}
\fi}
\end{verbatim}

\begin{verbatim}
\markboth
The definition of \markboth is equivalent to that of \markright, except that we need two token registers. The document classes report and book define and set the headings for the page. While doing so they also store a copy of \markboth in \@mkboth. Therefore we need to check whether \@mkboth has already been set. If so we need to do that again with the new definition of \markboth. (As of Oct 2019, \LaTeX stores the definition in an intermediate macro, so it's not necessary anymore, but it's preserved for older versions.)
\end{verbatim}

\begin{verbatim}
\if\@mkboth\markboth
\def\bbl@tempc{\let@mkboth\markboth}
\else
\def\bbl@tempc{}
\fi
\end{verbatim}
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}
\bbl@trace{Preventing clashes with other packages}
\bbl@xin{(R)}\bbl@opt@safe
\ifin@
  \AtBeginDocument{%
    @ifpackage{loaded}{ifthen}{% 

    Then we can redefine \ifthenelse:

\begin{verbatim}
\bbl@redefine@long\ifthenelse#1#2#3{%
\end{verbatim}

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\bbl@temp@pref\pageref
\let\pageref\org@pageref
\let\bbl@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@actives@true
\org@ifthenelse{#1}%
\let\pageref\bbl@temp@pref
\let\ref\bbl@temp@ref
%@safe@actives@false
\end{verbatim}

119
11.4.2 varioref

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

\AtBeginDocument{%\ifpackageloaded{varioref}{%\bbl@redefine\@@vpageref#1[#2]#3{%\@safe@activestrue\org@@@vpageref{#1}[#2]{#3}\@safe@activesfalse}%\bbl@redefine\vrefpagenum#1#2{%\@safe@activestrue\org@vrefpagenum{#1}{#2}\@safe@activesfalse}%\}}}{}}

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

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

11.4.3 hhline

\hhline Delaying the activation of the shorthand characters has introduced a problem with the hhline package. The reason is that it uses the `:` character which is made active by the french support in babel. Therefore we need to reload the package when the `:` is an active character.

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

\AtEndOfPackage{%\AtBeginDocument{%\ifpackageloaded{hhline}%

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

{\expandafter\ifx\csname normal@char\endcsname: normal@char: \string: \endcsname: \relax \else

In that case we simply reload the package. Note that this happens after the category code of the @-sign has been changed to other, so we need to temporarily change it to letter again.

\makeatletter
\def@currname{hhline}\input{hhline.sty}\makeatother
{}}}
11.4.4 hyperref

\pdfstringdefDisableCommands

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

\AtBeginDocument{%
\ifx\pdfstringdefDisableCommands\@undefined\else
\pdfstringdefDisableCommands{\languageshorthands{system}}\fi%
}

11.4.5 fancyhdr

\FOREIGNLANGUAGE

The package fancyhdr treats the running head and footer 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.

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

\substitutefontfamily

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

\def\substitutefontfamily#1#2#3{%\lowercase{\immediate\openout15=#1#2.fd\relax}%\immediate\write15{\string\ProvidesFile{#1#2.fd}[^\the\year/\two@digits{\the\month}/\two@digits{\the\day}]\string\DeclareFontFamily{#1}{#2}{}\string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{}\string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{}\string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{}\string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{}\string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{}\string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{}\string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{}\string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{}%}
\closeout15}

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

\@onlypreamble{\substitutefontfamily

11.5 Encoding and fonts

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

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

\ensureasciie

\bbl@trace{Encoding and fonts}
\texttt{\newcommand{\BabelNonASCII}{LGR,X2,OT2,OT3,OT6,LHE,LW\textsc{n},LMA,LMC,LMS,LMU,PU,PD1}}
\texttt{\newcommand{\BabelNonText}{TS1,T3,TS3}}
\texttt{\let\org@TeX\TeX}
\texttt{\let\org@LaTeX\LaTeX}
\texttt{\let\ensureascii\@firstofone}
\texttt{\AtBeginDocument{}}
\texttt{\in@false}
\texttt{\bbl@foreach\BabelNonASCII{% is there a text non-ascii enc?}}
\texttt{\ifin@else}
\texttt{\lowercase{\bbl@xin@{,#1enc.def,}{,@filelist,}}}\texttt{%}
\texttt{\fi}\
\texttt{\ifi\% if a text non-ascii has been loaded}
\texttt{\def\ensureascii#1{{\fontencoding{OT1}\selectfont#1}}}\texttt{%}
\texttt{\DeclareTextCommandDefault\TeX\{\org@TeX\}}\texttt{%}
\texttt{\DeclareTextCommandDefault\LaTeX\{\org@LaTeX\}}\texttt{%}
\texttt{\def\bbl@tempc#1@\@{\uppercase\bbl@tempc#1@ENC.DEF\empty@@}}\texttt{%}
\texttt{\def\bbl@tempc@ENC.DEF#2@@{}}\texttt{%}
\texttt{\iffalse\empty@2\else}
\texttt{\bbl@ifunset{T@#1}@%}
\texttt{\%}
\texttt{\{\bbl@xin@{,#1,}{,\BabelNonASCII,\BabelNonText,},%}
\texttt{\ifi\%}
\texttt{\DeclareTextCommand\TeX\{#1\}{\ensureascii{\org@TeX}}}\texttt{%}
\texttt{\DeclareTextCommand\LaTeX\{#1\}{\ensureascii{\org@LaTeX}}}\texttt{%}
\texttt{\else}
\texttt{\def\ensureascii#1{{\fontencoding{#1}\selectfont#1}}}\texttt{%}
\texttt{\fi}\
\texttt{\bbl@foreach\@filelist{\bbl@tempb#1@\@@}% TODO - @@ de mas??}
\texttt{\bbl@xin@{,\cf@encoding,}{,\BabelNonASCII,\BabelNonText,},%}
\texttt{\ifi\%}
\texttt{\edef\ensureascii#1{{\noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}}}\texttt{%}
\texttt{\fi}\
\texttt{\fi}

Now comes the old deprecated stuff (with a little change in 3.9l, for fontspec). The first thing we need to do is to determine, at \begin{document}, which latin fontencoding to use.\texttt{\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.\texttt{\AtEndOfPackage{\edef\latinencoding{\cf@encoding}}}\texttt{\%}

But this might be overruled with a later loading of the package fontenc. Therefore we check at the execution of \begin{document} whether it was loaded with the T1 option. The normal way to do this (using @ifpackageloaded) is disabled for this package. Now we have to revert to parsing the internal macro @filelist which contains all the filenames loaded.\texttt{\AtBeginDocument{}}
\texttt{\@ifpackageloaded{fontspec}{%}
\texttt{\{\xdef\latinencoding{%}
\texttt{\iffx\UTFencname\undefined}
\texttt{\EU\iffcase\bbl@engine\or2\or1\fi}
\texttt{\else}
\texttt{\UTFcase\bbl@engine\or2\or1\fi\}}}\texttt{\%}
\texttt{\{\gdef\latinencoding{OT1}{}}}\texttt{\%}
Then we can define the command \latintext which is a declarative switch to a Latin font-encoding. Usage of this macro is deprecated.

\DeclareRobustCommand\latintext{\fontencoding{\latinencoding}\selectfont\def\encodingdefault{\latinencoding}}

\textlatin This command takes an argument which is then typeset using the requested font encoding. In order to avoid many encoding switches it operates in a local scope.

11.6 Basic bidi support

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

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

- pdftex provides a minimal support for bidi text, and it must be done by hand. Vertical typesetting is not possible.
- xetex is somewhat better, thanks to its font engine (even if not always reliable) and a few additional tools. However, very little is done at the paragraph level. Another challenging problem is text direction does not honour \TeX grouping.
- luatex can provide the most complete solution, as we can manipulate almost freely the node list, the generated lines, and so on, but bidi text does not work out of the box and some development is necessary. It also provides tools to properly set left-to-right and right-to-left page layouts. As Lua\TeX-ja shows, vertical typesetting is possible, too.
The following command is executed only if there is a right-to-left script (once). It activates the `\everypar` hack for xetex, to properly handle the par direction. Note text and par dirs are decoupled to some extent (although not completely).

A tool for weak \(L\) (mainly digits). We also disable warnings with hyperref.

11.7 Local Language Configuration

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

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

\loadlocalcfg

\DeclareRobustCommand\babelsublr[1]{\leavevmode{\bbl@textdir\z@#1}}
\AtBeginDocument{%}
\ifx\pdfstringdefDisableCommands\undefined\else
\pdfstringdefDisableCommands\relax\else
\pdfstringdefDisableCommands\@firstofone\fi
\fi
\fi

\loadlocalcfg

\trace{Local Language Configuration}
Just to be compatible with \TeX 2.09 we add a few more lines of code:
\ifx\@unexpandable@protect\@undefined
\def\@unexpandable@protect{\noexpand\protect\noexpand}
\long\def\protected@write#1#2#3{% 
\begingroup
\let\thepage\relax
#2%
\let\protect\@unexpandable@protect
\edef\reserved@a{\write#1{#3}}%
\reserved@a
\endgroup
\if@nobreak\ifvmode
\nobreak\fi\fi}
\fi
⟨/core⟩
⟨∗kernel⟩

12 Multiple languages (switch.def)

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.

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

\adddialect
The macro \adddialect can be used to add the name of a dialect or variant language, for which an already defined hyphenation table can be used.
\def\bbl@version{⟨⟨version⟩⟩}
\def\bbl@date{⟨⟨date⟩⟩}
\def\adddialect#1#2{% 
\global\chardef#1#2\relax
\bbl@usehooks{adddialect}{{#1}{#2}}%
\begingroup
\count@#1\relax
\bbl@elt##1##2##3##4{% 
\ifnum\count@=##2\relax
\bbl@info{\string#1 = using hyphenrules for ##1\% (\string\language\the\count@)}%
\bbl@elt####1####2####3####4{}}%
\fi}
\@nameuse{bbl@languages}%
\endgroup}
\bbl@iflanguage executes code only if the language l@ exists. Otherwise raises and error. The argument of \bbl@fixname has to be a macro name, as it may get “fixed” if casing (lc/uc) is wrong. It’s intended to fix a long-standing bug when \foreignlanguage and the like appear in a \MakeXXXcase. However, a lowercase form is not imposed to improve
backward compatibility (perhaps you defined a language named MYLANG, but unfortunately mixed case names cannot be trapped). Note @l is encapsulated, so that its case does not change.

\def\bbl@fixname#1{%  
\begingroup  
\def\bbl@tempe{l@}  
\edef\bbl@tempd{\noexpand\@ifundefined{\noexpand\bbl@tempe#1}}  
\bbl@tempd  
{\lowercase\expandafter{\bbl@tempd}  
{\uppercase\expandafter{\bbl@tempd}  
\@empty  
{\edef\bbl@tempd{\def\noexpand#1{#1}}  
{\uppercase\expandafter{\bbl@tempd}}}  
\@empty  
{\edef\bbl@tempd{\endgroup\def\noexpand#1{#1}}}}  
\bbl@tempd}

\def\bbl@iflanguage#1{%  
\@ifundefined{l@#1}{\@nolanerr{#1}\@gobble}\@firstofone

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

\def\iflanguage#1{%  
\bbl@iflanguage{#1}{%  
\ifnum\csname l@#1\endcsname=\language  
\expandafter\@firstoftwo  
\else  
\expandafter\@secondoftwo  
\fi}}

\selectlanguage The macro \selectlanguage checks whether the language is already defined before it performs its actual task, which is to update \language and activate language-specific definitions.

To allow the call of \selectlanguage either with a control sequence name or with a simple string as argument, we have to use a trick to delete the optional escape character. To convert a control sequence to a string, we use the \string primitive. Next we have to look at the first character of this string and compare it with the escape character. Because this escape character can be changed by setting the internal integer \escapechar to a character number, we have to compare this number with the character of the string. To do this we have to use \TeX's backquote notation to specify the character as a number. If the first character of the \string'ed argument is the current escape character, the comparison has stripped this character and the rest in the ‘then’ part consists of the rest of the control sequence name. Otherwise we know that either the argument is not a control sequence or \escapechar is set to a value outside of the character range 0–255.

If the user gives an empty argument, we provide a default argument for \string. This argument should expand to nothing.

\let\bbl@select@type\z@  
\edef\selectlanguage{%  
\noexpand\protect  
\expandafter\noexpand\csname selectlanguage \endcsname}
Because the command `\selectlanguage` could be used in a moving argument it expands to `\protect\selectlanguage`. Therefore, we have to make sure that a macro `\protect` exists. If it doesn’t it is `\let` to `\relax`.

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.

`\bbl@pop@language` 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:

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.

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.
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@last(0) % No real need for a new counter
\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@nameuse{bbl@id@@languagename}}

The unprotected part of \selectlanguage.

\def\BabelContentsFiles{toc,lof,lot}
def\bbl@set@language{#1} {% from selectlanguage, pop@
def\languagename{%
\ifnum\escapechar=\expandafter`\string\@empty
\else\string\@empty\fi}%
\select@language{\languagename}%
% write to auxs
\expandafter\ifx\csname date\languagename\endcsname\relax\else
\if@filesw
\ifx\babel@aux\@gobbletwo\else % Set if single in the first, redundant
\protected@write\@auxout{}{\string\babel@aux{\languagename}{}}%\fi
\fi
\bbl@usehooks{write}{}%
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 \textbackslash lefthyphenmin and \textbackslash righthyphenmin is somewhat different. First we save their current values, then we check if \textbackslash lang\textbackslash hyphenmins is defined. If it is not, we set default values (2 and 3), otherwise the values in \textbackslash lang\textbackslash hyphenmins will be used.

\newif\iifbbl@usedategroup
\def\bbl@switch#1{% from select@, foreign@
  % make sure there is info for the language if so requested
  \bbl@ensureinfo(#1)%
  % restore
  \originalTeX
  \expandafter\def\expandafter\originalTeX\expandafter{%
  \csname noextras#1\endcsname
  \let\originalTeX@empty
  \bbl@beginsave}%
  \bbl@usehooks{afterreset}{%}
  \languageshorthands{none}%
  % set the locale id
  \bbl@id@assign
  % switch captions, date
  \ifcase\bbl@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
    \ifbbl@usedategroup % if foreign... within \textbackslash lang\textbackslash date
      \bbl@usedategroupfalse
    \fi
    \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
  % switch extras
  \bbl@usehooks{beforeextras}{}}%
  \csname extras#1\endcsname\relax
  \bbl@usehooks{afterextras}{}}%
  % babel-ensure
  % babel-sh-<short>
  % babel-bidi
  % babel-fontspec
  % hyphenation - case mapping
  \ifcase\bbl@opt@hyphenmap
    \def\BabelLower##1##2{\lccode##1=##2}\relax
    \ifnum\bbl@hymapsel>4
      \csname\languagename @bbl@hyphenmap\endcsname
    \fi
    \chardef\bbl@opt@hyphenmap\z@
  \else
    \chardef\bbl@opt@hyphenmap\z@
  \fi
otherlanguage The otherlanguage environment can be used as an alternative to using the \selectlanguage declarative command. When you are typesetting a document which mixes left-to-right and right-to-left typesetting you have to use this environment in order to let things work as you expect them to. The \ignorespaces command is necessary to hide the environment when it is entered in horizontal mode.

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

\foreignlanguage The \foreignlanguage command is another substitute for the \selectlanguage command. This command takes two arguments, the first argument is the name of the language to use for typesetting the text specified in the second argument. Unlike \selectlanguage this command doesn't switch everything, it only switches the hyphenation rules and the extra definitions for the language specified. It does this within a group and assumes the \extras{lang} command doesn't make any \global changes. The coding is very similar to part of \selectlanguage.\bblbefore\foreign is a trick to fix a bug in bidi texts. \foreignlanguage is supposed to be a 'text' command, and therefore it must emit a \leavevmode, but it does not, and therefore the indent is placed on the opposite margin. For backward compatibility, however, it is done only if a right-to-left script is requested; otherwise, it is no-op.
(3.11) \foreignlanguage* is a temporary, experimental macro for a few lines with a
different script direction, while preserving the paragraph format (thank the braces around
\par, things like \hangindent are not reset). Do not use it in production, because its
semantics and its syntax may change (and very likely will, or even it could be removed
altogether). Currently it enters in vmode and then selects the language (which in turn sets
the paragraph direction).

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

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

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

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 have been set, too). \bbl@hyphenation@ is set to relax until the very first \babelhyphenation, so do nothing with this value. If the exceptions for a language (by its number, not its name, so that :ENC is taken into account) has been set, then use \hyphenation with both global and language exceptions and empty the latter to mark they must not be set again.

\let\bbl@patterns\@empty
\let\bbl@hyphenation@\relax
\let\bbl@pttnlist\@empty
\let\bbl@patterns\relax
\let\bbl@hyphlist\@empty
\def\bbl@patterns#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
  \@expandtwoargs\bbl@usehooks{patterns}{{#1}{\bbl@tempa}}%
  % > luatex
  @ifundefined{bbl@hyphenation@}{}{% Can be \relax!
    \begingroup
      \bbl@xin@,\number\language,}{,\bbl@hyphlist}%
    \ifin@else
      @expandtwoargs\bbl@usehooks{hyphenation}{{#1}{\bbl@tempa}}%
      \hyphenation(%
        \bbl@hyphenation@
        \@ifundefined{bbl@hyphenation@}{% Can be \relax!
          \empty
          {\space\csname bbl@hyphenation@\#1\endcsname}%
          \edef\bbl@hyphlist{\bbl@hyphlist\number\language,}%
          \fi
        }
      \endgroup}
    }

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

\def\hyphenrules#1{%
  \edef\bbl@tempf{#1}%
  \bbl@fixname\bbl@tempf
  \bbl@iflanguage\bbl@tempf{%
    \expandafter\bbl@patterns\expandafter{\bbl@tempf}%
    \languageshorthands{none}%
    \expandafter\if\csname bbl@tempf hyphenmins\endcsname\relax
    \set\hyphenmins\tw@\thr@\relax
    \else
The macro \providehyphenmins should be used in the language definition files to provide a default setting for the hyphenation parameters \lefthyphenmin and \righthyphenmin. If the macro \langle lang\rangle hyphenmins is already defined this command has no effect.

\providehyphenmins

\set@hyphenmins

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

\ProvidesLanguage

The identification code for each file is something that was introduced in \LaTeX2ε. 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.

\ifx\originalTeX@undefined\let\originalTeX@empty\fi
Because this part of the code can be included in a format, we make sure that the macro which initialises the save mechanism, \bbl@beginsave, is not considered to be undefined.

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

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

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

12.2 Errors

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

When the package was loaded without options not everything will work as expected. An error message is issued in that case.
When the format knows about \PackageError it must be \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.

\edef\bbl@nulllanguage{\string\language=0}
\if\PackageError\@undefined
\def\bbl@error#1#2{%
  \begingroup
  \newlinechar=`\^^J
  \def\{\^^J(babel) }
  \errhelp{#2}\errmessage{\#1}%
  \endgroup}
\def\bbl@warning#1{%
  \begingroup
  \newlinechar=`\^^J
  \message{\#1}%
  \endgroup}
\let\bbl@infowarn\bbl@warning
\def\bbl@info#1{%
  \begingroup
  \newlinechar=`\^^J
  \wlog{#1}%
  \endgroup}
\else
\def\bbl@error#1#2{%
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{verbatim}
\let\org\everyjob\everyjob
\def\everyjob#1{%
\org\everyjob{#1}%
\org\everyjob\expandafter{\the\org\everyjob\immediate\write16{%
hyphenation patterns for \the\loaded@patterns loaded.}}%
\let\everyjob\org\everyjob\let\org\everyjob\@undefined}
\end{verbatim}

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

There are some problems with this approach though.

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

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

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

To make sure that \TeX{} 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{\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}[\date{} \version{} 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
\end{verbatim}

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

\def\process@line#1#2 #3 #4 {%
\ifx=#1%
\process@synonym{#2}%
\else
\process@language{#1#2}{#3}{#4}%
\fi
\ignorespaces}

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

When no languages have been loaded yet, the name following the = will be a synonym for hyphenation register 0. So, it is stored in a token register and executed when the first pattern file has been processed. (The \relax just helps to the \if below catching synonyms without a language.) Otherwise the name will be a synonym for the language loaded last. We also need to copy the hyphenmin parameters for the synonym.
\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\rangle \hyphenmins macro. When no assignments were made we provide a default setting.
Some pattern files contain changes to the \lccode and \uccode arrays. Such changes should remain local to the language; therefore we process the pattern file in a group; the \patterns command acts globally so its effect will be remembered.
Then we globally store the settings of \lefthyphenmin and \righthyphenmin and close the group.
When the hyphenation patterns have been processed we need to see if a file with hyphenation exceptions needs to be read. This is the case when the third argument is not
empty and when it does not contain a space token. (Note however there is no need to save
hyphenation exceptions into the format.)
\bbl@languages saves a snapshot of the loaded languages in the form
\bbl@elt{⟨language-name⟩}{⟨number⟩}{⟨patterns-file⟩}{⟨exceptions-file⟩}. Note the last
2 arguments are empty in ‘dialects’ defined in language.dat with =. Note also the
language name can have encoding info.
Finally, if the counter \language is equal to zero we execute the synonyms stored.

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

\bbl@get@enc \bbl@hyph@enc
```

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

```
\def\bbl@get@enc#1:#2:#3\@@@{
  \def\bbl@hyph@enc{#2}}
```

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

```
\def\bbl@hook@everylanguage{\}
\def\bbl@hook@loadpatterns{\}
\let\bbl@hook@loadexceptions=\bbl@hook@loadpatterns
\let\bbl@hook@loadkernel=\bbl@hook@loadpatterns
\begingroup
  \def\AddBabelHook#1#2{%
    \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
      \def\next{\toks1}%
    \else
      \expandafter\expandafter\expandafter\AddBabelHook
      \csname bbl@hook@#2\endcsname
    \fi
  \fi
\endgroup
```

\bbl@get@enc \bbl@hyph@enc

141
The configuration file can now be opened for reading.

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

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 $-1$.

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

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

If the file has reached its end, exit from the loop here. If not, empty lines are skipped. Add 3 space characters to the end of \bbll@line. This is needed to be able to recognize the arguments of \process@line later on. The default language should be the very first one.
Check for the end of the file. We must reverse the test for `\ifeof` without `\else`. Then reactivate the default patterns.

```latex
\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
\fi
```

and close the configuration file.

```latex
\closein1
```

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

```latex
\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.

```latex
\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
\fi
```

Here the code for init\TeX ends.

### 14 Font handling with fontspec

Add the bidi handler just before luaoffload, 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].

```latex
\f@ifodd\bbl@engine
\DeclareOption{bidi=basic-r}%
  {\ExecuteOptions{bidi=basic}}
\ExecuteOptions{bidi=basic}%
```

% TODO - to locale_props, not as separate attribute
```latex
\newattribute\bbl@attr@dir
```

% I don't like it, hackish:
```latex
\frozen@everymath\expandafter{% 
  \expandafter\bbl@mathboxdir\the\frozen@everymath}%
\frozen@everydisplay\expandafter{% 
  \expandafter\bbl@mathboxdir\the\frozen@everydisplay}%
```
With explicit languages, we could define the font at once, but we don’t. Just wait and see if the language is actually activated. bbl@font replaces hardcoded font names inside \..family by the corresponding macro \..default.
If the family in the previous command does not exist, it must be defined. Here is how:

```latex
\def\bbl@providefam#1{% \\
  \newcommand\#1default{}% Just define it \\
  \bbl@add@list\bbl@font@fams\#1%
} %
\DeclareRobustCommand\#1family{% \\
  \not@math@alphabet\#1family\relax \\
  \fontfamily\#1default\selectfont} \\
\DeclareTextFontCommand{\text#1}{\#1family}
```

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

```latex
\def\bbl@nostdfont#1{% \\
  \bbl@ifunset\bbl@lsys\languagename{%\bbl@provide@lsys\{\languagename\}}% \\
  \expandafter\bbl@ifblank\expandafter{\bbl@tempa}{% ie bbl@rmdflt@lang / *scrt \\
    \bbl@csarg\edef\bbl@tempb{#1}{#2}}% save bbl@rmdflt@ & \\
  \bbl@exp{% \\
    \let\bbl@bblfont\bbl@tempb family >%
  }% For the default font, just in case:
} %
\def\bbl@providefam#1{% \\
  \newcommand\#1default{}% Just define it \\
  \bbl@add@list\bbl@font@fams\#1% \\
  \DeclareRobustCommand\#1family{% \\
    \not@math@alphabet\#1family\relax \\
    \fontfamily\#1default\selectfont} \\
  \DeclareTextFontCommand{\text#1}{\#1family}
```

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

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.
\input{\jobname}

\begin{Verbatim}
\def\bbl@font@rst#1#2#3#4{% \\
\bbl@csarg\def{famrst@#4}{\bbl@font@set{#1}#2#3}}
\end{Verbatim}

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

\begin{Verbatim}
\newcommand\babelfSstore[2][]{% \\
  \bbl@ifblank{#1}% \\
  {\bbl@csarg\def{sname@#2}{Latin}}% \\
  {\bbl@csarg\def{sname@#2}{#1}}% \\
  \bbl@provide@dirs{#2}}% \\
\end{Verbatim}

The old tentative way. Short and preserved for compatibility, but deprecated. Note there is no direct alternative for \babelfSfeatures. The reason in explained in the user guide, but essentially – that was not the way to go :-).
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.\TeX sets many "codes" just before loading hyphen.cfg. That is not a problem in luatex, but in xetex they must be reset to the proper value. Most of the work is done in xe(la)tex.ini, so here we just “undo” some of the changes done by \TeX. Anyway, for consistency Lua\TeX also resets the catcodes.

\begin{verbatim}
\let\bbl@langfeatures\@empty
\def\babelFSfeatures{% make sure \fontspec is redefined once
\let\bbl@ori@fontspec\fontspec
\renewcommand\fontspec[1][1]{%\bbl@ori@fontspec[1][\bbl@langfeatures##1]}
}\let\babelFSfeatures\bbl@FSfeatures
\def\bbl@FSfeatures#1#2{\expandafter\addto\csname extras#1\endcsname{\babel@save\bbl@langfeatures\edef\bbl@langfeatures{#2,}}}
\end{verbatim}

\section{Restore Unicode catcodes before loading patterns} ≡

\begin{verbatim}
\begingroup
% 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=999 \fi}
% Letter without case mappings:
\def\l ##1 {\L ##1 ##1 ##1 }
\end{verbatim}
Some more common code.

```latex
\input #1\relax

⟨⟨/Restore Unicode catcodes before loading patterns⟩⟩
```

\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{"\select@language{\bbl@mainlanguage}\bbl@fn@footnote{#2#1{\ignorespaces#4}#3}"
}

\def\bbl@footnote@o#1#2#3[#4]#5{"\select@language{\bbl@mainlanguage}\bbl@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}"
}

\def\BabelFootnote#1#2#3#4{\ifx\bbl@fn@footnote\@undefined\let\bbl@fn@footnote\footnote\fi\ifx\bbl@fn@footnotetext\@undefined\let\bbl@fn@footnotetext\footnotetext\fi\bbl@ifblank{#2}{{\def#1{\bbl@footnote{\@firstofone}{#3}{#4}}\@namedef{\bbl@stripslash#1text}{\bbl@footnotetext{\@firstofone}{#3}{#4}}}{{\def#1{\bbl@exp{\\bbl@footnote{\\foreignlanguage{#2}}}{#3}{#4}}\@namedef{\bbl@stripslash#1text}{\bbl@exp{\\bbl@footnotetext{\\foreignlanguage{#2}}}{#3}{#4}}}}}
```

⟨⟨/Footnote changes⟩⟩

```latex
`\input #1\relax`

⟨⟨/Restore Unicode catcodes before loading patterns⟩⟩
```
Now, the code.

3774 ⟨∗xetex⟩
3775 \def\BabelStringsDefault{unicode}
3776 \let\xebbl@stop\relax
3777 \AddBabelHook{xetex}{encodedcommands}{%
3778 \def\bbl@tempa{#1}%
3779 \ifx\bbl@tempa\@empty
3780 \XeTeXinputencoding"bytes"%
3781 \else
3782 \XeTeXinputencoding"#1"%
3783 \fi
3784 \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
3785 \AddBabelHook{xetex}{stopcommands}{%
3786 \xebbl@stop
3787 \let\xebbl@stop\relax}
3788 \def\bbl@intraspace#1 #2 #3\@@{%
3789 \bbl@csarg\gdef{xeisp@\bbl@cs{sbcp@\languagename}}%
3790 \{\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}%
3791 \def\bbl@intrapenalty#1\@@{%
3792 \bbl@csarg\gdef{xeipn@\bbl@cs{sbcp@\languagename}}%
3793 \{\XeTeXlinebreakpenalty #1\relax}%
3794 \def\bbl@provide@intraspace{%
3795 \bbl@xin@{\bbl@cs{sbcp@\languagename}}{Thai,Laoo,Khmr}%
3796 \ifin@ % sea (currently ckj not handled)
3797 \bbl@ifunset{bbl@intsp@\languagename}{}%
3798 \{\expandafter\ifx\csname bbl@intsp@\languagename\endcsname\@empty\else
3799 \bbl@exp{%
3800 \\\bbl@intraspace\bbl@cs{intsp@\languagename}\@@}%
3801 \fi
3802 \bbl@intrapenalty\@@
3803 \bbl@intrapenalty0\@@
3804 \bbl@intrapenalty\@@
3805 \fi}
3806 \fi
3807 \ifx\bbl@KVP@intraspace\nil\else % We may override the ini
3808 \expandafter\bbl@intraspace\bbl@KVP@intraspace\@\@@
3809 \fi
3810 \ifx\bbl@KVP@intrapenalty\nil\else
3811 \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@\@@
3812 \fi
3813 \ifx\bbl@ispacesize\undefined
3814 \AtBeginDocument{%
3815 \expandafter\bbl@add
3816 \csname selectfont \endcsname{\bbl@ispacesize}}%
3817 \def\bbl@ispacesize{\bbl@cs{xeisp@\bbl@cs{sbcp@\languagename}}}%
3818 \fi}
3819 \fi
3820 \AddBabelHook{xetex}{loadkernel}{%
3821 ⟨⟨Restore Unicode catcodes before loading patterns⟩⟩}
3822 \ifx\DisableBabelHook\undefined\endinput\fi
3823 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
3824 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ckeckstdfonts}
3825 \DisableBabelHook{babel-fontspec}
3826 ⟨⟨Font selection⟩⟩
3827 \input txtbabel.def
3828 ⟨/xetex⟩
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, \bbl@startskip\adim.

Consider \texttt{txt babel} as a shorthand for \emph{tex--xet babel}, which is the bidi model in both pdftex and xetex.
Implicitly reverses sectioning labels in \texttt{bidi=basic}, because the full stop is not in contact with L numbers any more. I think there must be a better way.

15.3 LuaTeX

The new loader for luatex is based solely on language.dat, which is read on the fly. The code shouldn’t be executed when the format is build, so we check if \texttt{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 \texttt{\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, \texttt{\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 reassign languages, which would require modifying the language list, and better error handling. We need catcode tables, but no format (targeted by babel) provide a command to allocate them (although there are packages like ctablestack). For the moment, a dangerous approach is used – just allocate a high random number and cross the fingers. To complicate things, etex.sty changes the way languages are allocated.

\begin{verbatim}
3905 ⟨∗luatex⟩
3906 \if\AddBabelHook@undefined
3907 \bbl@trace{Read language.dat}
3908 \if\bbl@readstream@undefined
3909 \csname newread\endcsname\bbl@readstream
3910 \fi
3911 \bgroup
3912 \toks@{}
3913 \count@ 0=start, 1=0th, 2=normal
3914 \def\bbl@process@line{#1#2 #3 #4}{%
3915  \ifx=#1%
3916  \bbl@process@synonym{#2}%
3917  \else
3918  \bbl@process@language{#1#2}{#3}{#4}%
3919  \fi
3920 \ignorespaces}
3921 \def\bbl@manylang{%
3922  \ifnum\bbl@last>1\relax
3923    \bbl@info{Non-standard hyphenation setup}%
3924  \fi
3925  \let\bbl@many\relax}
3926 \def\bbl@process@language#1#2#3{%
3927  \ifcase\count@
3928    \ifundefined{zth@#1}{\count@ tw}{\count@ one}%
3929    \or
3930    \count@ tw%
3931  \fi
3932  \ifnum\count@ 2=tw@
3933    \expandafter\newlanguage\csname l@#1\endcsname
3934 \language\allocationnumber
3935 \chardef\bbl@last\allocationnumber
3936 \bbl@manylang
3937 \let\bbl@elt\relax
3938 \def\bbl@languages{%
3939  \bbl@languages\bbl@elt{#1}{the\language}{#2}{#3}}%
3940 \fi
3941 \the\toks@
3942 \toks@{}}
3943 \def\bbl@process@synonym@aux#1#2{%
3944 \global\expandafter\chardef\csname l@#1\endcsname#2\relax
3945 \let\bbl@elt\relax
3946 \def\bbl@languages{%
3947 \bbl@languages\bbl@elt{#1}{#2}{}}%
3948 \def\bbl@process@synonym#1{%
3949 \ifcase\count@
3950 \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
3951 \or
3952 \ifundefined{zth@#1}{\bbl@process@synonym@aux{#1} 0}{%}
3953 \else
3954 \bbl@process@synonym@aux{#1}{the\bbl@last}%
\end{verbatim}
\begin{verbatim}
4014 \input #1\relax
4015 \catcodetable\babel\catcodetablenum\relax
4016 \endgroup
4017 \def\bbl@tempa{#2}\
4018 \ifx\bbl@tempa\@empty\else
4019 \input #2\relax
4020 \fi
4021 \egroup\%
4022 \def\bbl@patterns@lua#1{\
4023 \language=\expandafter\ifx\csname l@#1:f@encoding\endcsname\relax
4024 \csname l@#1\endcsname
4025 \edef\bbl@tempa{#1}\
4026 \else
4027 \csname l@#1:f@encoding\endcsname
4028 \edef\bbl@tempa{#1:f@encoding}\
4029 \fi\relax
4030 \@namedef{lu@texhyphen@loaded@the\language}{% Temp
4031 \@ifundefined{bbl@hyphendata@the\language}{}{\
4032 {\def\bbl@elt##1##2##3##4{\
4033 \ifnum##2=\csname l@#1:encoding\endcsname % #2=spanish, dutch:OT1...
4034 \def\bbl@tempb{##3}\
4035 \ifx\bbl@tempb\@empty\else % if not a synonymous
4036 \def\bbl@tempc{{##3}{##4}}\%
4037 \fi
4038 \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4039 \fi}%)%
4040 \bbl@languages
4041 \@ifundefined{bbl@hyphendata@the\language}{}{
4042 {\bbl@info{No hyphenation patterns were set for\
4043 language 'bl@tempa'. Reported}}%
4044 {\expandafter\expandafter\expandafter\bbl@luapatterns
4045 \csname bbl@hyphendata@the\language\endcsname}}}}
4046 \endinput\fi
4047 \begingroup
4048 \catcode`\%=12
4049 \catcode`\%=12
4050 \catcode`\%=12
4051 \catcode`\%=12
4052 \directlua{
4053 Babel = Babel or {}\
4054 function Babel.bytes(line)
4055 return line:gsub("(.)","",
4056 function (chr) return unicode.utf8.char(string.byte(chr)) end)
4057 end
4058 function Babel.begin_process_input()
4059 if luatexbase and luatexbase.add_to_callback then
4060 luatexbase.add_to_callback('process_input_buffer',
4061 Babel.bytes, 'Babel.bytes')
4062 else
4063 Babel.callback = callback.find('process_input_buffer')
4064 callback.register('process_input_buffer', Babel.bytes)
4065 end
4066 end
4067 function Babel.end_process_input ()
4068 if luatexbase and luatexbase.remove_from_callback then
4069 luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
4070 else
4071 callback.register('process_input_buffer', Babel.callback)
4072 end
4073 \end{verbatim}
4073  end
4074  function Babel.addpatterns(pp, lg)
4075      local lg = lang.new(lg)
4076      local pats = lang.patterns(lg) or ''
4077      lang.clear_patterns(lg)
4078      for p in pp:gmatch('[^%s]+') do
4079          ss = ''
4080          for i in string.utfcharacters(p:gsub('%d', '')) do
4081              ss = ss .. '%d?' .. i
4082          end
4083          ss = ss:gsub('^%%d%?%.', '%%.') .. '%d?'
4084          pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
4085          if n == 0 then
4086              tex.sprint(['\[\{\text{New pattern: '.. p .. '}\}\]'])
4087              pats = pats .. ' ' .. p
4088          else
4089              tex.sprint(['\[\{\text{Renew pattern: '.. p .. '}\]\}'])
4090          end
4091      end
4092  end
4093  lang.patterns(lg, pats)
4094  end
4095}
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 \belposthyphenation. In progress. Replace regular (ie, implicit) discretionaries by spaceskips, based on the previous glyph (which I think makes sense, because the hyphen and the previous char go always together). Other discretionaries are not touched. For the moment, only 3 SA languages are activated by default (see Unicode UAX 14).

\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
}

\begingroup
\catcode`%=12
\catcode`^=14
\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, Cyril
          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
end
end
}
\bbl@luahyphenate}
\catcode'\%=14
\gdef\bbl@cjkintraspaces\
\let\bbl@cjkintraspaces\par
\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 % )
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)
}%
\bl@luahyphenate}
\gdef\bl@luahyphenate{%
\let\bl@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
    end
  end
  if Babel.sea_enabled then
    Babel.sea_disc_to_space(head)
  end
}%
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.
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.

```plaintext
\directlua{
Babel.script_blocks = {
  ['Arab'] = {{0x0600, 0x06FF}, {0x08A0, 0x08FF}, {0x0750, 0x077F},
              {0xFE70, 0xFEFF}, {0xFB50, 0xFDFF}, {0x1EE00, 0x1EEFF}},
  ['Armn'] = {{0x0530, 0x058F}},
  ['Beng'] = {{0x0980, 0x09FF}},
  ['Cntr'] = {{0x013A0, 0x13FF}, {0xAB70, 0xABBF}},
  ['Cyr'] = {{0x0400, 0x04FF}, {0x0500, 0x052F}, {0x1C80, 0x1CBF},
             {0x2DE0, 0x2DFF}, {0xA640, 0xA69F}},
  ['Deva'] = {{0x0900, 0x097F}, {0xA8E0, 0xA8FF}},
  ['Ethi'] = {{0x1200, 0x137F}, {0x1380, 0x139F}, {0x2D80, 0x2DDF},
              {0x0AB00, 0x0AB2F}},
  ['Geor'] = {{0x0590, 0x05FF}},
  ['Hans'] = {{0x02E80, 0x2EFF}, {0x3000, 0x303F}, {0x31C0, 0x31EF},
              {0x3300, 0x33FF}, {0x3400, 0x4DBF}, {0x4E00, 0x9FF}},
  ['Hebr'] = {{0x1780, 0x17FF}, {0x0180, 0x024F}, {0x1E00, 0x1EFF},
              {0x2C60, 0x2C7F}, {0x0A720, 0x0A7FF}, {0xA8B00, 0xA8B2F}},
  ['Hant'] = Babel.script_blocks.Hans
  ['Hans'] = Babel.script_blocks.Hans
  ['Kana'] = Babel.script_blocks.Kana
  ['Khmr'] = {{0x01C80, 0x1CBF}, {0x2D80, 0x2DDF}, {0xA640, 0xA69F}},
  ['Kore'] = {{0x0180, 0x01FF}, {0x01E0, 0x1EFF}, {0x2C60, 0x2C7F},
              {0x0A720, 0x0A7FF}, {0xA8B00, 0xA8B2F}},
  ['Laoo'] = {{0x0E80, 0x0EFF}},
  ['Latn'] = {{0x02E80, 0x02FF}, {0x0980, 0x09FF}, {0x0750, 0x077F},
              {0x01C80, 0x01CBF}, {0x2D80, 0x2DDF}, {0xA640, 0xA69F}},
  ['Mahj'] = {{0x01150, 0x1117F}},
  ['Mlym'] = {{0x01150, 0x0D7F}},
  ['Mymr'] = {{0x01150, 0x0D7F}},
  ['Orya'] = {{0x01150, 0x0D7F}},
  ['Sinh'] = {{0x01150, 0x0D7F}},
  ['Tam'] = {{0x01150, 0x0D7F}},
  ['Tel'] = {{0x01150, 0x0D7F}},
  ['Tfng'] = {{0x01150, 0x0D7F}},
  ['Thai'] = {{0x01150, 0x0D7F}},
  ['Tib'] = {{0x01150, 0x0D7F}},
  ['Vaii'] = {{0x01150, 0x0D7F}},
  ['Yi'] = {{0x01150, 0x0D7F}},
}
}"
The code for \texttt{\textbackslash babelcharproperty} is straightforward. Just note the modified lua table can be different.

\begin{verbatim}
\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)}\%\
 \fi
}
\newcommand\bbl@chprop[3][\the\count@]{%
 \@tempcnta=#1\relax
 \ifnum\@tempcnta<#2\@tempcnta=#2\relax\fi
 \ifnum\@tempcnta>#3\@tempcnta=#3\relax\fi
 \ifnum\@tempcnta=-1\@tempcnta=0\relax\fi
}
\end{verbatim}
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: \texttt{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.); \texttt{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).

\texttt{post\_hyphenate\_replace} is the callback applied after \texttt{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.
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) and not(item.char == 61) 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
      return word_string, word_nodes, item, lang
    end
    item = item.next
  end
  return head
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
    &% pass
    else
      return word_string, word_nodes, item, lang
    end
    item = item.next
  end
  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
      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)
        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)
else if 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
end &% for words
return head
end

&% The following functions belong to the next macro
Babel.capture_maps = {}

&% This table stores capture maps, numbered consecutively
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
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, \{pre={1}{1}--\} becomes \begin{verbatim}
function(m) return m[1]..m[1]..'-' end
\end{verbatim}, 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 –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 @, we just avoid this character in macro names (which
explains the internal group, too).
\begin{verbatim}
catcode`\#=6
\glob\def\babelposthyphenation#1#2#3{&%
  \bbl@activateposthyphen
  \begingroup
    \def\babeltempa{\bbl@add@list\babeltempb}&%
    \let\babeltempb\@empty
    \bbl@foreach{#3}{&%
      \bbl@ifsamestring{##1}{remove}&%
      {\bbl@add@list\babeltempb{nil}}&%
      \directlua{
        local rep = [[##1]]
        rep = rep:gsub( '(no)%s*=%s*(\[^%s,\]*)', Babel.capture_func)
        rep = rep:gsub( '(pre)%s*=%s*(\[^%s,\]*)', Babel.capture_func)
        rep = rep:gsub( '(post)%s*=%s*(\[^%s,\]*)', Babel.capture_func)
        rep = rep:gsub('(string)%s*=%s*(\[^%s,\]*)', Babel.capture_func)
        tex.print([[\string\babeltempa{{\[
        .. rep .. "]}}}])}
    }
  \endgroup
}\glob\def\babelposthyphenation#1#2#3{&%
\end{verbatim
Work in progress.

Unlike \texttt{xetex}, \texttt{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 \texttt{bidi=basic}, without having to patch almost any macro where text direction is relevant. 

\texttt{@hangfrom} is useful in many contexts and it is redefined always with the \texttt{layout} option. There are, however, a number of issues where the text direction is not the same as the box direction (as set by \texttt{\bodydir}), and when \texttt{\parbox} and \texttt{\hangindent} are involved. Fortunately, latest releases of \texttt{luatex} simplify a lot the solution with \texttt{\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, \texttt{tabular} seems to work (at least in simple cases) with \texttt{array}, \texttt{tabularx}, \texttt{hhline}, \texttt{colortbl}, \texttt{longtable}, \texttt{booktabs}, etc. However, \texttt{dcolumn} still fails.

\begin{verbatim}
15.7 Layout

Work in progress.

Unlike \texttt{xetex}, \texttt{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 \texttt{bidi=basic}, without having to patch almost any macro where text direction is relevant.

\texttt{@hangfrom} is useful in many contexts and it is redefined always with the \texttt{layout} option. There are, however, a number of issues where the text direction is not the same as the box direction (as set by \texttt{\bodydir}), and when \texttt{\parbox} and \texttt{\hangindent} are involved. Fortunately, latest releases of \texttt{luatex} simplify a lot the solution with \texttt{\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, \texttt{tabular} seems to work (at least in simple cases) with \texttt{array}, \texttt{tabularx}, \texttt{hhline}, \texttt{colortbl}, \texttt{longtable}, \texttt{booktabs}, etc. However, \texttt{dcolumn} still fails.
\end{verbatim}
Implicitly reverses sectioning labels in bidi=basic-r, because the full stop is not in contact with L numbers any more. I think there must be a better way. 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.

\begin{verbatim}
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:

\begin{verbatim}
[0x25]={d='et'},
[0x26]={d='on'},
[0x27]={d='on'},
[0x28]={d='on', m=0x29},
[0x29]={d='on', m=0x28},
[0x2A]={d='on'},
[0x2B]={d='es'},
[0x2C]={d='cs'},
\end{verbatim}

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 (\langle l\rangle, \langle r\rangle or \langle al\rangle).

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.

4895 (\langle basic-r\rangle)
4896 Babel = Babel or {}
4897 Babel.bidi_enabled = true
4899 require('babel-data-bidi.lua')
4901 local characters = Babel.characters
4903 local ranges = Babel.ranges
4904 local(DIR = node.id("dir")
4906 local function dir_mark(head, from, to, outer)
4908 dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
4909 local d = node.new(DIR)
4910 d.dir = '+' .. dir
4911 node.insert_before(head, from, d)
4912 d = node.new(DIR)
4913 d.dir = '-' .. dir
4914 node.insert_after(head, to, d)
4915 end
4917 function Babel.bidi(head, ispar)
4918 local first_n, last_n -- first and last char with nums
4919 local last_es -- an auxiliary 'last' used with nums
4920 local first_d, last_d -- first and last char in L/R block
4921 local dir, dir_real

Next also depends on script/lang (\langle al\rangle/\langle r\rangle). 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):
4922 local strong = ('TRT' == tex.pardir) and 'r' or 'l'
4923 local strong_lr = (strong == 'l') and 'l' or 'r'
4924 local outer = strong
4925 local new_dir = false
4927 local first_dir = false
4928 local innmath = false
4929 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
    dir = dir or 'l'
  end
  end
  if dir == 'nsm' then dir = strong end -- W1
end
if dir_real then
  attr_dir = 0
  for at in node.traverse(item.attr) do
    if at.number == luatexbase.registernumber'bbl@attr@dir' then
      attr_dir = at.value % 3
    end
  end
  if attr_dir == 1 then
    strong = 'r'
  elseif attr_dir == 2 then
    strong = 'al'
  else
    strong = 'l'
  end
  strong_lr = (strong == 'l') and 'l' or 'r'
  outer = strong_lr
  new_dir = false
end
if dir == 'nsm' then dir = strong end -- W1

Numbers. The dual <al>/<r> system for R is somewhat cumbersome.
dir_real = dir -- We need dir_real to set strong below
if dir == 'al' then dir = 'r' end -- W3
By W2, there are no \texttt{en}, \texttt{et}, \texttt{es} if \texttt{strong} == \texttt{al}, only \texttt{an}. Therefore, there are not \texttt{et en} nor \texttt{en et}, W5 can be ignored, and W6 applied:

```plaintext
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.

```plaintext
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)
else
  dir = nil -- Not a char
end
```

Numbers in R mode. A sequence of \texttt{en}, \texttt{et}, \texttt{an}, \texttt{es} and \texttt{cs} is typeset (with some rules) in L mode. We store the starting and ending points, and only when anything different is found (including nil, i.e., a non-char), the textdir is set. This means you cannot insert, say, a whatsit, but this is what I would expect (with luacolor you may colorize some digits). Anyway, this behavior could be changed with a switch in the future. Note in the first branch only \texttt{an} is relevant if \texttt{al}.

```plaintext
if dir == 'en' or dir == 'an' or dir == 'et' then
  if dir ~= 'et' then
    type_n = dir
  end
  first_n = first_n or item
  last_n = last_es or item
  last_es = nil
elseif dir == 'es' and last_n then -- W3+W6
  last_es = item
elseif dir == 'cs' then -- it's right - do nothing
elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
  if strong_lr == 'r' and type_n == '' then
    dir_mark(head, first_n, last_n, 'r')
  elseif strong_lr == 'l' and first_d and type_n == 'an' then
    dir_mark(head, first_n, last_n, 'r')
    dir_mark(head, first_d, last_d, outer)
    first_d, last_d = nil, nil
  elseif strong_lr == 'l' and type_n == '' then
    last_d = last_n
    type_n = ''
  first_n, last_n = nil, nil
end
```

R text in L, or L text in R. Order of \texttt{dir_mark}’s are relevant: d goes outside n, and therefore it's emitted after. See \texttt{dir_mark} to understand why (but is the nesting actually necessary or is a flat dir structure enough?). Only L, R (and AL) chars are taken into account – everything else, including spaces, whatsits, etc., are ignored:

```plaintext
if dir == 'l' or dir == 'r' then
  if dir ~= outer then
    first_d = first_d or item
    last_d = item
  elseif first_d and dir ~= strong_lr then
```

174
```lua
if dir and not last_lr and dir ~= 'l' and outer == 'r' then
    item.char = characters[item.char] and
    characters[item.char].m or item.char
elseif (dir or new_dir) and last_lr ~= item then
    local mir = outer .. strong_lr .. (dir or outer)
    if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
        for ch in node.traverse(node.next(last_lr)) do
            if ch == item then break end
            if ch.id == node.id'glyph' and characters[ch.char] then
                ch.char = characters[ch.char].m or ch.char
            end
        end
    end
end
end
end
end
end

Save some values for the next iteration. If the current node is 'dir', open a new sequence.
Since dir could be changed, strong is set with its real value (dir_real).
if dir == 'l' or dir == 'r' then
    last_lr = item
    strong = dir_real -- Don't search back - best save now
    strong_lr = (strong == 'l') and 'l' or 'r'
elseif new_dir then
    last_lr = nil
end
end
end
end
end
end

Mirror the last chars if they are no directed. And make sure any open block is closed, too.
if last_lr and outer == 'r' then
    for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
        if characters[ch.char] then
            ch.char = characters[ch.char].m or ch.char
        end
    end
end
end
end
end
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
end
end
end
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)
  end
  new_state.sim, new_state.eim = nil, nil
  return head, new_state
end

local function insert_numeric(head, state)
  local new_state = state
  if state.san and state.ean and state.san ~= state.ean then
    local d = node.new(DIR)
    d.dir = '+TLT'
    _, new = node.insert_before(head, state.san, d)
    if state.san == state.sim then state.sim = new end
    local d = node.new(DIR)
    d.dir = '-TLT'
    _, new = node.insert_after(head, state.ean, d)
    if state.ean == state.eim then state.eim = new end
  end
  new_state.san, new_state.ean = nil, nil
  return head, new_state
end

-- TODO - \hbox with an explicit dir can lead to wrong results
-- \hbox{TLT{<R>}} and \hbox{TRT{<L>}}. A small attempt
-- was made to improve the situation, but the problem is the 3-dir
-- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
-- well.

function Babel.bidi(head, ispar, hdir)
local d -- d is used mainly for computations in a loop
local prev_d = ''
local new_d = false
local nodes = {}
local outer_first = nil
local inmath = false
local glue_d = nil
local glue_i = nil
local has_en = false
local first_et = nil
local ATDIR = luatexbase.registernumber'bbl@attr@dir'
local save_outer
local temp = node.get_attribute(head, ATDIR)
if temp then
  temp = temp % 3
  save_outer = (temp == 0 and 'l') or (temp == 1 and 'r') or (temp == 2 and 'al')
elseif ispar then -- Or error? Shouldn't happen
  save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
else -- Or error? Shouldn't happen
  save_outer = ('TRT' == 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 last = outer
-- 'al' is only taken into account in the first, current loop
if save_outer == 'al' then save_outer = 'r' end
local fontmap = Babel.fontmap
for item in node.traverse(head) do
  -- In what follows, #node is the last (previous) node, because the
  -- current one is not added until we start processing the neutrals.
  -- three cases: glyph, dir, otherwise
  if item.id == GLYPH
    or (item.id == 7 and item.subtype == 2) then
    local d_font = nil
    local item_r
    if item.id == 7 and item.subtype == 2 then
      item_r = item.replace -- automatic discs have just 1 glyph
    else
      item_r = item
    end
    local chardata = characters[item_r.char]
    d = chardata and chardata.d or nil
    if not d or d == 'nsm' then

for nn, et in ipairs(ranges) do
  if item_r.char < et[1] then
    break
  elseif item_r.char <= et[2] then
    if not d then d = et[3] end
    if d == 'nsm' then d_font = et[3] end
    break end
  end
end

d = d or 'l'

-- A short 'pause' in bidi for mapfont
d_font = d_font or d

if d_font and fontmap and fontmap[d_font][item_r.font] then
  item_r.font = fontmap[d_font][item_r.font]
end

if new_d then
  table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
  if inmath then
    attr_d = 0
  else
    attr_d = node.get_attribute(item, ATDIR) % 3
  end
  if attr_d == 1 then
    outer_first = 'r'
    last = 'r'
  elseif attr_d == 2 then
    outer_first = 'l'
    last = 'l'
  else
    outer_first = 'l'
    last = 'l'
  end
  outer = last
  has_en = false
  first_et = nil
  new_d = false
end

if glue_d then
  if (d == 'l' and 'l' or 'r') ~= glue_d then
    table.insert(nodes, {glue_i, 'on', nil})
  end
  glue_d = nil
  glue_i = nil
end

elseif item.id == DIR then
  d = nil
  new_d = true
elseif item.id == node.id'glue' and item.subtype == 13 then
  glue_d = d
  glue_i = item
  d = nil
elseif item.id == node.id'math' then
  inmath = (item.subtype == 0)
else
  d = nil
end

-- AL <= EN/ET/ES -- W2 + W3 + W6
if last == 'al' and d == 'en' then
  d = 'an' -- W3
elseif last == 'al' and (d == 'et' or d == 'es') then
  d = 'on' -- W6
end

-- EN + CS/ES + EN -- W4
if d == 'en' and #nodes >= 2 then
  if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
    and nodes[#nodes-1][2] == 'en' then
    nodes[#nodes][2] = 'en'
end
end

-- AN + CS + AN -- W4 too, because uax9 mixes both cases
if d == 'an' and #nodes >= 2 then
  if (nodes[#nodes][2] == 'cs')
    and nodes[#nodes-1][2] == 'an' then
    nodes[#nodes][2] = 'an'
end
end

-- ET/EN -- W5 + W7->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
end
for e = first_et, #nodes do
  if nodes[e][1].id == GLYPH then nodes[e][2] = temp
  end
end
first_et = nil
has_en = false
end
if d then
    if d == 'al' then
        d = 'r'
        last = 'al'
    elseif d == 'l' or d == 'r' then
        last = d
    end
    prev_d = d
    table.insert(nodes, {item, d, outer_first})
end
outer_first = nil
end
-- TODO -- repeated here in case EN/ET is the last node. Find a
-- better way of doing things:
if first_et then -- dir may be nil here!
    if has_en then
        if last == 'l' then
            temp = 'l' -- W7
        else
            temp = 'en' -- W5
        end
    else
        temp = 'on' -- W6
    end
    for e = first_et, #nodes do
        if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
    end
end
-- dummy node, to close things
table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})

-- NEUTRAL -----------------
outer = save_outer
last = outer
local first_on = nil
for q = 1, #nodes do
    local item
    local outer_first = nodes[q][3]
    outer = outer_first or outer
    last = outer_first or last
    local d = nodes[q][2]
    if d == 'an' or d == 'en' then d = 'r' end
    if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
    if d == 'on' then
        first_on = first_on or q
    elseif first_on then
        if last == d then
            temp = d
        else
            temp = d
        end
    else
        temp = d
    end
end
temp = outer
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 -- W1
    if d == 'en' then d = 'an' end
    local isdir = (d == 'r' or d == 'l')
    if outer == 'l' and d == 'an' then
        state.san = state.san or item
        state.ean = item
    elseif state.san then
        head, state = insert_numeric(head, state)
    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
        end
    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
            state.san, state.ean = nil, nil
        end
    else
        if d == 'an' or d == 'l' then
        end
    end
end
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
end
return node.prev(head) or head end

\ProvidesLanguage{nil}[⟨⟨date⟩⟩ ⟨⟨version⟩⟩] Nil language
\LdfInit{nil}{datenil}

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

\ift\l@nil\@undefined
\newlanguage\l@nil
\edef\bbl@languages{% Add it to the list of languages

\section{Data for CJK}

It is a boring file and it is not shown here (see the generated file), but here is a sample:

\begin{verbatim}
[0x0021]={c='ex'},
[0x0024]={c='pr'},
[0x0025]={c='po'},
[0x0028]={c='op'},
[0x0029]={c='cp'},
[0x002B]={c='pr'},
\end{verbatim}

For the meaning of these codes, see the Unicode standard.

\section{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.

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

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.
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
\let\captionsnil\@empty
\let\datenil\@empty

The macro \ldf@finish takes care of looking for a configuration file, setting the main language to be switched on at \begin{document} and resetting the category code of @ to its original value.

\ldf@finish(nil)
\(/nil\)

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 diddle with hyphen.tex (or plain.tex except to preload additional fonts).

The files bplain.tex and blplain.tex can be used as replacement wrappers around plain.tex and lplain.tex to achieve the desired effect, based on the babel package. If you load each of them with iniTeX, 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 initex sees, we need to set some category codes just to be able to change the definition of \input

\let\a\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...

\openin 0 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

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

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

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

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

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

### 18.2 Emulating some \LaTeX features

The following code duplicates or emulates parts of \LaTeX\textsuperscript{2e} 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 }
\immediate\write16{ * }
\fi 
\input #1.cfg\relax 
\@endofldf}

### 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\long#1}}

\long\def\firstoftwo#1\@nil{#1}
\long\def\secondoftwo#1\@nil{\@nil}
\def\gobbletwo#1\@nil{\@nil}
\def\ifstar#1\@nil{\ifnextchar *{\@firstoftwo{#1}}}
\def\star@or@long#1\@nil{\@ifstar{\@nil}{\long#1}}

\let\l@ngrel@x\relax
\let\l@ngrel@x\relax

\let\l@ngrel@x\relax
\let\l@ngrel@x\relax

184
LATEX has the command \@onlypreamble which adds commands to a list of commands that are no longer needed after \begin{document}.

\ifx\@preamblecmds\@undefined
\def\@preamblecmds{}
\fi
\@onlypreamble\@preamblecmds\global\let\do\noexpand
\ifx\@begindocumenthook\@undefined
\def\@begindocumenthook{}
\fi
\@onlypreamble\@begindocumenthook\def\AtBeginDocument{\g@addto@macro\@begindocumenthook}

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

\def\AtEndOfPackage#1{\g@addto@macro\@endofldf{#1}}
\LaTeX{} needs to be able to switch off writing to its auxiliary files; plain doesn’t have them by default.

\begin{verbatim}
552 \ifx\if@filesw@undefined
553\expandafter\let\csname if@filesw\expandafter\endcsname
554\csname iffalse\endcsname
\fi

Mimick \LaTeX{}’s commands to define control sequences.
\begin{verbatim}
556 \def\newcommand{\@star@or@long\new@command}
557 \def\new@command#1{%
558 \@testopt{\@newcommand#1}0
559 \@newcommand#1[#2][]{%
560 \@argdef#1[#2]#3{%
561 \@xargdef#1[#2][#3]#4{%
562 \long:def@argdef#1[#2][#3][#4]xxx#5
563 \long:def@argdef#1[#2][#3][#4]xxx#5
564 \long:def@argdef#1[#2][#3][#4]xxx#5
565 \long:def@argdef#1[#2][#3][#4]xxx#5
566 \long:def@argdef#1[#2][#3][#4]xxx#5
567 \long:def@argdef#1[#2][#3][#4]xxx#5
568 \long:def@argdef#1[#2][#3][#4]xxx#5
569 \long:def@argdef#1[#2][#3][#4]xxx#5
570 \long:def@argdef#1[#2][#3][#4]xxx#5
571 \long:def@argdef#1[#2][#3][#4]xxx#5
572 \long:def@argdef#1[#2][#3][#4]xxx#5
573 \long:def@argdef#1[#2][#3][#4]xxx#5
574 \long:def@argdef#1[#2][#3][#4]xxx#5
575 \long:def@argdef#1[#2][#3][#4]xxx#5
576 \long:def@argdef#1[#2][#3][#4]xxx#5
577 \long:def@argdef#1[#2][#3][#4]xxx#5
578 \long:def@argdef#1[#2][#3][#4]xxx#5
579 \long:def@argdef#1[#2][#3][#4]xxx#5
580 \long:def@argdef#1[#2][#3][#4]xxx#5
581 \long:def@argdef#1[#2][#3][#4]xxx#5
582 \long:def@argdef#1[#2][#3][#4]xxx#5
583 \long:def@argdef#1[#2][#3][#4]xxx#5
584 \long:def@argdef#1[#2][#3][#4]xxx#5
585 \long:def@argdef#1[#2][#3][#4]xxx#5
586 \long:def@argdef#1[#2][#3][#4]xxx#5
587 \long:def@argdef#1[#2][#3][#4]xxx#5
588 \long:def@argdef#1[#2][#3][#4]xxx#5
589 \long:def@argdef#1[#2][#3][#4]xxx#5
590 \long:def@argdef#1[#2][#3][#4]xxx#5
591 \long:def@argdef#1[#2][#3][#4]xxx#5
592 \long:def@argdef#1[#2][#3][#4]xxx#5
593 \long:def@argdef#1[#2][#3][#4]xxx#5
594 \long:def@argdef#1[#2][#3][#4]xxx#5
595 \long:def@argdef#1[#2][#3][#4]xxx#5
596 \long:def@argdef#1[#2][#3][#4]xxx#5
597 \long:def@argdef#1[#2][#3][#4]xxx#5
598 \long:def@argdef#1[#2][#3][#4]xxx#5
599 \long:def@argdef#1[#2][#3][#4]xxx#5
600 \long:def@argdef#1[#2][#3][#4]xxx#5
601 \long:def@argdef#1[#2][#3][#4]xxx#5
602 \long:def@argdef#1[#2][#3][#4]xxx#5
603 \long:def@argdef#1[#2][#3][#4]xxx#5
604 \long:def@argdef#1[#2][#3][#4]xxx#5
605 \long:def@argdef#1[#2][#3][#4]xxx#5
\end{verbatim}
\end{verbatim}

\begin{verbatim}
186
\end{verbatim}
\end{verbatim}
The following little macro \in@ is taken from \texttt{latex.ltx}; it checks whether its first argument is part of its second argument. It uses the boolean \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\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 \TeX we assume that the user wants them to be active by default. Therefore the only thing we do is execute the third argument (the code for the true case).

\begin{verbatim}
\def\@ifpackagewith#1#2#3#4{#3}
\end{verbatim}

The \LaTeX macro \texttt{\@ifl@aded} checks whether a file was loaded. This functionality is not needed for plain \TeX but we need the macro to be defined as a no-op.

\begin{verbatim}
\def\@ifl@aded#1#2#3#4{}
\end{verbatim}

For the following code we need to make sure that the commands \texttt{\newcommand} and \texttt{\providecommand} exist with some sensible definition. They are not fully equivalent to their \LaTeX\,2\,\epsilon\,versions; just enough to make things work in plain \TeX\,environments.

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

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

\begin{verbatim}
\ifx\bye\@undefined
\advance\count10 by -2\relax
\fi
\ifx\ifnextchar\@undefined
\def\ifnextchar#1\#2\#3{%
\let\reserved@d=#1%
\def\reserved@a{#2}\def\reserved@b{#3}%
\end{verbatim}
18.4 Encoding related macros

Code from ltoutenc.dtx, adapted for use in the plain \LaTeX\ environment.
Currently we only use the \LaTeX method for accents for those that are known to be made active in some language definition file.
The following control sequences are used in babel.def but are not defined for plain \TeX.

\DeclareTextSymbol{\textquotedblleft}{OT1}{92}
\DeclareTextSymbol{\textquotedblright}{OT1}{"}
\DeclareTextSymbol{\textquoteleft}{OT1}{`}
\DeclareTextSymbol{\textquoteright}{OT1}{'}
\DeclareTextSymbol{i}{OT1}{16}
\DeclareTextSymbol{ss}{OT1}{25}

For a couple of languages we need the \TeX-control sequence \texttt{\scriptsize} to be available. Because plain \TeX doesn't have such a sophisticated font mechanism as \TeX has, we just let it to \sevenrm.

\ifx\scriptsize@undefined
\let\scriptsize\sevenrm
\fi
\langle/plain\rangle

19 Acknowledgements

I would like to thank all who volunteered as \textbeta-testers for their time. Michel Goossens supplied contributions for most of the other languages. Nico Poppelier helped polish the text of the documentation and supplied parts of the macros for the Dutch language. Paul Wackers and Werenfried Spit helped find and repair bugs.

During the further development of the babel system I received much help from Bernd Raichle, for which I am grateful.

References