Babel

Localization and internationalization

Unicode

TeX
pdfTeX
LuaTeX
XeTeX
# Contents

1 Identification and loading of required files

2 locale directory

3 Tools
   3.1 Multiple languages
   3.2 The Package File (L\TeX, babel.sty)
   3.3 base
   3.4 key=value options and other general option
   3.5 Conditional loading of shorthands
   3.6 Interlude for Plain

4 Multiple languages
   4.1 Selecting the language
   4.2 Errors
   4.3 Hooks
   4.4 Setting up language files
   4.5 Shorthands
   4.6 Language attributes
   4.7 Support for saving macro definitions
   4.8 Short tags
   4.9 Hyphens
   4.10 Multiencoding strings
   4.11 Macros common to a number of languages
   4.12 Making glyphs available
      4.12.1 Quotation marks
      4.12.2 Letters
      4.12.3 Shorthands for quotation marks
      4.12.4 Umlauts and tremas
   4.13 Layout
   4.14 Load engine specific macros
   4.15 Creating and modifying languages

5 Adjusting the Babel behavior
   5.1 Cross referencing macros
   5.2 Marks
   5.3 Preventing clashes with other packages
      5.3.1 if\then
      5.3.2 varioref
      5.3.3 hhline
   5.4 Encoding and fonts
   5.5 Basic bidi support
   5.6 Local Language Configuration
   5.7 Language options

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

7 Loading hyphenation patterns

8 Font handling with fontspec

9 Hooks for Xe\TeX{} and Lua\TeX{}
   9.1 Xe\TeX{}

13
The babel package 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 in real documents only as documented (except, of course, if you want to explore and test them).

1 Identification and loading of required files

*Code documentation is still under revision.*

The babel package after unpacking consists of the following files:

- **babel.sty** is the L\(\LaTeX\) package, which set options and load language styles.
- **babel.def** is loaded by Plain.
- **switch.def** defines macros to set and switch languages (it loads part babel.def).
- **plain.def** is not used, and just loads babel.def, for compatibility.
- **hyphen.cfg** is the file to be used when generating the formats to load hyphenation patterns.

There are some additional tex, def and lua files.

The babel installer extends docstrip with a few “pseudo-guards” to set “variables” used at installation time. They are used with \langle@name@\rangle at the appropriate places in the source code and defined with either \langle{name=\textit{value}}\rangle, or with a series of lines between \langle*\textit{name}\rangle and \langle/\textit{name}\rangle. The latter is cumulative (eg, with More package options). That brings a little bit of literate programming. The guards \texttt{<-name>} and \texttt{>+name>} have been redefined, too. See babel.ins for further details.

2 locale directory

A required component of babel is a set of ini files with basic definitions for about 250 languages. They are distributed as a separate zip file, not packed as dtx. 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, there are no geographic areas in Spanish). Not all include LICR variants.

- **babel-*.ini** files contain the actual data; **babel-*.tex** files are basically proxies to the corresponding ini files.

See Keys in ini files in the the babel site.

3 Tools

```
1 ⟨⟨version=24.3⟩⟩
2 ⟨⟨date=2024/03/29⟩⟩

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 \texttt{\bbl@afterfi}, will not change.

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

3 ⟨⟨Basic macros⟩⟩ ≡
4 \bbl@trace{Basic macros}
5 \def\bbl@strip\string\slash{\expandafter\@gobble\string}
6 \def\bbl@add#1#2{\ifunset\bbl@strip\slash#1\{\def#1{#2}\}\expandafter\def\expandafter#1\expandafter{#1#2}}
7 \def\bbl@in@{\@expandtwoargs\in@}
8 \def\bbl@carg#1#2{\expandafter#1\csname#2\endcsname}
9 \def\bbl@ncarg#1#2#3{\expandafter#1\expandafter#2\csname#3\endcsname}
10 \def\bbl@ccarg#1#2#3{\expandafter#1\csname#2\expandafter\endcsname\csname#3\endcsname}
11 \def\bbl@csarg#1#2{\expandafter\csname bbl@#1\endcsname}
12 \def\bbl@cs#1{\expandafter\csname bbl@#1\endcsname}
13 \def\bbl@cl#1{\csname bbl@#1@\languagename\endcsname}
```

3
\def\bbl@loop#1#2#3{\bbl@@loop#1{#3}#2,\@nnil,}
\def\bbl@loopx#1#2{\expandafter\bbl@loop\expandafter#1\expandafter{#2}}
\def\bbl@@loop#1#2#3,{% 
  \ifx\@nnil#3\relax% 
  \else\def#1{#3}#2\bbl@afterfi\bbl@@loop#1{#2}% 
  \fi}
\def\bbl@for#1#2#3{\bbl@loopx#1{#2}{\ifx#1\@empty\else#3\fi}}
\bbl@add@list
This internal macro adds its second argument to a comma separated list in its first argument. When the list is not defined yet (or empty), it will be initiated. It presumes expandable character strings.
\def\bbl@add@list#1#2{\edef#1{\ifunset\bbl@stripslash#1\{}\else\{\ifx#1\@empty\else#1,\fi\}\endif#2}}
\bbl@afterelse
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 \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\ldots\fi statement appears in one of the arguments and it is not enclosed in braces.
\def\bbl@afterelse#1\else#2\fi{\fi#1}
\def\bbl@afterfi#1\fi{\fi#1}
\bbl@exp
Now, just syntactical sugar, but it makes partial expansion of some code a lot more simple and readable. Here \\ stands for \noexpand, \<..> for \noexpand applied to a built macro name (which does not define the macro if undefined to \relax, because it is created locally), and \[..] for one-level expansion (where . is the macro name without the backslash). The result may be followed by extra arguments, if necessary.
\def\bbl@exp##1##2{\begingroup\let\noexpand\bbl@exp@en\let\noexpand\expandafter\bbl@exp@ue\edef\bbl@exp@aux{\endgroup#1}\bbl@exp@aux}
\def\bbl@exp@en#1>{\noexpand\csname#1\endcsname}
\def\bbl@exp@ue#1\{\unexpanded\expandafter\expandafter\expandafter{\csname#1\endcsname}}
\bbl@trim
The following piece of code is stolen (with some changes) from keyval, by David Carlisle. It defines two macros: \bbl@trim and \bbl@trimdef. 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.
\def\bbl@tempa{ }
\def\bbl@trim##1##2{\futurelet\bbl@trim@a\bbl@trim@c##2\@nil\@nil#1\@nil\relax{##1}}
\def\bbl@trim@c{\ifx\bbl@trim@a\@sptoken\expandafter\bbl@trim@b\else\expandafter\bbl@trim@b\expandafter#1\fi}
\def\bbl@trim@b#1##1\@nil\@nil#3{#3{#1}}
\def\bbl@trim@def#1{\bbl@trim{\def#1}}
\bbl@ifunset
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 does not waste
memory. Defined inside a group, to avoid \ifcsname being implicitly set to \relax by the \csname test.

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

\bbl@ifblank A tool from url, by Donald Arseneau, which tests if a string is empty or space. The companion macros tests if a macro is defined with some ‘real’ value, ie, not \relax and not empty.

\def\bbl@ifblank#1{\bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
\long\def\bbl@ifblank@i#1#2\@nil#3#4#5\@nil{#4}
\def\bbl@ifset#1#2#3{\bbl@ifunset{#1}{#3}{\bbl@exp{\bbl@ifblank{\@nameuse{#1}}}{#3}{#2}}}

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).

\def\bbl@forkv#1#2{\def\bbl@kvcmd##1##2##3{#2}\bbl@kvnext#1,\@nil,}
\def\bbl@kvnext#1,{\ifx\@nil#1\relax\else\bbl@ifblank{#1}{}{\bbl@forkv@eq#1=#2=#3\@nil#4}{\expandafter\bbl@kvnext}\fi}
\def\bbl@forkv@eq#1=#2=#3\@nil#4{\bbl@trim@def\bbl@forkv@a{#1}\bbl@trim{\expandafter\bbl@kvcmd\expandafter{\bbl@forkv@a}}{#2}{#4}}

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

\def\bbl@foreach#1{\expandafter\bbl@vforeach\expandafter{#1}}
\bbl@replace Returns implicitly \toks@ with the modified string.
\def\bbl@replace#1#2#3{% in #1 -> repl #2 by #3
\toks@{}\def\bbl@replace@aux#1#2#3#4{%
An extension to the previous macro. It takes into account the parameters, and it is string based (i.e., if you replace \relax by ho, then \rho 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@ifsamestring 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 in your language style if you want.
As a somewhat hackish tool (hence its name) to avoid spurious spaces in some contexts.

Another hackish tool, to apply case changes inside a protected macros. It's based on the internal \let's made by \MakeUppercase and \MakeLowercase between things like \oe and \OE.

The following adds some code to \extras... both before and after, while avoiding doing it twice. It's somewhat convoluted, to deal with #'s. Used to deal with alph, Alph and frenchspacing when there are already changes (with \babel@save).

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

### 3.1 Multiple languages

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.
Another counter is used to keep track of the allocated languages. TeX and \TeX{} reserves for this purpose the count 19.

This macro was introduced for TeX < 2. Preserved for compatibility.

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. 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.

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

Start with some “private” debugging tool, and then define macros for errors.

\begin{verbatim}
\@ifpackagewith{babel}{debug}{
{\providecommand\bbl@trace[1]{\message{^^J[ #1 ]}}%\let\bbl@debug\@firstofone
\ifx\directlua\@undefined\else\directlua{ Babel = Babel or {}
Babel.debug = true }\%\input{babel-debug.tex}\fi}
{\providecommand\bbl@trace[1]{}%\let\bbl@debug\@gobble
\ifx\directlua\@undefined\else\directlua{ Babel = Babel or {}
Babel.debug = false }\fi}
\def\bbl@error#1{% Implicit #2#3#4
\begingroup\catcode`\=0 \catcode`\=12 \catcode`\=12
\input errbabel.def\endgroup\bbl@error{#1}}
\def\bbl@warning#1{\begingroup\def\MessageBreak{\PackageWarning{babel}{#1}\
\endgroup}
\def\bbl@infowarn#1{\begingroup\def\MessageBreak{\PackageNote{babel}{#1}\
\endgroup}
\def\bbl@info#1{\begingroup\def\MessageBreak{\PackageInfo{babel}{#1}\
\endgroup}\
\end{verbatim}
This file also takes care of a number of compatibility issues with other packages and defines a few additional package options. Apart from all the language options below we also have a few options that influence the behavior of language definition files.

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

But first, include here the Basic macros defined above.

```latex
\@ifpackagewith{babel}{silent}
{\let\bbl@info\@gobble
 \let\bbl@infowarn\@gobble
 \let\bbl@warning\@gobble}
```

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. Also available with base, because it just shows info.

```latex
\ifx\bbl@languages\@undefined
\else
\begingroup
\catcode`^^I=12
\@ifpackagewith{babel}{showlanguages}{%
\begingroup
\def\bbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}%
\wlog{<*languages>}%
\bbl@languages%
\wlog{</languages>}%
\endgroup}{}
\endgroup
\def\bbl@elt#1#2#3#4{%\ifnum#2=\z@
\gdef\bbl@nulllanguage{#1}%
\def\bbl@elt##1##2##3##4{}%\fi}%
\bbl@languages%
\fi%
```

### 3.3 base

The first 'real' option to be processed is base, which set the hyphenation patterns then resets `ver@babel.sty` so that `ffpx` forgets about the first loading. After a subset of babel.def has been loaded (the old switch.def) and `AfterBabelLanguage` defined, it exits.

Now the base option. With it we can define (and load, with luatex) hyphenation patterns, even if we are not interested in the rest of babel.

```latex
\bbl@trace{Defining option 'base'}
\ifpackagewith{babel}{base}{}
\let\bbl@onlyswitch@empty
\let\bbl@provide@locale\relax
\input babel.def
\let\bbl@onlyswitch\@undefined
\ifx\directlua\@undefined
\DeclareOption*{\bbl@patterns{\CurrentOption}}%
\else
\input luababel.def
\DeclareOption*{\bbl@patterns@lua{\CurrentOption}}%
\fi
\input babel.sty
\input babel.sty
```
3.4 key=value options and other general option

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

\DeclareOption{KeepShorthandsActive}{}
\DeclareOption{activeacute}{}
\DeclareOption{activegrave}{}
\DeclareOption{debug}{}
\DeclareOption{noconfigs}{}
\DeclareOption{showlanguages}{}
\DeclareOption{silent}{}
\DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
\chardef\bbl@iniflag\z@
\DeclareOption{provide=*}{\chardef\bbl@iniflag\@ne} % main -> +1
\DeclareOption{provide+=*}{\chardef\bbl@iniflag\tw@} % add = 2
\DeclareOption{provide*=*}{\chardef\bbl@iniflag\thr@@} % add + main
% A separate option
\DeclareOption{selectors=off}{\bbl@singletrue}
\let\bbl@autoload@options\@empty
\DeclareOption{provide@=*}{\def\bbl@autoload@options{import}}% Don’t use. Experimental. TODO.
\newif\ifbbl@single
\DeclareOption{selectors=on}{\bbl@sectiontrue}
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\@nnil
\let\bbl@opt@config\@nnil
\let\bbl@opt@main\@nnil
\let\bbl@opt@headfoot\@nnil
\let\bbl@opt@layout\@nnil
\let\bbl@opt@provide\@nnil
```

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

```latex
\def\bbl@tempa#1=#2\bbl@tempa{\%
  \bbl@csarg\ifx{opt@#1}\@nnil
  \bbl@csarg\edef{opt@#1}{#2}%
  \else
  \bbl@error{bad-package-option}{#1}{#2}{}
  \fi}
```

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

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

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

```latex
\ProcessOptions*
```

```latex
\ifx\bbl@opt@provide\@nnil
  \let\bbl@opt@provide\@empty % %%% MOVE above
\else
  \chardef\bbl@iniflag\@ne
  \bbl@exp{\bl@forkv{\@nameuse{@raw@opt@babel.sty}}}{%\in@{,provide,}{,#1,}%
    \ifin@
      \def\bbl@opt@provide{#2}%
      \bbl@replace\bbl@opt@provide{;}{,}%
    \fi}
  \fi
```

### 3.5 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=...`.

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

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

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

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

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

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

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

```latex
\ifx\bbl@opt@headfoot\@nnil\else
\g@addto@macro\@resetactivechars{%\set@typeset@protect%%%%%%%%%%%%%%%%%%%%%\select@language@x\selectlanguage{\bbl@opt@headfoot}\\\let\protect\noexpand}\\\fi
```

For the option safe we use a different approach – \bbl@opt@safe says which macros are redefined (B for bibs and R for refs). By default, both are currently set, but in a future release it will be set to none.

```latex
\bbl@ifx\bbl@opt@safe@undefined
\def\bbl@opt@safe{BR}\\\let\bbl@opt@safe\@empty % Pending of \cite
\fi
```

For layout an auxiliary macro is provided, available for packages and language styles. Optimization: if there is no layout, just do nothing.

```latex
\bbl@trace{Defining IfBabelLayout}
\ifx\bbl@opt@layout\@nnil
\newcommand\IfBabelLayout[3]{#3}\\\else
\bbl@exp{\\\bbl@forkv{\@nameuse{@raw@opt@babel.sty}}}{\in@{,layout,}{,#1,}\\\ifin@\def\bbl@opt@layout{#2}\\\bbl@replace\bbl@opt@layout{ }{.}\\\fi}\\\newcommand\IfBabelLayout[1]{\\\inp{.}{#1.}{.\bbl@opt@layout.}}\\\ifin@\\\expandafter\@firstoftwo\\\else\\\expandafter\@secondoftwo\\\fi}\\\fi
```

⟨/package⟩
3.6 Interlude for Plain

Because of the way docstrip works, we need to insert some code for Plain here. However, the tools provided by the babel installer for literate programming makes this section a short interlude, because the actual code is below, tagged as *Emulate LaTeX*.

```latex
\ifx\ldf@quit\@undefined\else
\endinput\fi % Same line!
\langle⟨ Make sure ProvidesFile is defined \rangle⟩
\ProvidesFile{babel.def}[⟨⟨ date ⟩⟩\langle⟨ version ⟩⟩ Babel common definitions]
\ifx\AtBeginDocument\@undefined % TODO. change test.
\langle⟨ Emulate LaTeX \rangle⟩
\fi
\langle⟨ Basic macros ⟩⟩
```

That is all for the moment. Now follows some common stuff, for both Plain and \LaTeX. After it, we will resume the \TeX-only stuff.

444 ⟨/core⟩
445 ⟨+package | core⟩

4 Multiple languages

This is not a separate file (switch.def) anymore.
Plain \TeX version 3.0 provides the primitive \texttt{\language} that is used to store the current language. When used with a pre-3.0 version this function has to be implemented by allocating a counter.

```latex
\def\bbl@version{
\langle⟨ version \rangle⟩}
\def\bbl@date{
\langle⟨ date \rangle⟩}
\langle⟨ Define core switching macros \rangle⟩
```

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

```
\def\adddialect#1#2{%
  \global\chardef#1#2\relax
  \bbl@usehooks{adddialect}{{#1}{#2}}%
  \begingroup
    \count@#1\relax
    \def\bbl@elt##1##2##3##4{%
      \ifnum\count@=##2\relax
        \edef\bbl@tempa{\expandafter\@gobbletwo\string#1}%
        \bbl@info{Hyphen rules for '\expandafter\@gobble\bbl@tempa'\%
          \string\language\the\count@}. Reported}%
      \def\bbl@elt####1####2####3####4{}%
    \fi}%
  \bbl@cs{languages}%
  \endgroup}
```

\texttt{\bbl@iflanguage} executes code only if the language \texttt{l@} exists. Otherwise raises an error.
The argument of \texttt{\bbl@iflanguage} has to be a macro name, as it may get “fixed” if casing (lc/uc) is wrong. It’s an attempt to fix a long-standing bug when \texttt{\foreignlanguage} and the like appear in a \texttt{\MakeXXXcase}. However, a lowercase form is not imposed to improve backward compatibility (perhaps you defined a language named \texttt{MYLANG}, but unfortunately mixed case names cannot be trapped). Note \texttt{l@} is encapsulated, so that its case does not change.

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

```latex
\bbl@iflanguage
\langle⟨ Multiple languages \rangle⟩
```

13
After a name has been ‘fixed’, the selectors will try to load the language. If even the fixed name is not defined, will load it on the fly, either based on its name, or if activated, its BCP47 code.

We first need a couple of macros for a simple BCP 47 lookup. It also makes sure, with \bbl@bcp case, casing is the correct one, so that sr-latin-ba becomes fr-Latin-BA. Note #4 may contain some \@empty's, but they are eventually removed. \bbl@bcp lookup either returns the found ini or is \relax.

\def\bbl@bcpcase#1#2#3#4\@@#5{% #4 may contain some \@empty's, but they are eventually removed. \bbl@bcp lookup either returns the found ini or is \relax.
\ifx\@empty#3%
\uppercase{\def#5{#1#2}}%
\else
\uppercase{\def#5{#1}}%
\lowercase{\edef#5{#5#2#3#4}}%
\fi}

\def\bbl@bcplookup#1-#2-#3-#4\@@{% #4 may contain some \@empty's, but they are eventually removed. \bbl@bcp lookup either returns the found ini or is \relax.
\let\bbl@bcp\relax
\lowercase{\def\bbl@tempa{#1}}%
\ifx\@empty#2%
\IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
\else\ifx\@empty#3%
\bbl@bcpcase#2\@empty\@empty\@@\bbl@tempb
\IfFileExists{babel-\bbl@tempa-\bbl@tempb.ini}{}
\else\ifx\@empty#4%
\bbl@bcpcase#2\@empty\@empty\@@\bbl@tempb
\bbl@bcpcase#3\@empty\@empty\@@\bbl@tempc
\IfFileExists{babel-\bbl@tempa-\bbl@tempb-\bbl@tempc.ini}{}
\else\bbl@bcpcase#2\@empty\@empty\@@\bbl@tempb
\bbl@bcpcase#3\@empty\@empty\@@\bbl@tempc
\bbl@bcpcase#4\@empty\@empty\@@\bbl@tempd
\IfFileExists{babel-\bbl@tempa-\bbl@tempb-\bbl@tempc-\bbl@tempd.ini}{}
\else
\IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}
\fi
\fi
\fi
\fi
\let\bbl@initoload\relax
\let\bbl@auxname\languagename % Still necessary. TODO
\bbl@ifunset{bbl@bcp@map\languagename}{}% Move uplevel??
\edef\languagename{\@nameuse{bbl@bcp@map\languagename}}%
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.

4.1 Selecting the language

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

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`.

The following definition is preserved for backwards compatibility (e.g., `arabic`, `koma`). It is related to a trick for 2.09, now discarded.

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

But when the language change happens inside a group the end of the group doesn’t write anything to the auxiliary files. Therefore we need `\TeX`’s `\aftergroup` mechanism to help us. The command `\aftergroup` stores the token immediately following it to be executed when the current group is closed. So we define a temporary control sequence `\bbl@pop@language` to be executed at the end of the group. It calls `\bbl@set@language` with the name of the current language as its argument.
The previous solution works for one level of nesting groups, but as soon as more levels are used it is no longer adequate. For that case we need to keep track of the nested languages using a stack mechanism. This stack is called \bbl@language@stack and initially empty.

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

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:

\begin{verbatim}
\def\bbl@push@language{\ifx\languagename\@undefined\else\ifx\currentgrouplevel\@undefined\xdef\bbl@language@stack{\languagename+\bbl@language@stack}\else\ifnum\currentgrouplevel=\z@\xdef\bbl@language@stack{\languagename+}\else\xdef\bbl@language@stack{\languagename+\bbl@language@stack}\fi\fi\fi\fi}
\end{verbatim}

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.

\begin{verbatim}
\def\bbl@pop@lang#1+#2\@@{\edef\languagename{#1}\xdef\bbl@language@stack{#2}}\end{verbatim}

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).

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

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 \@... will be reserved for hyphenation patterns (so that two locales can share the same rules).

\begin{verbatim}
\chardef\localeid\z@\def\bbl@id@assign{% \bbl@ifunset{bbl@id@@\languagename}{\count@\bbl@id@last\relax\advance\count@\@ne\bbl@csarg\chardef{id@@\languagename}\count@\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 = '\language'\fi\fi\fi\fi}
\end{verbatim}
The unprotected part of `selectlanguage`. In case it is used as environment, declare \endselectlanguage, just for safety.

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

We also write a command to change the current language in the auxiliary files. \bbl@savelastskip is used to deal with skips before the write (whatsit (as suggested by U Fischer). Adapted from hyperref, but it might fail, so I'll consider it a temporary hack, while I study other options (the ideal, but very likely unfeasible except perhaps in \lualatex, is to avoid the \write altogether when not needed).

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

We also write a command to change the current language in the auxiliary files. \bbl@savelastskip is used to deal with skips before the write (whatsit (as suggested by U Fischer). Adapted from hyperref, but it might fail, so I'll consider it a temporary hack, while I study other options (the ideal, but very likely unfeasible except perhaps in \lualatex, is to avoid the \write altogether when not needed).
First, check if the user asks for a known language. If so, update the value of \language and call \originalTeX to bring TeX in a certain pre-defined state. The name of the language is stored in the control sequence \languagename. Then we have to redefine \originalTeX to compensate for the things that have been activated. To save memory space for the macro definition of \originalTeX, we construct the control sequence name for the \noextras{lang} command at definition time by expanding the \csname primitive. Now activate the language-specific definitions. This is done by constructing the names of three macros by concatenating three words with the argument of \selectlanguage, and calling these macros. The switching of the values of \lefthyphenmin and \righthyphenmin is somewhat different. First we save their current values, then we check if \noextras{lang} hyphenmins is defined. If it is not, we set default values (2 and 3), otherwise the values in \noextras{lang} hyphenmins will be used. No text is supposed to be added with switching captions and date, so we remove any spurious spaces with \bbl@bsphack and \bbl@esphack.
otherlanguage (en.) The otherlanguage environment can be used as an alternative to using the \selectlanguage declarative command. The \ignorespaces command is necessary to hide the environment when it is entered in horizontal mode.

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

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

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

otherlanguage* (en.) 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.

\expandafter\def\csname otherlanguage*\endcsname{% \ifnextchar[\bbl@otherlanguage@s\else\bbl@otherlanguage@s\]\fi\def\bbl@selectorname{other*}\ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi\def\bbl@select@opts{#1}\foreignlanguage{#2}%%}

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

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

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

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

(3.11) \foreignlanguage* is a temporary, experimental macro for a few lines with a different script direction, while preserving the paragraph format (thank the braces around \par, things like \hangindent are not reset). Do not use it in production, because its semantics and its syntax may change (and very likely will, or even it could be removed altogether). Currently it enters in vmode and then selects the language (which in turn sets the paragraph direction). (3.11) Also experimental are the hook foreign and foreign*. With them you can redefine \BabelText which by default does nothing. Its behavior is not well defined yet. So, use it in horizontal mode only if you do not want surprises.

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

\providecommand{\bbl@beforeforeign}{\edef{\foreignlanguage{}}{\noexpand{\protect{\expandafter\noexpand{\csname foreignlanguage \endcsname}}}\expandafter\def\csname foreignlanguage \endcsname{\%\noexpand{\protect{\expandafter\noexpand{\csname foreignlanguage \endcsname}}}}}\expandafter\def\csname foreignlanguage \endcsname{% \@ifstar{\bbl@foreign@x{\bbl@foreign@x{\bbl@foreign@x}}}\providecommand{\bbl@foreign@x[3]{}{}}{\begingroup{\def{\bbl@selectorname{foreign}}{}}{}}
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.

The following macro executes conditionally some code based on the selector being used.

This macro selects the hyphenation patterns by changing the \language register. If special hyphenation patterns are available specifically for the current font encoding, use them instead of the default. It also sets hyphenation exceptions, but only once, because they are global (here \language\lccode's has been set, too). \bbl@hyphenation@ is set to relax until the very first \babelhyphenation, so do nothing with this value. If the exceptions for a language (by its number, not its name, so that :ENC is taken into account) has been set, then use \hyphenation with both global and language exceptions and empty the latter to mark they must not be set again.
hyphenrules (em.) The environment hyphenrules can be used to select just the hyphenation rules. This environment does not change \languagename and when the hyphenation rules specified were not loaded it has no effect. Note however, \lcname's and font encodings are not set at all, so in most cases you should use otherlanguage*.

\def\hyphenrules#1{\edef\bbl@tempf{#1}\bbl@fixname\bbl@tempf\bbl@iflanguage\bbl@tempf{\expandafter\bbl@patterns\expandafter{\bbl@tempf}\ifx\languageshorthands\@undefined\else\languageshorthands{none}\fi\expandafter\expandafter\expandafter\set@hyphenmins\csname\bbl@tempf hyphenmins\endcsname\relax\set@hyphenmins{tw@th@}\relax\else\expandafter\expandafter\expandafter\set@hyphenmins\csname\bbl@tempf hyphenmins\endcsname\relax\fi}}\let\endhyphenrules\@empty

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

\def\providehyphenmins#1#2{\@namedef{#1hyphenmins}{#2}\fi}

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

\def\sethyphenmins#1#2{\lefthyphenmin#1\relax\righthyphenmin#2\relax}

ProvidesLanguage The identification code for each file is something that was introduced in \it XeLaTeX. When the command \ProvidesFile does not exist, a dummy definition is provided temporarily. For use in the language definition file the command \ProvidesLanguage is defined by babel. Depending on the format, ie, on if the former is defined, we use a similar definition or not.

\ifx\ProvidesFile\@undefined\def\ProvidesLanguage#1[#2 #3 #4]{\wlog{Language: #1 #4 #3 <#2>}}\fi
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.

Because this part of the code can be included in a format, we make sure that the macro which initializes the save mechanism, \babel@beginsave, is not considered to be undefined.

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

\providecommand\setlocale{bl@error{not-yet-available}{}{}{}}
\let\uselocale\setlocale
\let\locale\setlocale
\let\selectlocale\setlocale
\let\textlocale\setlocale
\let\textlanguage\setlocale
\let\languagetext\setlocale

4.2 Errors

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

\@noopterr When the package was loaded without options not everything will work as expected. An error message is issued in that case. When the format knows about \PackageError it must be \TeX\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}
\def\bbl@nocaption{i\string\protect\bbl@nocaption@i}
\def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
\global\@namedef{#2}{\textbf{?#1?}}\@nameuse{#2}\edef\bbl@tempa{#1}\bbl@sreplace\bbl@tempa{name}{}\bbl@warning{\@backslashchar#1 not set for '\languagename'. Please,\\ define it after the language has been loaded\(\%\) (typically in the preamble) with:\\ \string\setlocalecaption{\languagename}{\bbl@tempa}{..}\bbl@warning{% Reported}}\def\bbl@tentative{\protect\bbl@tentative@i}
\def\bbl@tentative@i#1{\bbl@warning{Some functions for '#1' are tentative.\%\ They might not work as expected and their behavior\%\ could change in the future.\%}
\def\@nolanerr#1{\bbl@error{undefined-language}{#1}{}{}}
\def\@nopatterns#1{\%
\bbl@warning{No hyphenation patterns were preloaded for\% the language '#1' into the format.\%
Please, configure your TeX system to add them and\%
rebuild the format. Now I will use the patterns\%
preloaded for \bbl@nulllanguage\space instead}
\let\bbl@usehooks\@gobbletwo
\ifx\bbl@onlyswitch\@empty\endinput\fi
% Here ended switch.def
Here ended the now discarded switch.def. Here also (currently) ends the base option.
\ifx\directlua\@undefined\else
\ifx\bbl@luapatterns\@undefined
\input luababel.def
\fi\fi
\bbl@trace{Compatibility with language.def}
\ifx\bbl@languages\@undefined
\ifx\directlua\@undefined
\openin1 = language.def % TODO. Remove hardcoded number
\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
\addto{It takes two arguments, a \textit{control sequence} and TeX-code to be added to the \textit{control sequence}. If the \textit{control sequence} has not been defined before it is defined now. The control sequence could also expand to \texttt{\relax}, in which case a circular definition results. The net result is a stack overflow. Note there is an inconsistency, because the assignment in the last branch is global.
\def\addto#1#2{\%
\ifx#1\@undefined
\def#1{#2}\
\else
\ifx#1\relax
\def#1{#2}\
\else
\toks\expandafter{#1#2}\
\xdef#1{\the\toks@}}\
\fi\fi
\fi
\\texttt{\initiate@active@char} below takes all the necessary actions to make its argument a shorthand character. The real work is performed once for each character. But first we define a little tool.
\def\bbl@withactive#1#2{\%
\begingroup

\\texttt{\addto}
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). A macro named \macro will be saved new control sequences named \orgmacro.

\def\bbl@redefine#1{\edef\bbl@tempa{\bbl@stripslash#1}\expandafter\let\csname org@\bbl@tempa\endcsname#1\expandafter\def\csname\bbl@tempa\endcsname}
\@onlypreamble\bbl@redefine
\bbl@redefine@long
This version of \babel@redefine can be used to redefine \long commands such as ifthenelse.

\def\bbl@redefinerobust#1{\edef\bbl@tempa{\bbl@stripslash#1}\bbl@ifunset{\bbl@tempa\space}{\expandafter\let\csname org@\bbl@tempa\endcsname#1\bbl@exp{\def\#1{\protect\bbl@tempa\space}}}\bbl@exp{\let\org@\bbl@tempa\bbl@tempa\space}}
\@onlypreamble\bbl@redefinerobust

4.3 Hooks

Admittedly, the current implementation is a somewhat simplistic and does very little to catch errors, but it is meant for developers, after all. \bbl@usehooks is the commands used by babel to execute hooks defined for an event.

\def\bbl@usehooks{\bbl@usehooks@lang\languagename}
\@onlypreamble\bbl@usehooks

\def\bbl@usehooks@lang#1#2#3{% Test for Plain
\ifx\UseHook\@undefined\else\UseHook{babel/*/#2}\fi
\def\bbl@elth##1{\bbl@cs{hk@##1}{\bbl@cs{ev@##1@#2@#1}#3}}
\bbl@cs{ev@#2@#1}%
\fi}
To ensure forward compatibility, arguments in hooks are set implicitly. So, if a further argument is added in the future, there is no need to change the existing code. Note events intended for hyphen.cfg are also loaded (just in case you need them for some reason).

```
1016 \def\bbl@evargs{% <- don't delete this comma
1017   everylanguage=1, loadkernel=1, loadpatterns=1, loadexceptions=1, %
1018   addddialect=2, patterns=2, defaultcommands=0, encodedcommands=2, write=0, %
1019   beforeextras=0, afterextras=0, stopcommands=0, stringprocess=0, %
1020   hyphenation=2, initiateactive=3, afterreset=0, foreign=0, foreign*=0, %
1021   beforerestart=0, languagename=2, begindocument=1}
1022 \ifx\NewHook\@undefined\else % Test for Plain (?)
1023 \def\bbl@tempa#1=#2\@@{
1024 \bbl@foreach\bbl@evargs\{
1025 \def\bbl@tempb##1{\ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}
1026 \edef\bbl@tempa{\bbl@tempb#1\@empty}
1027 \def\bbl@tempb##1=#2\@@{\@namedef{bbl@ens@##1}{##2}}
1028 \bbl@foreach\bbl@tempa\{
1029 \def\bbl@tempc{\bbl@ensure}
1030 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{\expandafter{\bbl@ens@include}}
1031 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{\expandafter{\bbl@ens@exclude}}
1032 \toks@\expandafter{\bbl@tempc\expandafter{}}}
1033 \fi}
1034 \bbl@trace{Defining babelensure}
1035 \newcommand\babelensure[2][3][]{% 1: include 2: exclude 3: fontenc
1036 \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
1037 \ifx##1\@undefined % 3.32 - Don't assume the macro exists
1038 \edef##1{\noexpand\bbl@nocaption{\bbl@stripslash##1}{\languagename\bbl@stripslash##1}}%
1039 \fi
1040 \ifx##1\@empty\else
1041 \in@{##1}{#2}%
1042 \ifin@%
1043 \bbl@ifunset{bbl@ensure@languagename}\
1044 {\DeclareRobustCommand\bbl@ensure@languagename[1]{%}
1045 \foreignlanguage{##1}{%}
1046 \fontencoding{#3}\
1047 \selectfont%
1048 \fi
1049 \fi
1050 \bbl@trace{The user command just parses the optional argument and creates a new macro named \bbl@ensure@language). We register a hook at the afterextras event which just executes this macro in a “complete” selection (which, if undefined, is \relax and does nothing). This part is somewhat involved because we have to make sure things are expanded the correct number of times. The macro \bbl@ensure@language contains \bbl@ensure{include}\{include\}\{exclude\}\{fontenc\}, which in in turn loops over the macros names in \bbl@captionslist, excluding (with the help of \in@ those in the exclude list. If the fontenc is given (and not \relax), the \fontencoding is also added. Then we loop over the include list, but if the macro already contains \foreignlanguage, nothing is done. Note this macro (1) is not restricted to the preamble, and (2) changes are local.
1051 \bbl@exp{\begin{verbatim}
1052 \bbl@trace{The user command just parses the optional argument and creates a new macro named \bbl@ensure@language). We register a hook at the afterextras event which just executes this macro in a “complete” selection (which, if undefined, is \relax and does nothing). This part is somewhat involved because we have to make sure things are expanded the correct number of times. The macro \bbl@ensure@language contains \bbl@ensure{include}\{include\}\{exclude\}\{fontenc\}, which in in turn loops over the macros names in \bbl@captionslist, excluding (with the help of \in@ those in the exclude list. If the fontenc is given (and not \relax), the \fontencoding is also added. Then we loop over the include list, but if the macro already contains \foreignlanguage, nothing is done. Note this macro (1) is not restricted to the preamble, and (2) changes are local.
1053 \bbl@exp{\end{verbatim}
1054 \endgroup}
1055 \edef\bbl@@language{\langle language\rangle).
1056 \bbl@trace{The user command just parses the optional argument and creates a new macro named \bbl@ensure@language). We register a hook at the afterextras event which just executes this macro in a “complete” selection (which, if undefined, is \relax and does nothing). This part is somewhat involved because we have to make sure things are expanded the correct number of times. The macro \bbl@ensure@language contains \bbl@ensure{include}\{include\}\{exclude\}\{fontenc\}, which in in turn loops over the macros names in \bbl@captionslist, excluding (with the help of \in@ those in the exclude list. If the fontenc is given (and not \relax), the \fontencoding is also added. Then we loop over the include list, but if the macro already contains \foreignlanguage, nothing is done. Note this macro (1) is not restricted to the preamble, and (2) changes are local.
1057 \bbl@exp{\end{verbatim}
1058 \endgroup}
1059 \edef\bbl@@language{\langle language\rangle).
1060 \bbl@trace{The user command just parses the optional argument and creates a new macro named \bbl@ensure@language). We register a hook at the afterextras event which just executes this macro in a “complete” selection (which, if undefined, is \relax and does nothing). This part is somewhat involved because we have to make sure things are expanded the correct number of times. The macro \bbl@ensure@language contains \bbl@ensure{include}\{include\}\{exclude\}\{fontenc\}, which in in turn loops over the macros names in \bbl@captionslist, excluding (with the help of \in@ those in the exclude list. If the fontenc is given (and not \relax), the \fontencoding is also added. Then we loop over the include list, but if the macro already contains \foreignlanguage, nothing is done. Note this macro (1) is not restricted to the preamble, and (2) changes are local.
1061 \bbl@exp{\end{verbatim}
1062 \endgroup}
1063 \edef\bbl@@language{\langle language\rangle).
1064 \bbl@trace{The user command just parses the optional argument and creates a new macro named \bbl@ensure@language). We register a hook at the afterextras event which just executes this macro in a “complete” selection (which, if undefined, is \relax and does nothing). This part is somewhat involved because we have to make sure things are expanded the correct number of times. The macro \bbl@ensure@language contains \bbl@ensure{include}\{include\}\{exclude\}\{fontenc\}, which in in turn loops over the macros names in \bbl@captionslist, excluding (with the help of \in@ those in the exclude list. If the fontenc is given (and not \relax), the \fontencoding is also added. Then we loop over the include list, but if the macro already contains \foreignlanguage, nothing is done. Note this macro (1) is not restricted to the preamble, and (2) changes are local.
1065 \fi
4.4 Setting up language files

\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.

\LdfInit \ldfInit \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.
This macro interrupts the processing of a language definition file.
\def\ldf@quit#1{%
  \expandafter\main@language\expandafter{#1}%
  \catcode`@=\atcatcode \let\atcatcode\relax
  \catcode`@=\eqcatcode \let\eqcatcode\relax
  \endinput
}\ldf@finish
This macro takes one argument. It is the name of the language that was defined in the language definition file.

We load the local configuration file if one is present, we set the main language (taking into account that the argument might be a control sequence that needs to be expanded) and reset the category code of the @-sign.

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

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

We also have to make sure that some code gets executed at the beginning of the document, either when the aux file is read or, if it does not exist, when the \AtBeginDocument is executed. Languages do not set \pagedir, so we set here for the whole document to the main \bodydir.
A bit of optimization. Select in heads/foots the language only if necessary.

\bbl@add@special

\bbl@remove@special

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

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

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

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

\begin{verbatim}
\def\bbl@active@def#1#2#3#4{% 
  \@namedef{#3#1}{% 
    \expandafter\ifx\csname#2@sh@#1@\endcsname\relax 
    \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1} \fi}% 
  \else \bbl@afterfi\csname#2@sh@#1@\endcsname \fi}}% 
\end{verbatim}

\texttt{\initiate@active@char} calls \texttt{\@initiate@active@char} with 3 arguments. Allof them are the same character with different catcodes: active, other (\texttt{\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 {}\fi}
\end{verbatim}

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\def\@namedef{#3#1}{% 
  \expandafter\ifx\csname#2#1@\endcsname\relax 
  \bbl@afterelse\bbl@sh@select#2#1{#3#1\string#1\#1} \else \bbl@afterfi\csname#2#1@\endcsname \fi}}% 
\end{verbatim}

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

\begin{verbatim}
\def\@initiate@active@char#1{% 
  \bbl@ififset{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 treatment to avoid making them \texttt{\relax} and preserving some degree of protection).

\begin{verbatim}
\def\@initiate@active@char#1#2#3{% 
  \bbl@csarg\edef\oricat@#2\{\catcode`#2=\the\catcode`#2\relax\}% 
  \ifx#1\undefined \else \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 \texttt{\normalchar (char)} to expand to the character in its default state. If the character is mathematically active when \texttt{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 \texttt{8000 a posteriori}).

To prevent problems with the loading of other packages after \texttt{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 .aux file. Otherwise some citations may give unexpected results in the printout when a shorthand was used in the optional argument of \texttt{\bibitem} for example. Then we make it active (not strictly necessary, but done for backward compatibility).

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

(where \active@char (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.

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

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

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.
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.

```latex
\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}
```

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 on \ifincsname is available. If there is, the expansion will be more robust.

```latex
\begingroup
\bbl@ifunset{ifincsname}% TODO. Ugly. Correct? Only Plain?
{\gdef\active@prefix#1{\ifx\protect\@typeset@protect\else\ifx\protect\@unexpandable@protect\noexpand#1\else\protect#1\fi\expandafter\@gobble\fi}}
{\gdef\active@prefix#1{\ifincsname\string#1\expandafter\@gobble\else\ifx\protect\@typeset@protect\else\ifx\protect\@unexpandable@protect\noexpand#1\else\protect#1\fi\expandafter\expandafter\expandafter\@gobble\fi\fi}}
\endgroup
```

In some circumstances it is necessary to be able to reset the shorthand to its 'normal' value (usually the character with catcode 'other') on the fly. For this purpose the switch @safe@actives is available. The setting of this switch should be checked in the first level expansion of \active\char{char}. When this expansion mode is active (with \@safe@activate, something like "13" becomes "12" in an \edef (in other words, shorthands are 'string'ed). This contrasts with \protected@edef, where catcodes are always left unchanged. Once converted, they can be used safely even after this expansion mode is deactivated (with \@safe@activetrue).

```latex
\if@safe@actives In some circumstances it is necessary to be able to reset the shorthand to its 'normal' value (usually the character with catcode 'other') on the fly. For this purpose the switch @safe@actives is available. The setting of this switch should be checked in the first level expansion of \active\char{char}. When this expansion mode is active (with \@safe@activetrue, something like "13" becomes "12" in an \edef (in other words, shorthands are 'string'ed). This contrasts with \protected@edef, where catcodes are always left unchanged. Once converted, they can be used safely even after this expansion mode is deactivated (with \@safe@activetrue).

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

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.

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

\chardef\bbl@activated\z@
\def\bbl@activate#1{%
  \chardef\bbl@activated@ne
  \bbl@withactive{\expandafter\let\expandafter}{#1}%
  \csname bbl@active@\string#1\endcsname}
\def\bbl@deactivate#1{%
  \chardef\bbl@activated@tw@
  \bbl@withactive{\expandafter\let\expandafter}{#1}%
  \csname bbl@normal@\string#1\endcsname}

These macros are used only as a trick when declaring shorthands.

\def\bbl@firstcs#1#2{\csname#1\endcsname}
\def\bbl@scndcs#1#2#3#4{\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.

The auxiliary macro \babel@texpdf improves the interoperativity with hyperref and takes 4 arguments: (1) The TeX code in text mode, (2) the string for hyperref, (3) the TeX code in math mode, and (4), which is currently ignored, but it's meant for a string in math mode, like a minus sign instead of an hyphen (currently hyperref doesn't discriminate the mode). This macro may be used in \ldf files.

\def\bbl@tempa{#3}\
\ifx\bbl@tempa\@empty
  \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs\bbl@ifunset{#1@sh@\string#2@}{%}
\else
  \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa
  \else
    \bbl@info{Redefining #1 shorthand \string#2\string#3\% in language \CurrentOption}%
  \fi
  \@namedef{#1@sh@\string#2@\string#3@}{#4}%
\fi
\else
  \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs\bbl@ifunset{#1@sh@\string#2}@{}{%}
\else
  \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
  \else
    \bbl@info{Redefining #1 shorthand \string#2\string#3\% in language \CurrentOption}%
  \fi
  \@namedef{#1@sh@\string#2@\string#3@}{#4}%
\fi
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.

\textormath{\ifmmode\expandafter\@secondoftwo\else\expandafter\@firstoftwo\fi}

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

\def\user@group{user}
\def\language@group{english} % TODO. I don't like defaults
\def\system@group{system}

This is the user level macro. It initializes and activates the character for use as a shorthand character (ie, it's active in the preamble). Languages can deactivate shorthands, so a starred version is also provided which activates them always after the language has been switched.

\def\useshorthands{\@ifstar\bbl@usesh@s{\bbl@usesh@x{}}}
\def\bbl@usesh@s#1{\bbl@usesh@x{\AddBabelHook{babel-sh-\string#1}{afterextras}{\bbl@activate{#1}}}{#1}}
\def\bbl@usesh@x#1#2{\bbl@ifshorthand{#2}{\def\user@group{user}\initiate@active@char{#2}{#1}\bbl@activate{#2}}{\bbl@error{shorthand-is-off}{}{#2}{}}}

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

\def\bbl@set@user@generic#1#2{\bbl@ifunset{user@generic@active#1}{\bbl@active@def#1\user@language@group{user@active}{user@generic@active}\bbl@active@def#1\user@group{user@generic@active}{language@active}\expandafter\edef\csname#2@sh@#1@\protect@\endcsname{\expandafter\noexpand\csname normal@char#1\endcsname}\expandafter\edef\csname#2@sh@#1@\string\protect\endcsname{\noexpand\csname user@active#1\endcsname}}\@empty}}
\newcommand\defineshorthand[3][user]{\edef\bbl@tempa{\zap@space#1 \@empty}\bbl@for\bbl@tempb\bbl@tempa{\if*\expandafter\@car\bbl@tempb\@nil\edef\bbl@tempb{user@\expandafter\@gobble\bbl@tempb}\@expandtwoargs\bbl@set@user@generic{\expandafter\string\@car#2\@nil}\bbl@tempb}{\bbl@error{shorthand-is-off}{}{#2}{#3}}}}

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 to fix it in the same way languages names are fixed. [TODO].

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

Deprecated. 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 latter to \active@char".

\def\aliasshorthand#1#2{%
  \bbl@ifshorthand{#2}%
  {\expandafter\ifx\csname active@char\string#2\endcsname\relax
   \@notshorthand{#2}%
  {\ifx\document\@notprerr
   \else
   \begin{active@char}{#2}%
   \bbl@ccarg\let{active@char\string#2}{active@char\string#1}%
   \bbl@ccarg\let{normal@char\string#2}{normal@char\string#1}%
   \bbl@activate{#2}%
   \fi
   \fi}%

\@notshorthand

\def\@notshorthand#1\bbl@error{not-a-shorthand}{#1}{}{}

\shorthandon The first level definition of these macros just passes the argument onto \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#2\@nnil\else
  \bbl@ifunset{bbl@active@char\string#2}%
  {\bbl@error{not-a-shorthand-b}{}{#2}{}%}
  {\ifcase#1% off, on, off*
   \catcode`#212\relax
   \or
   \catcode`#2\active
   \bbl@ifunset{bbl@shdef@char\string#2}%
   {\bbl@withactive{\expandafter\let\expandafter}{#2}%%
    \csname bbl@shdef@char\string#2\endcsname
    \bbl@csarg\let{shdef@char\string#2}{}%
    \ifcase\bbl@activated
    \or
    \bbl@activate{#2}%
    \else
    \bbl@deactivate{#2}%
    \fi
    \fi
    \or
    \bbl@ifunset{bbl@shdef@char\string#2}%
    {\bbl@withactive{\bbl@csarg\let{shdef@char\string#2}}#2}%
    {\csname bbl@ori@char\string#2\endcsname
     \csname bbl@ori@def@char\string#2\endcsname
    }%
    \fi}
  \afterfi{\bbl@switch@sh#1%}
  \fi
}
Note the value is that at the expansion time; eg, in the preamble shorthands are usually deactivated.

\def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
\def\bbl@putsh#1{% 
  {\bbl@ifunset{bbl@active@\string#1}{\bbl@putsh@i#1\@empty\@nnil}{\csname bbl@active@\string#1\endcsname}}
\def\bbl@putsh@i#1#2\@nnil{% 
  \csname\language@group @sh@\string#1@\ifx\@empty#2\else\string#2\fi\endcsname}
\ifx\bbl@opt@shorthands\@nnil\else
  \let\bbl@s@initiate@active@char\initiate@active@char
  \def\initiate@active@char#1{% 
    \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
  \let\bbl@s@switch@sh\bbl@switch@sh
  \def\bbl@switch@sh#1#2{% 
    \ifx#2\@nnil\else\bbl@afterfi\bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}\bbl@switch@sh#1\fi\fi}
  \let\bbl@s@activate\bbl@activate
  \def\bbl@activate#1{% 
    \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
  \let\bbl@s@deactivate\bbl@deactivate
  \def\bbl@deactivate#1{% 
    \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
\fi

You may want to test if a character is a shorthand. Note it does not test whether the shorthand is on or off.
\newcommand\ifbabelshorthand[3]{\bbl@ifunset{bbl@active@\string#1}{#3}{#2}}
\bbl@prim@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.

\def\bbl@prim@s{\prime\futurelet\@let@token\bbl@pr@m@s}
\def\bbl@if@primes#1#2{% 
  \ifx#1\@let@token\expandafter\@firstoftwo\else\ifx#2\@let@token\bbl@afterelse\expandafter\@firstoftwo\else\bbl@afterfi\expandafter\expandafter\@secondoftwo\fi\fi
\begingroup
  \catcode`^=7 \catcode`*=12 \catcode`\^=12 \lccode`\^=`\^ \lccode`\*=\^ \lowercase%
  \gdef\bbl@pr@m@s{% 
    \bbl@if@primes\pr@@@s
    {\bbl@if@primes\pr@@@t\egroup}}
\endgroup

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).
\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

4.6 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.

\if\bbl@known@attribs\@undefined
\in@false
\else
\bbl@xin@{,\bbl@tempc-##1,}{,\bbl@known@attribs,}
\ifin@
\bbl@warning{You have more than once selected the attribute '##1' for language \bbl@tempc. Reported}\
\else
When we end up here the attribute is not selected before. So, we add it to the list of selected attributes and execute the associated \TeX-code.
\bbl@exp{% \bbl@add@list\bbl@known@attribs{\bbl@tempc-##1}}%
\edef\bbl@tempa{\bbl@tempc-##1}\
\expandafter\bbl@ifknown@trib\expandafter{\bbl@tempa}\bbl@attributes{%\csname\bbl@tempc @attr@##1\endcsname}\
\else\@attrerr{\bbl@tempc}{##1}\fi}}
\@onlypreamble\languageattribute

The error text to be issued when an unknown attribute is selected.
\newcommand*{\@attrerr}[2]{{\bbl@error{unknown-attribute}{#1}{#2}{}}}

\bbl@declare@tribute This command adds the new language/attribute combination to the list of known attributes.

Then it defines a control sequence to be executed when the attribute is used in a document. The result of this should be that the macro \extras... for the current language is extended, otherwise the attribute will not work as its code is removed from memory at \begin{document}.

38
\def\bbl@declare@attribute#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{\
  \ifx\bbl@known@attribs\@undefined\in@false\else\bbl@xin{,#1-#2,}{,\bbl@known@attribs,}\fi\
  \ifin@\bbl@afterelse#3\else\bbl@afterfi#4\fi}

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

\def\bbl@ifknown@trib#1#2{\
  \let\bbl@tempa\@secondoftwo\bbl@loopx\bbl@tempb{#2}{\
    \expandafter\in@\expandafter{\expandafter,\bbl@tempb,}{,#1,}\
    \ifin@\let\bbl@tempa\@firstoftwo\else\fi}\bbl@tempa}

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

\def\bbl@clear@tribs{\
  \ifx\bbl@attributes\@undefined\else\bbl@loopx\bbl@tempa{\bbl@attributes}{\bbl@loopx\bbl@tembp{\bbl@attributes}{\bbl@clear@tribs}}}\

4.7 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.
The initialization of a new save cycle: reset the counter to zero.
\def\babel@beginsave\{\edef\babel@savecnt\z@\}

Before it’s forgotten, allocate the counter and initialize all.
\newcount\babel@savecnt
\def\babel@beginsave\{\edef\babel@savecnt\z@\}

The macro \texttt{\babel@save\{csname\}} saves the current meaning of the control sequence \texttt{\langle csname\rangle} to \originalTeX. To do this, let the current meaning to a temporary control sequence, the restore commands are appended to \originalTeX and the counter is incremented. The macro \texttt{\babel@savevariable\{variable\}} saves the value of the variable. \texttt{\langle variable\rangle} can be anything allowed after the \texttt{\the} primitive. To avoid messing saved definitions up, they are saved only the very first time.
\def\babel@save#1\{\def\bbl@tempa{,#1,}\expandafter\bbl@add\expandafter\bbl@tempa\expandafter{,\bbl@savedextras,}\expandafter\in@\bbl@tempa\ifin@\else\bbl@add\bbl@savedextras{,#1,}\bbl@carg\let{babel@\number\babel@savecnt}#1\relax\toks@\expandafter{\originalTeX\let#1=}\bbl@exp{\def\originalTeX{\the	oks@\langle babel@\number\babel@savecnt\rangle\relax}}\advance\babel@savecnt\@ne\fi\}
\def\babel@savevariable#1\{\toks@\expandafter{\originalTeX #1=}\bbl@exp{\def\originalTeX{\the	oks@\the#1\relax}}\}

Some languages need to have \texttt{\frenchspacing} in effect. Others don't want that. The command \texttt{\bbl@frenchspacing} switches it on when it isn’t already in effect and \texttt{\bbl@nonfrenchspacing} switches it off if necessary. A more refined way to switch the catcodes is done with ini files. Here an auxiliary macro is defined, but the main part is in \texttt{\babelprovide}. This new method should be ideally the default one.
\def\bbl@frenchspacing\{\ifnum\the\sfcode`\string.\@m\let\bbl@nonfrenchspacing\relax\else\frenchspacing\let\bbl@nonfrenchspacing\nonfrenchspacing\fi\}
\let\bbl@nonfrenchspacing\nonfrenchspacing
\let\bbl@elt\relax
\edef\bbl@fs@chars{\bbl@elt{\string.}\@m{3000}\bbl@elt{\string?}\@m{3000}\bbl@elt{\string!}\@m{3000}\bbl@elt{\string:}\@m{2000}\bbl@elt{\string;}\@m{1500}\bbl@elt{\string,}\@m{1250}}\def\bbl@pre@fs{\bbl@save@sfcodes\edef\bbl@tempa{\bbl@cl{frspc}}\edef\bbl@tempa{\expandafter\@car\bbl@tempa\@nil}\if u\bbl@tempa % do nothing\else\if n\bbl@tempa % non french\bbl@elt\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\string\str
4.8 Short tags
\babeltags
This macro is straightforward. After zapping spaces, we loop over the list and define the macros \text{⟨tag⟩} and ⟨tag⟩. Definitions are first expanded so that they don't contain \csname but the actual macro.

4.9 Hyphens
\babelhyphenation
This macro saves hyphenation exceptions. Two macros are used to store them: \bbl@hyphenation@ for the global ones and \bbl@hyphenation<lang> for language ones. See \bbl@patterns above for further details. We make sure there is a space between words when multiple commands are used.
This macro makes hyphenation possible. Basically its definition is nothing more than \nobreak \\
hskip 0pt plus 0pt².

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

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

\newcommand\babelnullhyphen{\char\hyphenchar\font}
\def\babelhyphen{\active@prefix\babelhyphen\bbl@hyphen}
\def\bbl@hyphen{i\@empty}
\def\bbl@hyphen@i#1#2{\bbl@ifunset{bbl@hy@#1#2\@empty}{\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}\bbl@hy@#1#2\@empty\@empty}

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

The following macro inserts the hyphen char.

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

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

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

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

\def\bbl@disc@#1@2{\nobreak\discretionary{#2}{#1}{#1}}

²pX begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.
4.10 Multiencoding strings

The aim following commands is to provide a common interface for strings in several encodings. They also contain several hooks which can be used by luatex and xetex. The code is organized here with pseudo-guards, so we start with the basic commands.

**Tools** But first, a tool. It makes global a local variable. This is not the best solution, but it works.

```latex
\bbl@trace{Multiencoding strings}
\def\bbl@toglobal#1{\global\let#1#1}
```

The following option is currently no-op. It was meant for the deprecated \SetCase.

```latex
\langle\langle∗ More package options ⟨/ More package options⟩⟩ ={\DeclareOption{nocase}{}}\rangle
```

The following package options control the behavior of \SetString.

```latex
\langle\langle∗ More package options ⟨/ More package options⟩⟩ =\let\bbl@opt@strings\@nnil % accept strings=value
\langle\langle∗ More package options ⟨/ More package options⟩⟩ ={\let\bbl@opt@strings\relax}
\langle\langle∗ More package options ⟨/ More package options⟩⟩ ={\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.

```latex
\@onlypreamble\StartBabelCommands
\def\StartBabelCommands{%
\begingroup
@tempcnta="7F
\def\bbl@tempa{%
\ifnum@tempcnta="FF\else
\catcode@tempcnta=11
\advance@tempcnta=\\ne
\expandafter\bbl@tempa
\fi}\bbl@tempa
\langle\langleMacros local to BabelCommands\rangle⟩
\def\bbl@provstring##1##2{%\providecommand##1{##2}%
\bbl@toglobal##1}\global\let\bbl@scafter\@empty
\let\StartBabelCommands\bbl@startcmds
\ifx\BabelLanguages\relax
\let\BabelLanguages\CurrentOption
\fi\begingroup
\ifstar{\ifx\bbl@opt@strings\@nnil
\let\bbl@opt@strings\BabelStringsDefault
\fi\bbl@startcmds@i}%
\bbl@startcmds@i}
```

```latex
\def\bbl@startcmds@i#1#2{%\edef\bbl@L{\zap@space#1 \@empty}%
\def\bbl@opt@strings\@nnil % local flag - disable 1st stopcommands
\StartBabelCommands}
```

```
\def\bbl@startcmds{%
\ifx\bbl@opt@strings\@nnil
\def\bbl@startcmds@i{%\let\bbl@opt@strings\BabelStringsDefault
\fi
\bbl@startcmds@i}
\\bbl@startcmds@i}
\def\bbl@startcmds@i#1#2{%
\edef\bbl@L{\zap@space#1 \@empty}%
```

43
Parse the encoding info to get the label, input, and font parts. Select the behavior of \SetString. There are two main cases, depending on if there is an optional argument: without it and strings=encoded, strings are defined always; otherwise, they are set only if they are still undefined (i.e., 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 (i.e., 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 \(\text{date}⟨\text{language}\rangle\) is defined (after babel has been loaded). There are also two versions of \bbl@forlang. The first one skips the current iteration if the language is not in \BabelLanguages (used in ldfs), and the second one skips undefined languages (after babel has been loaded).

\def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{}{#2}}}\def\bbl@scswitch{\bbl@forlang\bbl@tempa{\ifx\bbl@G\@empty\else\ifx\SetString\@gobbletwo\else\edef\bbl@GL{\bbl@G\bbl@tempa}\bbl@xin@{,\bbl@GL,}{,\bbl@screset,}\ifin@\else\global\expandafter\let\csname\bbl@GL\endcsname\@undefined\xdef\bbl@screset{\bbl@screset,\bbl@GL}\fi\fi\fi}}\AtEndOfPackage{\def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@setstring#1#2{%\edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}\bbl@ifunset{\bbl@LC}{\bbl@exp{\global\bbl@add\\bbl@G\bbl@tempa\langle\bbl@scset\#1\bbl@LC\rangle}}}}}%\endgroup\let\bbl@scafter\relax\let\bbl@endcommands\EndBabelCommands\let\EndBabelCommands\bbl@scafter\let\bbl@endcommands\EndBabelCommands\endgroup\bbl@scafter\endgroup}

Now we define commands to be used inside \StartBabelCommands.

**Strings**  
The following macro is the actual definition of \SetString when it is “active” First save the “switcher”. Create it if undefined. Strings are defined only if undefined (i.e., like \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.

\def\bbl@setstring#1#2{% eg, \prefacename{<string>}\bbl@forlang\bbl@tempa{%\edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}\bbl@ifunset{\bbl@LC}{\bbl@exp%\global\bbl@add\langle\bbl@G\bbl@tempa\langle\\bbl@scset\#1\langle\bbl@LC\rangle\rangle\rangle\}}}%\endgroup\let\bbl@setstring\bbl@setstringdef\bbl@BabelString\bbl@BabelString{#2}\def\bbl@usehooks{stringprocess}{%\expandafter\bbl@stringdef\csname\bbl@LC\expandafter\endsname\expandafter{\BabelString}}}A little auxiliary command sets the string. TODO: Formerly used with casing. Very likely no longer necessary, although it's used in \setlocalecaption.
Define \texttt{\SetStringLoop}, which is actually set inside \texttt{\StartBabelCommands}. The current definition is somewhat complicated because we need a count, but \texttt{\count@} is not under our control (remember \texttt{\SetString} may call hooks). Instead of defining a dedicated count, we just “pre-expand” its value.

\[
\langle\langle\text{Macros local to BabelCommands}\rangle\rangle\equiv
\def\bbl@scset#1#2{\def#1{#2}}
\]

\[
\def\bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
\advance\count@\@ne
\toks@\expandafter{\bbl@tempa}%
\bbl@exp{\
\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
\count@=\the\count@\relax}}%
\]

\[
\langle\langle/\text{Macros local to BabelCommands}\rangle\rangle
\]

Delaying code  Now the definition of \texttt{\AfterBabelCommands} when it is activated.

\[
\langle\langle\text{Macros local to BabelCommands}\rangle\rangle
\]

Case mapping  The command \texttt{\SetCase} is deprecated. Currently it consists in a definition with a hack just for backward compatibility in the macro mapping.

\[
\langle\langle\text{Macros local to BabelCommands}\rangle\rangle\equiv
\newcommand\SetCase[3][]{%
\def\bbl@tempa####1####2{%
\ifx####1\@empty\else
\bbl@carg\bbl@add{extras\CurrentOption}{%
\bbl@carg\babel@save{c__text_uppercase_\string####1_tl}%
\bbl@carg\def{c__text_uppercase_\string####1_tl}{{####2}}%
\bbl@carg\babel@save{c__text_lowercase_\string####2_tl}%
\bbl@carg\def{c__text_lowercase_\string####2_tl}{{####1}}}%
\expandafter\bbl@tempa
\fi}%
\expandafter\bbl@tempa\@empty\@empty%\bbl@carg\bbl@toglobal{extras\CurrentOption}}%
\langle\langle/\text{Macros local to BabelCommands}\rangle\rangle
\]

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.

\[
\langle\langle\text{Macros local to BabelCommands}\rangle\rangle\equiv
\newcommand\SetHyphenMap[1][]{%
\bbl@forlang\bbl@tempa{%
\expandafter\bbl@stringdef\csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
\langle\langle/\text{Macros local to BabelCommands}\rangle\rangle
\]

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

\[
\langle\langle\text{Macros local to BabelCommands}\rangle\rangle\equiv
\newcommand\BabelLower[2][1]{% one to one.
\ifnum\lccode#1=#2\else
\babel@savevariable{\lccode#1}%
\lccode#1=#2\relax
\fi}
\newcommand\BabelLowerMM[4][1]{% many-to-many
\@tempcnta=#1\relax
\@tempcntb=#4\relax
\def\bbl@tempa{%
\ifnum\@tempcnta>#2\else
\@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
\advance\@tempcnta#3\relax
\fi}
\langle\langle/\text{Macros local to BabelCommands}\rangle\rangle
\]

\[
\text{46}
\]
The following package options control the behavior of hyphenation mapping.

\begin{itemize}
\item \texttt{hyphenmap=off}\texttt{\chardef\bbl@opt@hyphenmap\z@}
\item \texttt{hyphenmap=first}\texttt{\chardef\bbl@opt@hyphenmap@\@ne}
\item \texttt{hyphenmap=select}\texttt{\chardef\bbl@opt@hyphenmap@\tw@}
\item \texttt{hyphenmap=other}\texttt{\chardef\bbl@opt@hyphenmap@\thr@@}
\item \texttt{hyphenmap=other*}\texttt{\chardef\bbl@opt@hyphenmap4\relax}
\end{itemize}

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

\AtEndOfPackage{%}
\ifdef\bbl@opt@hyphenmap@\@undefined
\bbl@xin@{,}{\bbl@language@opts}%
\chardef\bbl@opt@hyphenmap@\ifin@4\else\@ne\fi
\fi}

This sections ends with a general tool for resetting the caption names with a unique interface. With the old way, which mixes the switcher and the string, we convert it to the new one, which separates these two steps.

\newcommand\setlocalecaption{% TODO. Catch typos.
\@ifstar\bbl@setcaption@s\bbl@setcaption@x}
\def\bbl@setcaption@x#1#2#3{% language caption-name string
\bbl@trim@def\bbl@tempa{#2}%
\bbl@xin@{.template}{\bbl@tempa}%
\ifin@
\bbl@ini@captions@template{#3}{#1}%
\else
\edef\bbl@tempd{%
\strip@prefix\expandafter\meaning\csname captions#1\endcsname}%
\bbl@xin@{\string\csname #2name\endcsname}{\bbl@tempd}%
\ifin@ % Renew caption
\bbl@xin@{\string\bbl@scset}{\bbl@tempd}%
\else % Old way converts to new way
\bbl@\@ifundefined{#1\#2name}%
\bbl@\exp{\bbl@ifsamestring{\bbl@tempa}{\languagename}%
\bbl@\ifsamestring{\bbl@scset\#2name}{\#1\#2name}}%
\bbl@\iffalse%
\else % New way sets new language
\edef\bbl@tempa{%
\bbl@ifundefined{\bbl@opt@hyphenmap\#2}{%
\edef\bbl@opt@hyphenmap\#2{}%}
\fi%
\fi}

\end{document}
4.11 Macros common to a number of languages

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

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

4.12 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.

4.12.1 Quotation marks

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

\quotesinglbase
We also need the single quote character at the baseline.
Make sure that when an encoding other than OT1 or T1 is used this glyph can still be typeset.

\provideTextCommandDefault{\quotesinglbase}{%}
\useTextSymbol{OT1}{\quotesinglbase}

\guillemetleft The guillemet characters are not available in OT1 encoding. They are faked. (Wrong names with o preserved for compatibility.)
\guillemetright

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

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

\provideTextCommand{\guilsinglleft}{OT1}{%}
\ifmmode
<%\fi
\provideTextCommand{\guilsinglright}{OT1}{%}
\ifmmode
>%\fi

Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.
4.12.2 Letters
\ij The Dutch language uses the letter ‘ij’. It is available in T1 encoded fonts, but not in the OT1 encoded fonts. Therefore we fake it for the OT1 encoding.
\IJ

\DeclareTextCommand{\ij}{OT1}{i\kern-0.02em\bbl@allowhyphens j}
\DeclareTextCommand{\IJ}{OT1}{I\kern-0.02em\bbl@allowhyphens J}
\DeclareTextCommand{\ij}{T1}{\char188}
\DeclareTextCommand{\IJ}{T1}{\char156}

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

\Providetextcommanddefault{\ij}{% \Usetextsymbol{OT1}{\ij}}
\Providetextcommanddefault{\IJ}{% \Usetextsymbol{OT1}{\IJ}}
\dj The Croatian language needs the letters \dj and \Dj; they are available in the T1 encoding, but not in the OT1 encoding by default.

Some code to construct these glyphs for the OT1 encoding was made available to me by Stipčević Mario, (stipcevic@olimp.irb.hr).
\def\crrtic@{\hrule height0.1ex width0.3em}
\def\crttic@{\hrule height0.1ex width0.33em}
\def\ddj@{\setbox0\hbox{d}\dimen@=\ht0
\advance\dimen@1ex \dimen@.45\dimen@
\dimen@ii\expandafter\rem@pt\the\fontdimen@\font\dimen@
\leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
\def\DDJ@{\setbox0\hbox{D}\dimen@=.55\ht0
\dimen@ii\expandafter\rem@pt\the\fontdimen@\font\dimen@
\advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
\dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
\leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
\DeclareTextCommand{\dj}{OT1}{\ddj@ d}
\DeclareTextCommand{\Dj}{OT1}{\DDJ@ D}

Make sure that when an encoding other than OT1 or T1 is used these glyphs can still be typeset.
\Providetextcommanddefault{\dj}{% \Usetextsymbol{OT1}{\dj}}
\Providetextcommanddefault{\Dj}{% \Usetextsymbol{OT1}{\Dj}}
\SS For the T1 encoding \SS is defined and selects a specific glyph from the font, but for other encodings it is not available. Therefore we make it available here.
\DeclareTextCommand{\SS}{OT1}{SS}
\Providetextcommanddefault{\SS}{% \Usetextsymbol{OT1}{\SS}}

4.12.3 Shorthands for quotation marks

Shorthands are provided for a number of different quotation marks, which make them usable both outside and inside math mode. They are defined with \Providetextcommanddefault, but this is very likely not required because their definitions are based on encoding-dependent macros.
\SSS
The ‘german’ single quotes.

\textbackslashProvideTextCommandDefault{\glq}{
\text.enquotebase{mbox{quoteringbase}}}

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

\textenquote{T1}{
\textquoteleft{mbox{textquoteleft}}}

\textenquote{TU}{
\textquoteleft{mbox{textquoteleft}}}

\textenquote{OT1}{
\textquoteleft{mbox{textquoteleft}}% 
\kern.07em}

\textProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}

The ‘german’ double quotes.

\textbackslashProvideTextCommandDefault{\glqq}{
\text.enquotebase{mbox{quoteringbase}}}

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

\textenquote{T1}{
\textquotedblleft{mbox{textquotedblleft}}}

\textenquote{TU}{
\textquotedblleft{mbox{textquotedblleft}}}

\textenquote{OT1}{
\textquotedblleft{mbox{textquotedblleft}}% 
\kern.07em}

\textProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}

The ‘french’ single guillemets.

\textbackslashProvideTextCommandDefault{\flq}{
\text.guilsinglleft{mbox{guilsinglleft}}}

\textbackslashProvideTextCommandDefault{\frq}{
\text.guilsinglright{mbox{guilsinglright}}}

The ‘french’ double guillemets.

\textbackslashProvideTextCommandDefault{\flqq}{
\text.guillemetleft{mbox{guillemetleft}}}

\textbackslashProvideTextCommandDefault{\frqq}{
\text.guillemetright{mbox{guillemetright}}}

4.12.4 Umlauts and tremas

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

\umlauthigh
\umlaute

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

\umlaute
\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 \langle dimen \rangle register.

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

Then we compute the new x-height in such a way that the umlaut character is lowered to the base character. The value of .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.

For all vowels we declare " to be a composite command which uses \bble@umlauta or \bble@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 \bble@umlauta and/or \bble@umlaute for a language in the corresponding ldf (using the babel switching mechanism, of course).

Finally, make sure the default hyphen rules are defined (even if empty). For internal use, another empty \language is defined. Currently used in Amharic.

4.13 Layout

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

Some macros are not defined in all engines, so, after loading the files define them if necessary to raise an error.

4.15 Creating and modifying languages

Continue with \texttt{\LaTeX} only.
\texttt{\LaTeX} provide 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.
\let\bbl@extend@ini\@gobble
\newcommand\babelprovide[2][]{%
\let\bbl@savelangname\languagename
\edef\bbl@savelocaleid{\the\localeid}%
% Set name and locale id
\edef\languagename{#2}%
\bbl@id@assign
% Initialize keys
\bbl@vforeach{captions, date, import, main, script, language,%
hyphenrules, linebreaking, justification, mapfont, maparabic,%
mapdigits, intraspace, intrapenalty, onchar, transforms, alph,%
Alph, labels, labels*, calendar, date, casing, interchar}%
\bbl@csarg\let{KVP@##1}\@nnil%
\global\let\bbl@release@transforms\@empty
\global\let\bbl@release@casing\@empty
\let\bbl@calendars\@empty
\global\let\bbl@inidata\@empty
\global\let\bbl@extend@ini\@gobble
\global\let\bbl@included@inis\@empty
\gdef\bbl@key@list{;}
\bbl@forkv{#1}{%
\in@{/}{##1}% With /, (re)sets a value in the ini
\ifin@
\global\let\bbl@extend@ini\bbl@extend@ini@aux
\bbl@renewinikey##1\@@{##2}%
\else
\bbl@csarg\ifx{KVP@##1}\@nnil\else
\bbl@error{unknown-provide-key}{##1}{}{}
\fi
\bbl@csarg\def{KVP@##1}{##2}%
\fi}%
\chardef\bbl@howloaded=0\none; 1:ldf without ini; 2:ini
\ifcase\bbl@howloaded
\let\bbl@lbkflag\@empty % new
\else
\if\bbl@KVP@hyphenrules\@nnil\else
\let\bbl@lbkflag\@empty
\fi
\if\bbl@KVP@import\@nnil\else
\let\bbl@lbkflag\@empty
\fi
\fi%
\chardef\bbl@howloaded=0\none; 1:ldf without ini; 2:ini
\if\bbl@KVP@date\@nnil\else
\fi%
% == init ==
\if\bbl@screset\@undefined
\bbl@ldfinit
\fi%
% == date (as option) ==
% \if\bbl@KVP@date\@nnil
% \fi%
% ==
\let\bbl@lbkflag\relax % @empty = do setup linebreak, only in 3 cases:
% \ifcase\bbl@howloaded
\let\bbl@lbkflag\@empty % new
\else
\if\bbl@KVP@hyphenrules\@nnil\else
\let\bbl@lbkflag\@empty
\fi
\if\bbl@KVP@import\@nnil\else
\let\bbl@lbkflag\@empty
\fi
\fi%
% == import, captions ==
% \if\bbl@KVP@import\@nnil
% \fi%
% ==
\let\bbl@lbkflag\relax % @empty = do setup linebreak, only in 3 cases:
% \ifcase\bbl@howloaded
\let\bbl@lbkflag\@empty % new
\else
\if\bbl@KVP@hyphenrules\@nnil\else
\let\bbl@lbkflag\@empty
\fi
\if\bbl@KVP@import\@nnil\else
\let\bbl@lbkflag\@empty
\fi
\fi%
% == import, captions ==
% \if\bbl@KVP@import\@nnil
% \fi%
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 set the basic parameters, but still loading the whole ini file.

```latex
{\let\bbl@KVP@date\@empty
{\let\bbl@KVP@captions@\bbl@KVP@captions} % TODO. A dirty hack
{\ifx\bbl@KVP@captions\@nnil
{\let\bbl@KVP@captions\bbl@KVP@import}
{\fi
{\ifx\bbl@KVP@transforms\@nnil\else
{\bbl@replace\bbl@KVP@transforms{{}{{}}}%
{\ifx\bbl@KVP@transforms\@nnil
{\let\bbl@KVP@transforms\bbl@KVP@import}
{\fi
{\ifcase\bbl@howloaded
{\bbl@provide@new{#2}}%
{\else
{\bbl@ifblank{#1}%
{\{\bbl@load@basic below
{\{\bbl@provide@renew{#2}}%
{\fi
{\ifx\bbl@KVP@transforms\@nnil
{\bbl@replace\bbl@KVP@transforms{ }{,}%
{\fi
{\ifcase\bbl@howloaded
{\bbl@load@basic{#2}%
{\% == Load ini ==
{\ifx\bbl@included@inis\@empty\else
{\bbl@replace\bbl@included@inis{ }{,}%
{\bbl@foreach\bbl@included@inis{%
{\openin\bbl@readstream=babel-##1.ini
{\bbl@extend@ini{#2}}%
{\closein\bbl@readstream
{\fi
{\% == ensure captions ==
{\ifx\bbl@KVP@captions\@nnil
{\bbl@ifunset{bbl@extracaps@#2}%
{\bbl@ifunset{bbl@ensure@\languagename}%
{\bbl@exp{\natex@bl@ ensuring[exclude=\	oday,include=\bbl@extracaps@#2]}%}
{\bbl@exp{\
\DeclareRobustCommand\bbl@ensure@\languagename[1]{%}
{\\foreignlanguage{\languagename}\text{####1}}}}%
{\bbl@exp{\
\bbl@ toglobal\bbl@ ensuring@\languagename}
{\bbl@ toglobal\bbl@ ensuring@\languagename\space}%
{\fi
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 set the basic parameters, but still loading the whole ini file.

{\bbl@load@basic{#2}%
{\% == script, language ==
{\ifx\bbl@KVP@script\@nnil
{\bbl@csarg\edef{sname@#2}{\bbl@KVP@script}%
{\fi
{\ifx\bbl@KVP@language\@nnil
{\bbl@csarg\edef{lname@#2}{\bbl@KVP@language}%
{\fi
{\ifcase\bbl@engine\or
55
```
\def\bbl@mapselect{\
\bbl@mapdir{\languagename}}% 
\fi 
\bbl@exp{\\bbl@add\\bbl@mapselect{\\bbl@mapdir{\languagename}}}% 
\fi 
% TODO - catch non-valid values 
\fi 
% For bidi texts, to switch the font based on direction 
\ifx\bbl@KVP@mapfont\@nnil\else 
\bbl@ifsamestring{\bbl@KVP@mapfont}{direction}{}% 
\bbl@ifndef{\bbl@lsys@\languagename}{\bbl@provide@lsys{\languagename}}{}% 
\bbl@ifndef{\bbl@wdir@\languagename}{\bbl@provide@dirs{\languagename}}{}% 
\ifx\bbl@mapselect\@undefined % TODO. See onchar. 
\AtBeginDocument{% 
\bbl@patchfont{{\bbl@mapselect}}% 
{\selectfont}}% 
\def\bbl@mapselect{% 
\let\bbl@mapselect\relax 
\edef\bbl@prefontid{\fontid\font}}% 
\def\bbl@mapdir##1{% 
{\def\languagename{##1}% 
\let\bbl@ifrestoring\@firstoftwo % avoid font warning 
\bbl@switchfont 
\directlua{Babel.fontmap\[
\the\csname bbl@wdir@##1\endcsname\]%,
\bbl@prefontid}}% 
\fi 
\bbl@exp{\\bbl@add\\bbl@mapselect{\\bbl@mapdir{\languagename}}}% 
\fi 
% == Line breaking: intraspace, intrapenalty == 
% For CJK, East Asian, Southeast Asian, if interspace in ini 
\ifx\bbl@KVP@intraspace\@nnil\else % We can override the ini or set 
\bbl@csarg\edef{intsp@#2}{\bbl@KVP@intraspace}% 
\fi 
\bbl@provide@intraspace 
% == Line breaking: CJK quotes == TODO -> @extras 
\ifcase\bbl@engine\or 
\bbl@xin@[/c]{/\bbl@cl{l\lbrk}}% 
\ifi
\bbl@ifundef{\bbl@quote@\languagename}{\bbl@quote@\languagename}{}% 
{"directlua{ 
Babel.locale_props[the\localeid].cjk_quotes = {} 
local cs = 'op' 
for c in string.utfvalues( 
[[\csname bbl@quote@\languagename\endcsname]]) do 
if Babel.cjk_characters[c].c == 'qu' then 
Babel.locale_props[the\localeid].cjk_quotes[c] = cs 
end 
end 
\bbl@quote|cl|op 
end 
}}% 
\fi 
% == Line breaking: justification == 
\ifx\bbl@KVP@justification\@nnil\else 
\let\bbl@KVP@linebreaking\bbl@KVP@justification 
\fi 
\ifx\bbl@KVP@linebreaking\@nnil\else 
\bbl@xin@,.\bbl@KVP@linebreaking,\bbl@KVP@linebreaking,\bbl@KVP@linebreaking,
Depending on whether or not the language exists (based on \date<language>), we define two macros. Remember \bbl@startcommands opens a group.

\def\bbl@provide@new#1{%
  \@namedef{date#1}{}% marks lang exists - required by \StartBabelCommands
  \@namedef{extras#1}{}%
  \@namedef{noextras#1}{}%
  \bbl@startcommands*{#1}{captions}%
  \ifx\bbl@KVP@captions\@nnil % and also if import, implicit
    \def\bbl@provide@new#1{% marks lang exists - required by \StartBabelCommands
      \@namedef{date#1}{}% marks lang exists - required by \StartBabelCommands
      \@namedef{extras#1}{}%
      \@namedef{noextras#1}{}%
      \bbl@startcommands*{#1}{captions}%
      \ifx\bbl@KVP@captions\@nnil % and also if import, implicit
      \else
      \fi
    }%}
  \fi}
%
\def\bbl@provide@renew#1{%
  \ifx\bbl@KVP@captions\@nnil
    \StartBabelCommands*{#1}{captions}%
    \bbl@read@ini{\bbl@KVP@captions}2% % Here letters cat = 11
  \else
    \StartBabelCommands*{#1}{captions}%
    \bbl@read@ini{\bbl@KVP@captions}2% % Here letters cat = 11
  \fi
%
\StartBabelCommands*{#1}{date}%
  \bbl@load@basic{#1}%
  \% == hyphenmins == (only if new)
  \ifx\bbl@KVP@captions\@nnil
    \gdef\<#1hyphenmins>{%$
      \ifx\bbl@ifunset{\bbl@lfthm@#1}{2}{\bbl@cs{lfthm@#1}}}%
      \ifx\bbl@ifunset{\bbl@rgthm@#1}{3}{\bbl@cs{rgthm@#1}}}%
      \% == hyphenrules (also in renew) ==
      \bbl@provide@hyphens{#1}%
      \ifx\bbl@KVP@main\@nnil
      \else
      \fi
      \expandafter\main@language\expandafter{#1}%
      \StartBabelCommands*{#1}{captions}%
      \bbl@read@ini{\bbl@KVP@captions}2% % Here all letters cat = 11
      \fi
}
\EndBabelCommands
\fi
\ifx\bbl@KVP@date\@nnil\else
\StartBabelCommands*{#1}{date}\bl@savetoday\bl@savedate\EndBabelCommands\fi
% == hyphenrules (also in new) ==
\ifx\bbl@lbkflag\@empty
\bbl@provide@hyphens{#1}\i}

Load the basic parameters (ids, typography, counters, and a few more), while captions and dates are left out. But it may happen some data has been loaded before automatically, so we first discard the saved values. (TODO. But preserving previous values would be useful.)

\def\bbl@load@basic#1{%
\ifcase\bbl@howloaded\or\or
\ifcase\csname bbl@llevel@\languagename\endcsname
\bbl@csarg\let{lname@\languagename}\relax
\fi
\bbl@ifunset{bbl@lname@#1}{}%\bbl@read@ini{##1}1%
\ifx\bbl@initoload\relax\endinput\fi
\endgroup}
\ifx\bbl@initoload\relax
\bbl@input@texini{#1}%
\else
\setbox\z@\hbox{\BabelBeforeIni{\bbl@initoload}{}%}
\bbl@read@ini{##1}1%
\ifx\bbl@initoload\relax\endinput\fi
\endgroup
\endgroup%
\begingroup % boxed, to avoid extra spaces:
\ifx\bbl@initoload\relax
\bbl@input@texini{#1}%
\else
\setbox\z@\hbox{\BabelBeforeIni{\bbl@initoload}{}%}
\bbl@read@ini{##1}1%
\ifx\bbl@initoload\relax\endinput\fi
\endgroup
\endgroup%
\endgroup%
\begingroup % boxed, to avoid extra spaces:
\ifx\bbl@initoload\relax
\bbl@input@texini{#1}%
\else
\setbox\z@\hbox{\BabelBeforeIni{\bbl@initoload}{}%}
\bbl@read@ini{##1}1%
\ifx\bbl@initoload\relax\endinput\fi
\endgroup
\endgroup%

The hyphenrules option is handled with an auxiliary macro. This macro is called in three cases:
when a language is first declared with \babelprovide, with hyphenrules and with import.

\def\bbl@provide@hyphens#1{%
\@tempcnta\m@ne % a flag
\ifx\bbl@KVP@hyphenrules\@nnil\else
\bbl@replace\bbl@KVP@hyphenrules{ }{,}\bl@foreach\bbl@KVP@hyphenrules{%
\ifnum\@tempcnta=\m@ne % if not yet found
\bbl@ifsamestring{##1}{+}%
\bbl@carg\addlanguage{l@##1}%
}\bbl@ifunset{l@##1}{}% After a possible +
\bbl@ifsamestring{##1}{+}%
\bbl@IFS@l@##1}%
\bbl@ifunset{l@##1}{}%
\bbl@ifsamestring{##1}{+}%
\bbl@ifnum\@tempcnta=\m@ne
\bbl@warning{\ Requested \texttt{hyphenrules}\ for \texttt{\languagename}\ not found:\\%\bbl@KVP@hyphenrules.\%
Using the default value. Reported}%
\fi
\fi
\ifnum\@tempcnta=\m@ne % if no opt or no language in opt found
\ifx\bbl@KVP@captions@\@nnil \TODO. Hackish. See above.
\bbl@ifunset{bbl@hyphr@#1}{}% use value in ini, if exists
\fi

\fi
\fi
\fi
The reader of babel-...tex files. We set temporarily some catcodes.

\def\bbl@input@texini#1{\bbl@bsphack\bbl@exp{\catcode`\\%=14 \catcode`\\\=0 \catcode`\\{=1 \catcode`\\}=2 \lowercase{\InputIfFileExists{babel-#1.tex}{}{}} \catcode`\\%=	he\catcode`\% \catcode`\\\=\the\catcode`\\\relax \catcode`\\{=	he\catcode`\\} \catcode`\\}=\the\catcode`\\} \bbl@esphack}

The following macros read and store ini files (but don’t process them). For each line, there are 3 possible actions: ignore if starts with ;; switch section if starts with [, and store otherwise. There are used in the first step of \bbl@read@ini.

Now, the ‘main loop’, which **must be executed inside a group**. At this point, \bbl@inidata may contain data declared in \bbl@provide, with ‘slashed’ keys. There are 3 steps: first read the ini file and store it; then traverse the stored values, and process some groups if required (date, captions, labels, counters); finally, ‘export’ some values by defining global macros (identification, typography, characters, numbers). The second argument is 0 when called to read the minimal data for fonts; with \bbl@provide it’s either 1 or 2.
A variant to be used when the ini file has been already loaded, because it's not the first \babel\provide for this language.
A somewhat hackish tool to handle calendar sections. TODO. To be improved.

A key with a slash in \babelprovide replaces the value in the ini file (which is ignored altogether). The mechanism is simple (but suboptimal): add the data to the ini one (at this point the ini file has not yet been read), and define a dummy macro. When the ini file is read, just skip the corresponding key and reset the macro (in \bbl@inistore above).
The previous assignments are local, so we need to export them. If the value is empty, we can provide a default value.

2895 \def\bbl@exportkey{i2#3}{\%}
2896 \bbl@ifunset{bbl@@kv@i\%}{\%}
2897 {\bbl@csarg\def[i\%]{\languagename}{\%}{}\%}
2898 {\expandafter\ifx\csname bbl@@kv@i\%\endcsname\empty
2899 \bbl@exp{\global\let\bbl@\%\languagename\bbl@@kv@i\%}{}\%}
2900 \else
2901 \bbl@exp{\global\let\bbl@\%\languagename\bbl@@kv@i\%=\%}{}\%
2902 \fi}\%}

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@ini@exports is called always (via \bbl@inisec), while \bbl@after@ini must be called explicitly after \bbl@read@ini if necessary. Although BCP 47 doesn't treat ‘-x-’ as an extension, the CLDR and many other sources do (as a private use extension). For consistency with other single-letter subtags or ‘singletons’, here is considered an extension, too.

2903 \def\bbl@iniwarning{i\%}
2904 \bbl@ifunset{bbl@@kv@identification.warningi\%}{}\%
2905 \ifcase\bbl@engine
2906 \bbl@iniwarning{.pdflatex}{}\%
2907 \or \bbl@iniwarning{.lualatex}{}\%
2908 \or \bbl@iniwarning{.xelatex}{}\%
2909 \fi\%
2910 \let\bbl@release@transforms@empty\%
2911 \let\bbl@release@casing@empty\%
2912 \def\bbl@ini@exports{i\%}
2913 \% Identification always exported
2914 \bbl@iniwarning{}\%
2915 \ifcase\bbl@engine\%
2916 \bbl@iniwarning{.pdflatex}{}\%
2917 \or \bbl@iniwarning{.lualatex}{}\%
2918 \or \bbl@iniwarning{.xelatex}{}\%
2919 \fi\%
2920 \bbl@exportkey[llevel]{identification.load.level}{\%}
2921 \bbl@exportkey[elname]{identification.name.english}{\%}
2922 \bbl@exp{\bbl@exportkey[lname]{identification.name.opentype}{\%}
2923 {\csname bbl@elname@\languagename\endcsname}}\%
2924 \bbl@exportkey[tbcp]{identification.tag.bcp47}{\%}
2925 % Somewhat hackish. TODO:
2926 \bbl@exportkey[cbcp]{identification.tag.bcp47}{\%}
2927 \bbl@exportkey[bcp]{identification.language.tag.bcp47}{\%}
2928 \bbl@exportkey[lotf]{identification.tag.opentype}{DFLT}%
2929 \bbl@exportkey[esname]{identification.script.name}{\%}
2930 \bbl@exp{\bbl@exportkey[sname]{identification.script.name.opentype}{\%}
2931 {\csname bbl@esname@\languagename\endcsname}}\%
2932 \bbl@exportkey[sbcp]{identification.script.tag.bcp47}{\%}
2933 \bbl@exportkey[sbcp]{identification.script.tag.bcp47}{\%}
2934 \bbl@exportkey[soft]{identification.script.tag.opentype}{DFLT}%
2935 \bbl@exportkey[rbcp]{identification.region.tag.bcp47}{\%}
2936 \bbl@exportkey[vbcp]{identification.variant.tag.bcp47}{\%}
2937 \bbl@exportkey[extt]{identification.extension.t.tag.bcp47}{\%}
2938 \bbl@exportkey[extu]{identification.extension.u.tag.bcp47}{\%}
2939 \bbl@exportkey[extx]{identification.extension.x.tag.bcp47}{\%}
2940 % Also maps bcp47 -> languagename
2941 \if\bbl@bcp47@\%
2942 \bbl@csarg\xdef{bcp@map@\bbl@cl{\{tbcp\}}}{\languagename}\%
2943 \fi
2944 \ifcase\bbl@engine\%
2945 \directlua{Babel.locale_props[\the\bbl@cs{id@\languagename}].script = '\bbl@cl{sbcp}'}\%

65
A shared handler for key=val lines to be stored in \bbl@@kv@<section>.<key>.

By default, the following sections are just read. Actions are taken later.

The characters section also stores the values, but casing is treated in a different fashion. Much like transforms, a set of commands calling the parser are stored in \bbl@release@casing, which is executed in \babelprovide.

Additive numerals require an additional definition. When .1 is found, two macros are defined – the basic one, without .1 called by \localenumeral, and another one preserving the trailing .1 for the 'units'.

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.

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

```latex
\texttt{\begin{verbatim}
\def\bbl@ini@captions#1#2{% string language tempa=capt-name
  \def\bbl@replace{\bbl@tempa}{#1}\%\ifin@#1\else#2\fi
  \expandafter\expandafter\global\expandafter\let\expandafter\bbl@tempa
\endverbatim}
```
Labels. Captions must contain just strings, no format at all, so there is new group in ini files.

\def\bbl@list@the{%
  part, chapter, section, subsection, subsubsection, paragraph, %
  subparagraph, enumi, enumii, enumiii, enumiv, equation, figure, %
  table, page, footnote, mpfootnote, mpfn}
\def\bbl@map@cnt#1{% #1: roman, etc, // #2: enumi, etc
  \bbl@ifunset{bbl@map@#1@languagename}{{@nameuse{#1}}}{{@nameuse{bbl@map@#1@languagename}}}
}\def\bbl@inikv@labels#1#2{%
  \in@{.map}{#1}\
  \ifin@
    \ifx\bbl@KVP@labels@languagename\@nnil\else
      \bbl@xin@{ map }{ \bbl@KVP@labels \space}\
      \ifin@
        \def\bbl@tempc{#1}\
        \bbl@replace\bbl@tempc{.map}{}\
        \in@{,#2,}{arabic, roman, Roman, alph, Alph, fnsymbol,}\
        \\bbl@exp{\gdef\bbl@map@\bbl@tempc @languagename{%\localecounter{#2}}}}\
      \\bbl@foreach\bbl@list@the{\bbl@ifunset{the##1}{}{\bbl@exp{\
        \bbl@add<\bbl@extracaps@languagename>{\bbl@tempa name}}}\
        \\bbl@toglobal<\bbl@extracaps@languagename>}{\bbl@tempd}{\bbl@map@cnt{\bbl@tempd}{##1}}}\
      \expandafter\ifx\csname the##1\endcsname\bbl@tempd\else
        \toks@\expandafter\expandafter\expandafter{\csname the##1\endcsname}\
        \expandafter\if\csname the##1\endcsname\expandafter{\bbl@map@cnt{\bbl@tempd}{##1}}
        \expandafter\if\csname the##1\endcsname\expandafter{\bbl@map@cnt{\bbl@tempd}{##1}}
      \fi
    \else
      % The following code is still under study. You can test it and make
      % suggestions. Eg. enumerate.2 = ([enumi]).([enumii]). It's
      % language dependent.
      \in@{enumerate.}{#1}\
      \ifin@
        \def\bbl@tempa{#1}\
        \bbl@replace\bbl@tempa{enumerate.}{}\
        \def\bbl@toreplace{#2}\
        \bbl@replace\bbl@toreplace{[ ]}{\csname the}\
        \bbl@replace\bbl@toreplace{[}{\csname the}\
        \bbl@replace\bbl@toreplace{]}{\endcsname{}}\
        \expandafter\if\csname the##1\endcsname\expandafter{\bbl@map@cnt{\bbl@tempd}{##1}}
        \expandafter\if\csname the##1\endcsname\expandafter{\bbl@map@cnt{\bbl@tempd}{##1}}
      \fi
    \fi
  \fi
}%
\def\bbl@map@cnt{enumerate.}{#1}\
\def\bbl@map@enumerate.roman#1#2{\bbl@map@cnt{\bbl@enumerate.roman@#1}{#2}}\
\def\bbl@enumerate.roman@#1#2{\bbl@map@enumerate.roman@#1{#2}}\
\def\bbl@enumerate.roman@#1{#2}{\bbl@enumerate.roman@#1{#2}}\
\def\bbl@enumerate.roman@#1{#2}{\bbl@enumerate.roman@#1{#2}}\fi
\def\bbl@list@the{%
  part, chapter, section, subsection, subsubsection, paragraph, %
  subparagraph, enumi, enumii, enumiii, enumiv, equation, figure, %
  table, page, footnote, mpfootnote, mpfn}
\def\bbl@map@cnt#1{% #1: roman, etc, // #2: enumi, etc
  \bbl@ifunset{bbl@map@#1@languagename}{{@nameuse{#1}}}{{@nameuse{bbl@map@#1@languagename}}}
}\def\bbl@inikv@labels#1#2{%
  \in@{.map}{#1}\
  \ifin@
    \ifx\bbl@KVP@labels@languagename\@nnil\else
      \bbl@xin@{ map }{ \bbl@KVP@labels \space}\
      \ifin@
        \def\bbl@tempc{#1}\
        \bbl@replace\bbl@tempc{.map}{}\
        \in@{,#2,}{arabic, roman, Roman, alph, Alph, fnsymbol,}\
        \\bbl@exp{\gdef\bbl@map@\bbl@tempc @languagename{%\localecounter{#2}}}}\
      \\bbl@foreach\bbl@list@the{\bbl@ifunset{the##1}{}{\bbl@exp{\
        \bbl@add<\bbl@extracaps@languagename>{\bbl@tempa name}}}\
        \\bbl@toglobal<\bbl@extracaps@languagename>}{\bbl@tempd}{\bbl@map@cnt{\bbl@tempd}{##1}}}\
      \expandafter\if\csname the##1\endcsname\expandafter{\bbl@map@cnt{\bbl@tempd}{##1}}
      \expandafter\if\csname the##1\endcsname\expandafter{\bbl@map@cnt{\bbl@tempd}{##1}}
    \else
      % The following code is still under study. You can test it and make
      % suggestions. Eg. enumerate.2 = ([enumi]).([enumii]). It's
      % language dependent.
      \in@{enumerate.}{#1}\
      \ifin@
        \def\bbl@tempa{#1}\
        \bbl@replace\bbl@tempa{enumerate.}{}\
        \def\bbl@toreplace{#2}\
        \bbl@replace\bbl@toreplace{[ ]}{\csname the}\
        \bbl@replace\bbl@toreplace{[}{\csname the}\
        \bbl@replace\bbl@toreplace{]}{\endcsname{}}\
        \expandafter\if\csname the##1\endcsname\expandafter{\bbl@map@cnt{\bbl@tempd}{##1}}
        \expandafter\if\csname the##1\endcsname\expandafter{\bbl@map@cnt{\bbl@tempd}{##1}}
      \fi
    \fi
  \fi
}%
\def\bbl@map@cnt{enumerate.}{#1}\
\def\bbl@map@enumerate.roman#1#2{\bbl@map@cnt{\bbl@enumerate.roman@#1}{#2}}\
\def\bbl@enumerate.roman@#1#2{\bbl@map@enumerate.roman@#1{#2}}\
\def\bbl@enumerate.roman@#1{#2}{\bbl@enumerate.roman@#1{#2}}\fi
\def\bbl@enumerate.roman@#1{#2}{\bbl@enumerate.roman@#1{#2}}
To show correctly some captions in a few languages, we need to patch some internal macros, because
the order is hardcoded. For example, in Japanese the chapter number is surrounded by two string,
while in Hungarian is placed after. These replacement works in many classes, but not all. Actually,
the following lines are somewhat tentative.

\def\bbl@chaptype{chapter}
\ifx@makechapterhead\undefined
 \let\bbl@patchchapter\relax
\else\ifx thechapter \undefined
 \let\bbl@patchchapter\relax
\else\ifx ps@headings \undefined
 \let\bbl@patchchapter\relax
\else
 \def\bbl@patchchapter{\global\let\bbl@patchchapter\relax
 \gdef\bbl@chfmt{\bbl@ifunset{bbl@bbl@chaptype fmt@\languagename}{\@chapapp\space\thechapter}{\@nameuse{bbl@bbl@chaptype fmt@\languagename}}}
 \bbl@add\appendix{\def\bbl@chaptype{appendix}}% Not harmful, I hope
 \bbl@sreplace\ps@headings{\@chapapp\ 	hechapter}{\bbl@chfmt}
 \bbl@sreplace\chaptermark{\@chapapp\ 	hechapter}{\bbl@chfmt}
 \bbl@sreplace@makechapterhead{\@chapapp\space\thechapter}{\bbl@chfmt}
 \bbl@toglobal\appendix
 \bbl@tgloba...
\begin{lstlisting}[language=TeX]
brforeach@tempb##1\@@}%
brreplace@ld@calendar{gregorian}{}%
\ifx\bbl@ld@calendar\@empty\else
  \ifx\bbl@ld@convert\relax\else
    \babelcalendar[\bbl@th-bl@th-bl@th]\bbl@th\bbl@th\bbl@th
  \fi
\fi
\@nameuse{bbl@precalendar}% Remove, eg, +, -civil (-ca-islamic)
edef\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
\bbl@trim@def\bbl@tempa{#1.#2}%
\bbl@ifsamestring{\bbl@tempa}{months.wide}% to savedate
  \bbl@trim\toks@{#3}%
  \@temptokena\expandafter{\bbl@savedate}%
  \bbl@exp{% Reverse order - in ini last wins
    \def\bbl@savedate{\bbl@savedate{\SetString\month\roman{#5}{#6}}%}
  }%
\bbl@ifsamestring{\bbl@tempa}{date.long}% defined now
  {\lowercase{\def\bbl@tempb{#6}}%\bbl@trim\toks@{#5}%
    \bbl@TG@date
    \global\@nameuse{bbl@date\languagename @\bbl@calendar}%
    \bbl@th\bbl@th\bbl@th
    \bbl@exp% TDOO. Move to a better place.
    \AfterBabelCommands{%
      \def<\languagename date>{\protect<\languagename date >}%
      \newcommand<\languagename date >[4][]{%\bbl@usesdategrouptrue
        \@ensure\languagename>=%
        \localdate[####1][####2][####3][####4]{\bbl@savedate{#4}{#3}{#2}{#1}}}}%
    \bbl@exp\bbl@savedate%
    \bbl@TG@date
    \global\@nameuse{bbl@date\languagename @\bbl@tempb}\bbl@toreplace
  \ifx\bbl@savetoday\@empty
    }% 

\let\bbl@calendar\@empty
\newcommand\babelcalendar[2]{\the\year-\the\month-\the\day}{%\bbl@th\bbl@th\bbl@th%
  \@nameuse{bbl@ca@#2}#1\@@}
\newcommand\BabelDateSpace{\nobreakspace}
\newcommand\BabelDateDot{.\@} % TDOO. let instead of repeating
\newcommand\BabelDated[1]{\ifnum#1<10 0\fi\number#1}
\newcommand\BabelDateM[1]{\ifnum#1<10 0\fi\number#1}
\newcommand\BabelDateMM[1]{\number#1}
\newcommand\BabelDateMMM[1]{\number#1}
\end{lstlisting}

\textbf{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. Note after \bbl@replace \toks@ contains the resulting string, which is used by \bbl@replace@finish@iii (this implicit behavior doesn't seem a good idea, but it's efficient).

\begin{verbatim}
\let\bbl@calendar\@empty
\newcommand\babelcalendar[2]\the\year-\the\month-\the\day{%\bbl@th\bbl@th\bbl@th%
  \@nameuse{bbl@ca@#2}#1\@@}
\newcommand\BabelDateSpace{\nobreakspace}
\newcommand\BabelDateDot{.\@} % TDOO. let instead of repeating
\newcommand\BabelDated[1]{\ifnum#1<10 0\fi\number#1}
\newcommand\BabelDateM[1]{\ifnum#1<10 0\fi\number#1}
\newcommand\BabelDateMM[1]{\number#1}
\newcommand\BabelDateMMM[1]{\number#1}
\end{verbatim}

70
\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{limit-two-digits}{}{}{}%  
\fi\fi\fi\fi}}%  
\newcommand\BabelDateyyyy[1]{{\number#1}} % TODO - add leading 0  
\newcommand\BabelDateU[1]{{\number#1}}%  
\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}}\%  
\bbl@replace\bbl@toreplace{[U]}{\BabelDateU{####1}}\%  
\bbl@replace\bbl@toreplace{[y|}{\bbl@datecntr[####1|}\%  
\bbl@replace\bbl@toreplace{[U|}{\bbl@datecntr[####1|}\%  
\bbl@replace\bbl@toreplace{[m|}{\bbl@datecntr[####2|}\%  
\bbl@replace\bbl@toreplace{[d|}{\bbl@datecntr[####3|}\%  
\bbl@replace@finish@iii\bbl@toreplace}%  
\def\bbl@datecntr{\expandafter\bbl@xdatecntr\expandafter}  
\def\bbl@xdatecntr[#1|#2]{\localenumeral{#2}{#1}}\%  
\bbl@csarg\let{inikv@transforms.prehyphenation}\bbl@inikv\%  
\bbl@csarg\let{inikv@transforms.posthyphenation}\bbl@inikv\%  
\def\bbl@transforms@aux#1#2#3,#4,#5\relax{\%  
#1[#2]{#3}{#4}{#5}}\%  
\begingroup % A hack. TODO. Don’t require an specific order  
\catcode\%=12\%  
\catcode\&=14\%  
\gdef\bbl@transforms#1#2{&\%  
\directlua{\%  
local str = \[[==\[#2]==\]\]\%  
str = str:gsub('%.%d+%.%d+$', '')\%  
token.set_macro('babeltempa', str)\%  
}\%  
\def\babeltempc{}&\%  
\bbl@xin{,\babeltempa,}{,\bbl@KVP@transforms,}&\%  
\ifin@ &\%  
\bbl@xin{:\babeltempa,}{,\bbl@KVP@transforms,}&\%  
\fi\%  
\ifin@\%  
\bbl@foreach\bbl@KVP@transforms{&\%  
\bbl@xin{:\babeltempa,}{,##1,}&\%  
\ifin@ &% font:font:transform syntax  
\directlua{\%  
local t = {}\%  
for m in string.gmatch('##1'..'','') do\%  
table.insert(t, m)\%  
end\%  
\bbl@foreach\bbl@KVP@transforms{&\%  
\bbl@xin{:\babeltempa,}{,##1,}&\%  
\ifin@ &% font:font:transform syntax  
\directlua{\%  
local t = {}\%  
for m in string.gmatch('##1'..'','') do\%  
table.insert(t, m)\%  
end\%  
Transforms.
table.remove(t)
    token.set_macro('babeltempc', ',fonts=' .. table.concat(t, ' '))
    }
  
\endgroup

Language and Script values to be used when defining a font or setting the direction are set with the following macros.
\def\bbl@provide@lsys#1{%
  \bbl@ifunset{bbl@lname@#1}{\bbl@load@info{#1}}{}
  \bbl@csarg\let{lsys@#1}@empty
  \bbl@ifunset{bbl@sname@#1}{\bbl@csarg\gdef{sname@#1}{Default}}{}
  \bbl@ifunset{bbl@sotf@#1}{\bbl@csarg\gdef{sotf@#1}{DFLT}}{}
  \bbl@csarg\bbl@add@list{lsys@#1}{Script={sname@#1}}{}
  \bbl@ifunset{bbl@lname@#1}{}\bbl@csarg\bbl@add@list{lsys@#1}{Language={bbl@cs{lname@#1}}}{}
  \ifcase\bbl@engine\or\or
    \bbl@ifunset{bbl@prehc@#1}{}\bbl@exp{\ifblank{bbl@cs{prehc@#1}}{}\ifx\AtBeginDocument@notperr\expandafter\@secondoftwo % to execute right now
      \AtBeginDocument{%
        \bbl@patchfont\bbl@xenohyph%d
        \expandafter\select@language\expandafter{\languagename}\expandafter\select@language\expandafter{\languagename}}%
      \expandafter\select@language\expandafter{\languagename}}%
    \fi}
  \fi
  \bbl@csarg\bbl@toglobal{lsys@#1}
The following ini reader ignores everything but the identification section. It is called when a font is defined (i.e., when the language is first selected) to know which script/language must be enabled. This means we must make sure a few characters are not active. The ini is not read directly, but with a proxy .tex file named as the language (which means any code in it must be skipped, too).

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. Non-digits characters are kept.

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

\newcommand\localenumeral[2]{\bbl@cs{cntr@#1@languagename}{#2}}
\def\bbl@localecntr#1#2{\localenumeral{#2}{#1}}
\newcommand\localecounter[2]{\expandafter\bbl@localecntr\expandafter{\number\csname c@#2\endcsname}{#1}}
\def\bbl@alphnumeral#1#2{\expandafter\bbl@alphnumeral@i\number#276543210\@@{#1}}
\def\bbl@alphnumeral@i#1#2#3#4#5#6#7#8\@@#9{\ifcase\@car#8\@nil\or % Currently <10000, but prepared for bigger
\bbl@alphnumeral@ii{#9}000000#1\or \bbl@alphnumeral@ii{#9}00000#1#2\or \bbl@alphnumeral@ii{#9}0000#1#2#3\or \bbl@alphnumeral@ii{#9}000#1#2#3#4\else \bbl@alphnum@invalid{>9999}\fi}
\def\bbl@alphnumeral@ii#1#2#3#4#5#6#7#8{\bbl@ifunset{bbl@cntr@#1.F.@number#5#6#7#8@languagename}{\bbl@cs{cntr@#1.4@languagename}{#5} \bbl@cs{cntr@#1.3@languagename}{#6} \bbl@cs{cntr@#1.2@languagename}{#7} \bbl@cs{cntr@#1.1@languagename}{#8}}\ifnum#6#7#8>\z@ % TODO. An ad hoc rule for Greek. Ugly.
\bbl@ifunset{bbl@cntr@#1.S.321@languagename}{}\else \bbl@cs{cntr@#1.F.@number#5#6#7#8@languagename}\fi}
\def\bbl@alphnum@invalid#1{\bbl@error{alphabetic-too-large}{#1}{}{}}

The information in the identification section can be useful, so the following macro just exposes it
with a user command.
\def\bbl@localeinfo#1{%
\bbl@ifunset{bbl@info@#2}{#1}
\{\bbl@cs{cntr@#1.F.@number#5#6#7#8@languagename}\}
\bbl@error{no-ini-info}{}{}{}}
\newcommand\localeinfo[1]{\bbl@ifunset{bbl@info@#2}{#1}}
\newcommand\localeinfo[1]{\bbl@error{no-ini-info}{}{}{}}

The information in the identification section can be useful, so the following macro just exposes it
with a user command.
LATEX needs to know the BCP 47 codes for some features. For that, it expects `BCPdata` to be defined. While language, region, script, and variant are recognized, extension ⟨s⟩ for singletons may change.

\begin{verbatim}
\ifcase\bbl@engine % Converts utf8 to its code (expandable)
  \def\bbl@utftocode#1{\the\numexpr\decode@UTFviii#1\relax}
\else
  \def\bbl@utftocode#1{\expandafter`\string#1}
\fi
\providecommand\BCPdata{}
\ifx\renewcommand\@undefined\else % For plain. TODO. It’s a quick fix
  \renewcommand\BCPdata[1]{\bbl@bcpdata@i#1\@empty}
\def\bbl@bcpdata@i#1#2#3#4#5#6\@empty{
  \@nameuse{str_if_eq:nnTF}{#1#2#3#4#5}{main.}{
    \bbl@bcpdata@ii{#6}\bbl@main@language
  }{\bbl@bcpdata@ii{#1#2#3#4#5#6}\languagename}}
\def\bbl@bcpdata@ii#1#2{
  \bbl@ifunset{bbl@info@#1.tag.bcp47}{}{
    \bbl@ifunset{bbl@csname bbl@info@#1.tag.bcp47@#2}{}{
      \bbl@cs{bbl@csname bbl@info@#1.tag.bcp47@#2}}}\
\fi
\@namedef{bbl@info@casing.tag.bcp47}{casing}
\newcommand\BabelUppercaseMapping[3]{\DeclareUppercaseMapping[\@nameuse{bbl@casing@#1}]{#2}{#3}}
\newcommand\BabelTitlecaseMapping[3]{\DeclareTitlecaseMapping[\@nameuse{bbl@casing@#1}]{#2}{#3}}
\newcommand\BabelLowercaseMapping[3]{\DeclareLowercaseMapping[\@nameuse{bbl@casing@#1}]{#2}{#3}}
The parser for casing and casing ⟨variant⟩.
\end{verbatim}
With version 3.75 \BabelEnsureInfo is executed always, but there is an option to disable it.

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

\newcommand\getlocaleproperty{\@ifstar\bbl@getproperty@s\bbl@getproperty@x}
\def\bbl@getproperty@s#1#2#3{\let#1\relax\def\bbl@elt##1##2##3{\bbl@ifsamestring{##1/##2}{#3}{\providecommand#1{##3}\def\bbl@elt####1####2####3{}}}}\bbl@cs{inidata@#2}}\def\bbl@getproperty@x#1#2#3{\bbl@getproperty@s{#1}{#2}{#3}\ifx#1\relax\bbl@error{unknown-locale-key}{#1}{#2}{#3}\fi}
\let\bbl@ini@loaded\@empty\newcommand\LocaleForEach{\bbl@foreach\bbl@ini@loaded}
\def\ShowLocaleProperties#1{\typeout{}\typeout{*** Properties for language '#1' ***}\def\bbl@elt##1##2##3{\typeout{##1/##2 = ##3}}\@nameuse{bbl@inidata@#1}\typeout{*******}}

5 Adjusting the Babel behavior

A generic high level interface is provided to adjust some global and general settings.
\def\bbl@adjust@lua#1#2{\ifvmode\ifnum\currentgrouplevel=\z@\directlua{Babel.#2}\expandafter\expandafter\expandafter\@gobble\fi\fi{\bbl@error{adjust-only-vertical}{#1}{}}}% Gobbled if everything went ok.
\@namedef{bbl@ADJ@bidi.mirroring@on}{\bbl@adjust@lua{bidi}{mirroring_enabled=true}}
\@namedef{bbl@ADJ@bidi.mirroring@off}{\bbl@adjust@lua{bidi}{mirroring_enabled=false}}
\@namedef{bbl@ADJ@bidi.text@on}{\bbl@adjust@lua{bidi}{bidi_enabled=true}}
\@namedef{bbl@ADJ@bidi.text@off}{\bbl@adjust@lua{bidi}{bidi_enabled=false}}
\@namedef{bbl@ADJ@bidi.math@on}{\let\bbl@noamsmath@empty}
\@namedef{bbl@ADJ@bidi.math@off}{\let\bbl@noamsmath\relax}
\@namedef{bbl@ADJ@bidi.mapdigits@on}{\bbl@adjust@lua{bidi}{digits_mapped=true}}
\@namedef{bbl@ADJ@bidi.mapdigits@off}{\bbl@adjust@lua{bidi}{digits_mapped=false}}
\@namedef{bbl@ADJ@linebreak.sea@on}{\bbl@adjust@lua{linebreak}{sea_enabled=true}}
\@namedef{bbl@ADJ@linebreak.sea@off}{\bbl@adjust@lua{linebreak}{sea_enabled=false}}
\@namedef{bbl@ADJ@linebreak.cjk@on}{\bbl@adjust@lua{linebreak}{cjk_enabled=true}}
\@namedef{bbl@ADJ@linebreak.cjk@off}{\bbl@adjust@lua{linebreak}{cjk_enabled=false}}
\@namedef{bbl@ADJ@justify.arabic@on}{\bbl@adjust@lua{linebreak}{arabic.justify_enabled=true}}
\@namedef{bbl@ADJ@justify.arabic@off}{\bbl@adjust@lua{linebreak}{arabic.justify_enabled=false}}
\def\bbl@adjust@layout#1{\ifvmode#1\expandafter\@gobble\fi{\bbl@error{layout-only-vertical}{}{}}}% Gobbled if everything went ok.
\@namedef{bbl@ADJ@layout.tabular@on}{\bbl@adjust@layout{\let\@tabular\bbl@NL@@tabular}}
\@namedef{bbl@ADJ@layout.tabular@off}{\bbl@adjust@layout{\let\@tabular\bbl@OL@@tabular}}
\@namedef{bbl@ADJ@layout.lists@on}{\bbl@adjust@layout{\let\list\bbl@NL@list}}
\@namedef{bbl@ADJ@layout.lists@off}{\bbl@adjust@layout{\let\list\bbl@OL@list}}
\namedef{bbl@ADJ@autoload.bcp47@on}{\bbl@bcpallowedtrue}
\namedef{bbl@ADJ@autoload.bcp47@off}{\bbl@bcpallowedfalse}
\namedef{bbl@ADJ@autoload.bcp47.prefix}#1{\def\bbl@bcp@prefix{#1}}
\def\bbl@bcp@prefix{bcp47-}
\namedef{bbl@ADJ@autoload.options}#1{\def\bbl@autoload@options{#1}}
\let\bbl@autoload@bcpoptions\@empty
\namedef{bbl@ADJ@autoload.bcp47.options}#1{\def\bbl@autoload@bcpoptions{#1}}
\newif\ifbbl@bcptoname
\namedef{bbl@ADJ@bcp47.toname@on}{\bbl@bcptonametrue\BabelEnsureInfo}
\namedef{bbl@ADJ@bcp47.toname@off}{\bbl@bcptonamefalse}
\namedef{bbl@ADJ@prehyphenation.disable@nohyphenation}{\directlua{Babel.ignore_pre_char = function(node) return (node.lang == \the\csname l@nohyphenation\endcsname) end}}
\namedef{bbl@ADJ@prehyphenation.disable@off}{\directlua{Babel.ignore_pre_char = function(node) return false end}}
\namedef{bbl@ADJ@interchar.disable@nohyphenation}{\def\bbl@ignoreinterchar{\ifnum\language=\l@nohyphenation\expandafter\@gobble\else\expandafter\@firstofone\fi}}
\namedef{bbl@ADJ@interchar.disable@off}{\let\bbl@ignoreinterchar\@firstofone}
\namedef{bbl@ADJ@select.write@shift}{\let\bbl@restorelastskip\relax\def\bbl@savelastskip{\let\bbl@restorelastskip\relax\ifvmode\ifdim\lastskip=\z@\let\bbl@restorelastskip\nobreak\else\bbl@exp{\def\vskip\the\lastskip\nobreak \vskip-\the\lastskip \vskip\the\lastskip}\fi\fi}}
\namedef{bbl@ADJ@select.write@keep}{\let\bbl@restorelastskip\relax\let\bbl@savelastskip\relax\ifvmode\ifdim\lastskip=\z@\let\bbl@restorelastskip\nobreak\else\bbl@exp{\def\vskip\the\lastskip\nobreak \vskip-\the\lastskip \vskip\the\lastskip}\fi\fi}}
\namedef{bbl@ADJ@select.write@omit}{\AddBabelHook{babel-select}{beforestart}{\expandafter\babel@aux\expandafter{\bbl@main@language}{}}}\let\bbl@restorelastskip\relax\def\bbl@savelastskip#1\bbl@restorelastskip{}\namedef{bbl@ADJ@select.encoding@off}{\let\bbl@encoding@select@off\@empty}
5.1 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 following package options control which macros are to be redefined.

```
3695 ⟨⟨∗ More package options⟩⟩ ≡
3696 \DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
3697 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
3698 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
3699 \DeclareOption{safe=refbib}{\def\bbl@opt@safe{BR}}
3700 \DeclareOption{safe=bibref}{\def\bbl@opt@safe{BR}}
3701 ⟨⟨/More package options⟩⟩
```

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

```
3702 \bbl@trace{Cross referencing macros}
3703 \ifx\bbl@opt@safe\@empty\else % ie, if 'ref' and/or 'bib'
3704 \def\@newl@bel{\protect@activestrue}
3705 \bbl@ifunset{#1@#2}{}\relax
3706 \global\@namedef{#1@#2}{#3}
3707 \relax
3708 \edef\@multiplelabels{%
3709 \latex@warning{There were multiply-defined labels}}%
3710 \latex@warning{Label `#2' multiply defined}%
3711 \global\@namedef{#1@#2}{#3}}
```

\@testdef An internal \LaTeX\ macro used to test if the labels that have been written on the .aux file have changed. It is called by the \enddocument macro.

```
3712 \CheckCommand\@testdef[3]{% 3713 \def\reserved@a{#3}%
3714 \edef\@testdefa{#1#2\@testdefb}%
3715 \@tempswatrue
3716 \@tempswatrue
3717 \fi}
```

Now that we made sure that \@testdef still has the same definition we can rewrite it. First we make the shorthands ‘safe’. Then we use \bbl@tempa as an ‘alias’ for the macro that contains the label which is being checked. Then we define \bbl@tempb just as \@newl@bel does it. When the label is defined we replace the definition of \bbl@tempa by its meaning. If the label didn’t change, \bbl@tempa and \bbl@tempb should be identical macros.

```
3718 \def\@testdef{#1#2#3}% TODO. With @samestring?
3719 \def\@testdefa{#1#2\@testdefb}%
3720 \edef\@testdefa{#1#2\@testdefb}%
3721 \edef\@testdefb{#3}%
3722 \@tempswatrue
3723 \@tempswatrue
3724 \@tempswatrue
3725 \edef\@testdefb{#1#2\@testdefc}%
3726 \fi
3727 \edef\@testdefc{#3}%
3728 \def\@testdefc{#3}%
3729 \@tempswatrue
3730 \@tempswatrue
3731 \@tempswatrue
3732 \@tempswatrue
3733 \@tempswatrue
```
\ref The same holds for the macro \ref that references a label and \pageref to reference a page. We make them robust as well (if they weren't already) to prevent problems if they should become expanded at the wrong moment.

```latex
3733 \bbl@xin@{R}\bbl@opt@safe
3734 \ifin@
3735 \edef\bbl@tempc{\expandafter\string\csname ref code\endcsname}\
3736 \bbl@xin@{\expandafter\strip@prefix\meaning\bbl@tempc}\
3737 \expandafter\strip@prefix\meaning\ref\ifin@
3738 \else
3739 \bbl@redefine\@kernel@ref#1{%}
3740 \@safe@activestrue\org@@kernel@ref{#1}\@safe@activesfalse
3741 \bbl@redefine\@kernel@pageref#1{%}
3742 \@safe@activestrue\org@@kernel@pageref{#1}\@safe@activesfalse
3743 \bbl@redefine\@kernel@sref#1{%}
3744 \@safe@activestrue\org@@kernel@sref{#1}\@safe@activesfalse
3745 \bbl@redefine\@kernel@spageref#1{%}
3746 \@safe@activestrue\org@@kernel@spageref{#1}\@safe@activesfalse
3747 \else
3748 \bbl@redefinerobust\ref#1{%
3749 \@safe@activestrue\org@ref{#1}\@safe@activesfalse
3750 \bbl@redefinerobust\pageref#1{%
3751 \@safe@activestrue\org@pageref{#1}\@safe@activesfalse
3752 \fi
3753 \else
3754 \let\org@ref\ref
3755 \let\org@pageref\pageref
3756 \fi
```

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

```latex
3757 \bbl@xin@{B}\bbl@opt@safe
3758 \ifin@
3759 \bbl@redefine\@citex[#1]#2{%
3760 \@safe@activestrue\edef\bbl@tempa{#2}\@safe@activesfalse
3761 \org@@citex[#1]{\bbl@tempa}
3762 \else
3763 \bbl@redefinerobust\cite[#1]#2{%
3764 \@safe@activestrue\org@@citex[#1]{#2}\@safe@activesfalse
3765 \bbl@redefinerobust\pageref[#1]#2{%
3766 \@safe@activestrue\org@pageref[#1]{#2}\@safe@activesfalse
3767 \else
3768 \bbl@redefinerobust\cite#2{%
3769 \@safe@activestrue\edef\bbl@tempa{#2}\@safe@activesfalse
3770 \\org@cite[#1]{\bbl@tempa}
3771 \}\{}
3772 \}\{}
```

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.

```latex
3773 \AtBeginDocument{%
3774 \@ifpackageloaded{natbib}%
3775 \def\@citex[#1]#2{%
3776 \@safe@activestrue\edef\bbl@tempa{#2}\@safe@activesfalse
3777 \org@@citex[#1]{\bbl@tempa}
3778 \}@{}
```
The macro \nocite which is used to instruct BiBTeX to extract uncited references from the database.

\bbl@redefine\nocite#1{% \@safe@activestrue\org@nocite{#1}\@safe@activesfalse}

The macro that is used in the .aux file to define citation labels. When packages such as natbib or cite are not loaded its second argument is used to typeset the citation label. In that case, this second argument can contain active characters but is used in an environment where \@safe@activestrue is in effect. This switch needs to be reset inside the \hbox which contains the citation label. In order to determine during .aux file processing which definition of \bibcite is needed we define \bibcite in such a way that it redefines itself with the proper definition. We call \bbl@cite@choice to select the proper definition for \bibcite. This new definition is then activated.

\bbl@redefine\bibcite{% \bbl@cite@choice \bibcite}

The macro \bbl@bibcite holds the definition of \bibcite needed when neither natbib nor cite is loaded.
\def\bbl@bibcite#1#2{% \org@bibcite{#1}{\@safe@activesfalse#2}}

The macro \bbl@cite@choice determines which definition of \bibcite is needed. First we give \bibcite its default definition.
\def\bbl@cite@choice{% \global\let\bibcite\bbl@bibcite \@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{} \@ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{} \global\let\bbl@cite@choice\relax}

When a document is run for the first time, no .aux file is available, and \bibcite will not yet be properly defined. In this case, this has to happen before the document starts.
\AtBeginDocument{%\bbl@cite@choice}

One of the two internal \LaTeX macro called by \bibitem that write the citation label on the .aux file.
\def\bbl@bibitem#1{% \@safe@activestrue\org@@bibitem{#1}\@safe@activesfalse}
\else
\let\org@nocite\nocite
\let\org@citex\@citex
\let\org@bibcite\bibcite
\let\org@bibitem\@bibitem
\fi

\section{Marks}

Because the output routine is asynchronous, we must pass the current language attribute to the head lines. To achieve this we need to adapt the definition of \markright and \markboth somewhat. However, headlines and footlines can contain text outside marks; for that we must take some actions in the output routine if the 'headfoot' options is used. We need to make some redefinitions to the output routine to avoid an endless loop and to correctly handle the page number in bidi documents.
\bbl@trace{Marks}
\IfBabelLayout{sectioning}{\ifx\bbl@opt@headfoot\@nnil\g@addto@macro\@resetactivechars{\set@typeset@protect\expandafter\select@language@x\expandafter{\bbl@main@language}\let\protect\noexpand\ifcase\bbl@bidimode\else % Only with bidi. See also above\edef\thepage\noexpand\babelsublr{\unexpanded\expandafter{\thepage}}\fi}}
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.)

\[\textit{Preventing clashes with other packages}\]

\[5.3\] Preventing clashes with other packages

\[5.3.1\] ifthen

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:

\[
\textit{ifthenelse}{\isodd{\pageref{some:label}}}\\
\{\text{code for odd pages}\}\\
\{\text{code for even pages}\}
\]

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

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.

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.
5.3.2 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 needs to happen for \vrefpagenum.

\AtBeginDocument{\Ifpackageloaded{varioref}{\bbl@redefine\@vpageref#1[#2]#3{%\@safe@activestru\e\org@vpageref{#1}{#2}{#3}\@safe@activesfalse}\bbl@redefine\vrefpagenum#1#2{%\@safe@activestru\e\org@vrefpagenum{#1}{#2}\@safe@activesfalse}}\fi}

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

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

5.3.3 hhline

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

\AtEndOfPackage{\AtBeginDocument{\Ifpackageloaded{hhline}{\expandafter\ifx\csname normal@char\string:\endcsname\relax\else\makeatletter\def\currname{hhline}\input{hhline.sty}\makeatother\fi\{%}}}}
Deprecated. Use the tools provided by \TeX. The command \texttt{\substitutefontfamily} creates an .fd file on the fly. The first argument is an encoding mnemonic, the second and third arguments are font family names.

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

5.4 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. Requested encodings are currently stored in \texttt{@fontenc@load@list}. If a non-ASCII has been loaded, we define versions of \TeX and \LaTeX for them using \texttt{\ensureascii}. The default ASCII encoding is set, too (in reverse order): the “main” encoding (when the document begins), the last loaded, or OT1.

\begin{verbatim}
\ensureascii
\end{verbatim}

84
Now comes the old deprecated stuff (with a little change in 3.9l, for fontspec). The first thing we need to do is to determine, at \begin{document}, which latin font encoding to use.

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

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

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

\AtBeginDocument{%
\@ifpackageloaded{fontspec}{\xdef\latinencoding{\ifx\UTFencname\@undefined\EU\ifcase\bbl@engine\or2\or1\fi\else\UTFencname\fi}}%
\gdef\latinencoding{OT1}\ifx\cf@encoding\bbl@t@one\xdef\latinencoding{\bbl@t@one}\else\def\@elt#1{,#1,}\edef\bbl@tempa{\expandafter\@gobbletwo\@fontenc@load@list}\let\@elt\relax\bbl@xin{,T1,}\bbl@tempa\ifin@\xdef\latinencoding{\bbl@t@one}\fi\fi}}%

\latintext 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.

\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.

\Ifx{\undefined}{\DeclareTextFontCommand{\textlatin}[1]{\leavevmode{\latintext #1}}}{\else{\DeclareTextFontCommand{\textlatin}[1]{\latintext}}}{\fi}

For several functions, we need to execute some code with \selectfont. With \LaTeX 2021-06-01, there is a hook for this purpose.

\def\bbl@patchfont#1{\AddToHook{\selectfont}{#1}}

\section{Basic bidi support}

\textbf{Work in progress.} This code is currently placed here for practical reasons. It will be moved to the correct place soon, I hope.

It is loosely based on r\babel.\ def, but most of it has been developed from scratch. This babel module (by Johannes Braams and Boris Lavva) has served the purpose of typesetting R documents for two decades, and despite its flaws I think it is still a good starting point (some parts have been
copied here almost verbatim), partly thanks to its simplicity. I've also looked at \textsc{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 rl babel did), and by introducing a “middle layer” just below the user interface (sectioning, footnotes).

- \texttt{pdftex} provides a minimal support for bidi text, and it must be done by hand. Vertical typesetting is not possible.
- \texttt{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.
- \texttt{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 \texttt{Lu\TeX}-ja shows, vertical typesetting is possible, too.

\begin{verbatim}
3967\bbl@trace{Loading basic (internal) bidi support}
3968\iffidd\\bbl@engine
3969\else % TODO. Move to txtbabel
3970  \ifnum\\bbl@bidimode>100 \ifnum\\bbl@bidimode<200 % Any xe+lua bidi=
3971  \bbl@error{bidi-only-lua}{}{}{}%
3972  \let\\bbl@beforeforeign\leavevmode
3973  \AtEndOfPackage{%
3974  \EnableBabelHook{babel-bidi}%
3975  \bbl@xebidipar}
3976 \fi\fi
3977 \def\\bbl@loadxebidi#1{%
3978  \ifx\\RTLfootnotetext@undefined
3979  \AtEndOfPackage{%
3980  \EnableBabelHook{babel-bidi}%
3981  \bbl@loadfontspec % bidi needs fontspec
3982  \usepackage#1{bidi}%
3983  \let\\digitsdotdash\DigitsDotDashInterCharToks
3984  \def\DigitsDotDashInterCharToks{% See the 'bidi' package
3985  \ifnum@nameuse{bbl@wdir@\languagename}=	w@ % 'AL' bidi
3986  \bbl@digitsdotdash % So ignore in 'R' bidi
3987  \fi}%
3988 \fi}
3989 \ifnum\\bbl@bidimode>200 % Any xe bidi=
3990 \ifcase\expandafter\@gobbletwo\the\\bbl@bidimode\or
3991  \bbl@tentative{bidi=bidi}
3992  \bbl@loadxebidi()
3993 \or
3994  \bbl@loadxebidi[[rldocument}}
3995 \or
3996  \bbl@loadxebidi{}
3997 \fi
3998 \fi
3999 \fi
4000% TODO? Separate:
4001\iffnum\\bbl@bidimode=@ne % Any bidi= except default=1
4002 \let\\bbl@beforeforeign\leavevmode
4003 \iffidd\\bbl@engine
4004 \newattribute\\bbl@attr@dir
4005 \directlua{ Babel.attr_dir = luatexbase.registernumber'bbl@attr@dir' }
4006 \bbl@exp{\output{\bodydir\pagedir\the\output}}
4007 \fi
4008 \AtEndOfPackage{%
4009 \EnableBabelHook{babel-bidi}%
4010 \iffidd\\bbl@engine\else
4011 \bbl@xebidipar
4012 \fi}
4013\fi
\end{verbatim}

86
Now come the macros used to set the direction when a language is switched. First the (mostly) common macros.

4014 `\bbl@trace{Macros to switch the text direction}`
4015 `\def\bbl@alscripts{,Arabic,Syriac,Thai,Han,}
4016 \def\bbl@rscripts{,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,}
4017 Old Hungarian,Lydia,Mandaean,Manichaean,%
4018 Meroitic Cursive,Meroitic,Old North Arabian,%
4019 Nabataean,N’ko,Orkhon,Palmyrene,Inscriptional Pahlavi,%
4020 Psalter Pahlavi,Phoenician,Inscriptional Partian,Samaritan,%
4021 Old South Arabian,}
4022 `\def\bbl@provide@dirs#1{%
4023 \bbl@xin{\csname bbl@sname@#1\endcsname}{\bbl@alscripts}\bbl@rscripts}%
4024 \ifin@
4025 \global\bbl@csarg\chardef{wdir@#1}\@ne
4026 \bbl@xin{\csname bbl@sname@#1\endcsname}{\bbl@alscripts}%
4027 \ifin@
4028 \global\bbl@csarg\chardef{wdir@#1}\tw@
4029 \fi
4030 \else
4031 \global\bbl@csarg\chardef{wdir@#1}\z@
4032 \fi
4033 `\ifodd\bbl@engine`
4034 `\bbl@switchdir{%
4035 \bbl@csarg\ifcase{wdir@#1}%
4036 \directlua{ Babel.locale_props[\the\localeid].textdir = 'l' }%
4037 \or
4038 \directlua{ Babel.locale_props[\the\localeid].textdir = 'r' }%
4039 \or
4040 \directlua{ Babel.locale_props[\the\localeid].textdir = 'al' }%
4041 \fi`%
4042 \fi`%
4043 `\def\bbl@switchdir{%
4044 \bbl@ifsyst{\bbl@lsys\languagename}{\bbl@provide@sys\languagename}{}%
4045 \bbl@ifsyst{\bbl@wdir\languagename}{\bbl@provide@dirs\languagename}{}%
4046 \bbl@exp{\bbl@setdirs\bbl@cl\wdir}}%
4047 `\def\bbl@setdirs#1{%
4048 \bbl@textdir{#1}%
4049 \bbl@pardir{#1}%- Must precede \bbl@textdir
4050 \fi%
4051 \fi`%
4052 \ifodd\bbl@engine % luatex=1
4053 `\ifcase{wdir@#1}%
4054 \ifcase\bbl@select@type % TODO - strictly, not the right test
4055 \bbl@bodydir{#1}%
4056 \bbl@pardir{#1}%- Must precede \bbl@textdir
4057 \fi
4058 \fi`%
4059 \ifodd\bbl@engine % luatex=1
4060 % TODO. Only if \bbl@bidimode > 0?:
4061 % `\AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}`
4062 % `\DisableBabelHook{babel-bidi}`

Now the engine-dependent macros. TODO. Must be moved to the engine files.

4066 `\ifodd\bbl@engine % luatex=1
4067 `\else % pdftex=0, xetex=2
4068 \newcount\bbl@dirlevel
4069 \chardef\bbl@thetextdir\z@
4070 \chardef\bbl@thepardir\z@
4071 `\def\bbl@textdir{#1}{%
4072 `\ifcase#1\relax
4073 `\chardef\bbl@thetextdir\z@
4074 `\@nameuse{setlatin}%
4075 `\bbl@textdir@i\beginL\endL
4076 `\else
4077 `\immediate\write0{\bbl@textdir@i\beginR\endR}
4078 `\@nameuse{setnonlatin}%
4079 `\bbl@textdir@i\beginR\endR
4080 \fi`%
4081 `\def\bbl@textdir@i#1#2{%
4082 \ifhmode
4083 \chardef\bbl@thetextdir\z@
4084 \@nameuse{setlatin}%
4085 `\bbl@textdir@i\beginL\endL
4086 \else
4087 \immediate\write0{\bbl@textdir@i\beginR\endR}
4088 `\@nameuse{setnonlatin}%
4089 `\bbl@textdir@i\beginR\endR
4090 \fi`%
4091 `\def\bbl@textdir@i#1#2{%
4092 \ifhmode
4093 \chardef\bbl@thetextdir\z@
4094 \@nameuse{setlatin}%
4095 `\bbl@textdir@i\beginL\endL
4096 \else
4097 \immediate\write0{\bbl@textdir@i\beginR\endR}
4098 `\@nameuse{setnonlatin}%
4099 `\bbl@textdir@i\beginR\endR
4100 \fi`%
The following command is executed only if there is a right-to-left script (once). It activates the \everypar hack for \texttt{xetex}, to properly handle the par direction. Note text and par dirs are decoupled to some extent (although not completely).

\def\bbl@xebidipar{% 
\let\bbl@xebidipar\relax 
\TeXXeTstate\@ne 
\def\bbl@xeeverypar{% 
  \ifcase\bbl@thetextdir 
    \expandafter\bbl@tempa\expandafter{\BabelText{\LR{##1}}}% 
  \else 
    \expandafter\bbl@tempa\expandafter{\BabelText{\RL{##1}}}% 
  \fi} 
\def\bbl@pardir#1{\ifcase#1\relax\setLR\else\setRL\fi} 
\fi} 
\AtBeginDocument{% 
  \ifx\pdfstringdefDisableCommands\@undefined\else 
    \AddBabelHook{bidi}{foreign}{% 
      \def\tempa{\def\BabelText##1} 
      \ifcase\bbl@thetextdir 
        \expandafter\tempa\expandafter{\BabelText{\LR{##1}}}% 
      \else 
        \expandafter\tempa\expandafter{\BabelText{\RL{##1}}}% 
      \fi} 
  \fi\fi 
  \newtoks\everypar \everypar=\bbl@severypar 
  \bbl@severypar{\bbl@xeeverypar\the\everypar} 
  \newtoks\everypar \everypar=\bbl@severypar 
  \bbl@severypar{% \bbl@xeeverypar\the\everypar}}% 
\ifnum\bbl@bidimode>200 % Any xe bidi= 
\let\bbl@textdir\@gobbletwo 
\let\bbl@xebidipar\@empty 
\AddBabelHook{bidi}{foreign}{% 
  \def\bbl@tempa{\def\BabelText##1} 
  \ifcase\bbl@thetextdir 
    \expandafter\tempa\expandafter{\BabelText{\LR{##1}}}% 
  \else 
    \expandafter\tempa\expandafter{\BabelText{\RL{##1}}}% 
  \fi} 
\def\bbl@pardir#1{\ifcase#1\relax\setLR\else\setRL\fi} 
\fi 
\fi 

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

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

\fi 
\fi 

A tool for weak \texttt{L} (mainly digits). We also disable warnings with \texttt{hyperref}.
\loadlocalcfg At some sites it may be necessary to add site-specific actions to a language definition file. This can be done by creating a file with the same name as the language definition file, but with the extension .cfg. For instance the file norsk.cfg will be loaded when the language definition file norsk.ldf is loaded.

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

\def\bbl@try@load@lang#1#2#3{\IfFileExists{#1.ldf}{\bbl@load@language{#1}}{#3\bbl@load@language{#2}}}

5.7 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).

\def\bbl@try@load@lang#1#2#3{\IfFileExists{#1.ldf}{\bbl@load@language{#1}}{#3\bbl@load@language{#2}}}

% \DeclareOption{hebrew}{%
% \ifcase\bbl@engine\or
% \bbl@error{only-pdftex-lang}{hebrew}{luatex}{}%
% \else
\expandafter\let\expandafter\bbl@afterlang\csname\CurrentOption.ldf-h@@k\endcsname
% \fi}

Now, we set a few language options whose names are different from ldf files. These declarations are preserved for backwards compatibility, but they must be eventually removed. Use proxy files instead.

\def\bbl@try@load@lang#1#2#3{\IfFileExists{#1.ldf}{\bbl@load@language{#1}}{#3\bbl@load@language{#2}}}

% \DeclareOption{hebrew}{%
% \ifcase\bbl@engine\or
% \bbl@error{only-pdftex-lang}{hebrew}{luatex}{}%
% \else

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 bb@language@opts are assumed to be languages. If not declared above, the names of the option and the file are the same. We first pre-process the class and package options to determine the main language, which is processed in the third ‘main’ pass, except if all files are ldf and there is no main key. In the latter case (bb@opt@main is still \@nnil), the traditional way to set the main language is kept — the last loaded is the main language.

A few languages are still defined explicitly. They are stored in case they are needed in the ‘main’ pass (the value can be \relax).
Now define the corresponding loaders. With package options, assume the language exists. With class options, check if the option is a language by checking if the corresponding file exists.

4223 \foreach\@language@opts{% 
4224 \def\temp@{#1}% 
4225 \ifx\temp@\opt@main\else 
4226 \ifnum\iniflag<\tw@% \other = ldf 
4227 \ifunset{ds@#1}% 
4228 `{\ DeclareOption{#1}{\loadlanguage{#1}}}% 
4229 \}{}% 
4230 \else% \other = ini 
4231 \DeclareOption{#1}{% 
4232 \ldfinit 
4233 \babelprovide[import]{#1}% 
4234 \afterldf{}}% 
4235 \fi 
4236 \fi 
4237 \foreach\@classoptionslist{% 
4238 \def\temp@{#1}% 
4239 \ifx\temp@\opt@main\else 
4240 \ifnum\iniflag<\tw@% \other = ldf 
4241 \ifFileExists{#1.ldf}% 
4242 `{\ DeclareOption{#1}{\loadlanguage{#1}}}% 
4243 \}{}% 
4244 \else% \other = ini 
4245 \ifFileExists{babel-#1.tex}% 
4246 `{\ DeclareOption{#1}{% 
4247 \ldfinit 
4248 \babelprovide[import]{#1}% 
4249 \afterldf{}}% 
4250 \}{}% 
4251 \fi 
4252 \fi 
4253 \fi 
4254 \fi 

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

The options have to be processed in the order in which the user specified them (but remember class options are processes before):

4255 \def\AfterBabelLanguage@#1{% 
4256 \ifsamestring\CurrentOption{#1}{\global\add\afterlang}{}} 
4257 \DeclareOption*{} 
4258 \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. With some options in provide, the package luatexbase is loaded (and immediately used), and therefore \babelprovide can't go inside a \DeclareOption; this explains why it's executed directly, with a dummy declaration. Then all languages have been loaded, so we deactivate \AfterBabelLanguage.

4259 \trace{Option 'main'} 
4260 \foreach\opt@main\@nil 
4261 \edef\temp@{\classoptionslist,\language@opts} 
4262 \let\temp@=\empty 
4263 \edef\temp{,\loaded,} 
4264 \edef\temp\strip\meaning\temp 
4265 \for\temp\begin{meaning}\temp 
4266 \edef\tempd\strip\meaning\temp 
4267 \edef\tempd\strip\meaning\temp 
4268 \xint Diff{\tempd}{\temp} 
4269 \edef\temp{\begin{meaning}\temp} 
4270 \edef\temp{\end{meaning}} 
4271 \expandafter\temp
In order to catch the case where the user didn't specify a language we check whether \bbl@main@language has become defined. If not, the nil language is loaded.

\section{Error messages}

The kernel of Babel (\texttt{babel.def}, common)

The kernel of the babel system is currently stored in \texttt{babel.def}. The file \texttt{babel.def} contains most of the code. The file \texttt{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.

Because plain \TeX{} users might want to use some of the features of the babel system too, care has to be taken that plain \TeX{} can process the files. For this reason the current format will have to be checked in a number of places. Some of the code below is common to plain \TeX{} and \LaTeX{}, some of it is for the \LaTeX{} case only.

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

A proxy file for \texttt{switch.def}

\section{Error messages}

\section{Error messages}
They are loaded when \bbl@error is first called. To save space, the main code just identifies them with a tag, and messages are stored in a separate file. Since it can be loaded anywhere, you make sure some catcodes have the right value, although those for |\|, |`|, |\^^M|, |%| and |=| are reset before loading the file.

\catcode`\{=1 \catcode`\}=2 \catcode`\#=6
\catcode`:=12 \catcode`,:=12 \catcode`\-=12
\catcode`\'=12 \catcode`\(=12 \catcode`\)=12
\catcode`\@=11 \catcode`\^=7

\ifx\MessageBreak\@undefined
\gdef\bbl@error@i#1#2{%\begingroup
\newlinechar=`\^^J
\def\{\MessageBreak(babel) }%
\errhelp{#2}\errmessage{\#1}
\endgroup}
\else
\gdef\bbl@error@i#1#2{%\begingroup
\def\{\MessageBreak%
\PackageError{babel}{#1}{#2}%
\endgroup}
\fi
\def\bbl@errmessage#1#2#3{\expandafter\gdef\csname bbl@err@#1\endcsname##1##2##3{%\bbl@error@i{#2}{#3}}}
\bbl@errmessage{not-yet-available}{Not yet available}{Find an armchair, sit down and wait}
\bbl@errmessage{bad-package-option}{Bad option '#1=#2'. Either you have misspelled the key or there is a previous setting of '#1'. Valid keys are, among others, 'shorthands', 'main', 'bidi', 'strings', 'config', 'headfoot', 'safe', 'math'.}{See the manual for further details.}
\bbl@errmessage{base-on-the-fly}{For a language to be defined on the fly 'base' is not enough, and the whole package must be loaded. Either delete the 'base' option or request the languages explicitly}{See the manual for further details.}
\bbl@errmessage{undefined-language}{You haven't defined the language '#1' yet.}{Perhaps you misspelled it or your installation is not complete}{Your command will be ignored, type <return> to proceed}
\bbl@errmessage{shorthand-is-off}{I can't declare a shorthand turned off (\string#2)}{Sorry, but you can't use shorthands which have been turned off in the package options}{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}
\errormessage{not-a-shorthand-b}
\errmessage{This character is not a shorthand. Maybe you made a typing mistake? I will ignore your instruction.}
\errormessage{unknown-attribute}
\errmessage{The attribute \#2 is unknown for language \#1.}
\errmessage{Your command will be ignored, type \texttt{<return>} to proceed}
\errormessage{unknown-group}
\errmessage{Missing group for string \texttt{\string\#1}}
\errmessage{You must assign strings to some category, typically captions or extras, but you set none}
\errormessage{only-lua-xe}
\errmessage{This macro is available only in LuaLaTeX and XeLaTeX. Consider switching to these engines.}
\errormessage{only-lua}
\errmessage{This macro is available only in LuaLaTeX. Consider switching to that engine.}
\errormessage{unknown-provide-key}
\errmessage{Unknown key \#1 in \texttt{\string\babelprovide}}
\errmessage{See the manual for valid keys}
\errormessage{unknown-mapfont}
\errmessage{Option \texttt{\string\bbl@KVP@mapfont} unknown for \texttt{\string\mapfont}. Use \texttt{\string\direction}.}
\errmessage{no-ini-file}
\errmessage{There is no ini file for the requested language \#1: \texttt{\string\languagename}. Perhaps you misspelled it or your installation is not complete.}
\errmessage{digits-is-reserved}
\errmessage{The counter name 'digits' is reserved for mapping decimal digits}
\errormessage{no-ini-info}
\errmessage{I've found no info for the current locale. The corresponding ini file has not been loaded. Perhaps it doesn't exist.}
\errormessage{adjust-only-vertical}
\errmessage{Currently, \#1 related features can be adjusted only in the main vertical list.}
\errormessage{layout-only-vertical}
\errmessage{Currently, layout related features can be adjusted only in vertical mode.}
\errormessage{bidi-only-lua}
\errmessage{This macro is available only in LuaLaTeX.}
The bidi method 'basic' is available only in luatex. I'll continue with 'bidi=default', so expect wrong results.

See the manual for further details.

Multiple bidi settings inside a group

I'll insert a new group, but expect wrong results.

The language definition file \CurrentOption.ldf

was not found

Valid options are, among others: shorthands=, KeepShorthandsActive,
activeacute, activegrave, noconfigs, safe=, main=, math=
headfoot=, strings=, config=, hyphenmap=, or a language name.

Local config file \bbl@opt@config.cfg not found

Maybe there is a typo.

'{#1}' for '{\languagename}' cannot be enabled.

Maybe there is a typo.

'{#1}' for '{\languagename}' cannot be disabled.

Maybe there is a typo.

'{#1}' for '{\languagename}' cannot be enabled.

Maybe there is a typo or it's a font-dependent transform

Maybe there is a typo or it's a font-dependent transform

Year out of range.

The allowed range is #1

See the manual for further details.
The '#1' ldf style doesn't work with #2,\% but you can use the ini locale instead.\%
Try adding 'provide='" to the option list. You may\%
also want to set 'bidi=' to some value.)
{See the manual for further details.}

⟨/errors⟩
⟨∗patterns⟩

7 Loading hyphenation patterns
The following code is meant to be read by iniTeX because it should instruct TeX to read hyphenation patterns. To this end the docstrip option patterns is used to include this code in the file hyphen.cfg. Code is written with lower level macros.

\ProvidesFile{hyphen.cfg}[date\version Babel hyphens]
\xdef\bbl@format{\jobname}
\def\bbl@version{\version}
\def\bbl@date{\date}
\ifx\AtBeginDocument\@undefined
\def\@empty{}
\fi
⟨⟨Define core switching macros⟩⟩
\process@line
Each line in the file language.dat is processed by \process@line after it is read. The first thing this macro does is to check whether the line starts with =. When the first token of a line is an =, the macro \process@synonym is called; otherwise the macro \process@language will continue.

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

\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 `(lang)` 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.

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. `\loadkernel` currently loads nothing, but
define some basic macros instead.

\def\bbl@hook@everylanguage#1{}
\def\bbl@hook@loadpatterns#1{\input #1\relax}
\def\bbl@hook@loadexceptions\bbl@hook@loadpatterns
\def\addlanguage{\csname newlanguage\endcsname}\addlanguage{newlanguage\endcsname}%
\def\adddiasctext{}%\def\adddiasctext{a dialect from \string\language}
\global\chardef\string#1\relax
\wlog{\string#1 = a dialect from \string\language}
\def\iflanguage#1{\expandafter\ifx\csname l@#1\endcsname\relax
\@nolanerr{#1}\else
\expandafter\expandafter\expandafter\@firstoftwo
\else
\expandafter\expandafter\expandafter\@secondoftwo
\fi}
\def\providehyphenmins#1#2{\expandafter\ifx\csname ##1hyphenmins\endcsname\relax
\@namedef{##1hyphenmins}{##2}\fi}
\def\set@hyphenmins#1#2{\lefthyphenmin##1\relax\righthyphenmin##2\relax}
\def\selectlanguage{\errhelp{Selecting a language requires a package supporting it}\errmessage{Not loaded}}
\let\foreignlanguage\selectlanguage
\let\otherlanguage\selectlanguage
\expandafter\let\csname otherlanguage*\endcsname\selectlanguage
\def\bbl@usehooks{}% TODO. Temporary!!
\def\setlocale{\errhelp{Find an armchair, sit down and wait}\errmessage{(babel) Not yet available}}
\let\uselocale\setlocale
\let\locale\setlocale
\let\selectlocale\setlocale
\let\localename\setlocale
\let\textlocale\setlocale
\let\textlanguage\setlocale
\let\languagetext\setlocale
\begingroup
\def\AddBabelHook#1#2{\expandafter\ifx\csname bbl@hook@#2\endcsname\relax
\def\next{\toks1}\else
\def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname####1}\fi}
\ifx\directlua\@undefined
\ifx\XeTeXinputencoding\@undefined\else
\input xebabel.def
\else
\input luababel.def
\openin1 = babel-bl@format.cfg
\ifeof1
\else
\input babel-bl@format.cfg\relax
\fi
\fi
\ifx\directlua\@undefined
\ifx\XeTeXinputencoding\@undefined\else
\input xebabel.def
\else
\input luababel.def
\openin1 = babel-bl@format.cfg
\ifeof1
\else
\input babel-bl@format.cfg\relax
\fi
\fi
\readconfigfile

The configuration file can now be opened for reading.

\openin1 = language.dat
See if the file exists, if not, use the default hyphenation file hyphen.tex. The user will be informed about this.

\edef\languagename{english}\
\ififeof\else
Pattern registers are allocated using count register \texttt{last@language}. Its initial value is 0. The definition of the macro \texttt{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 \texttt{last@language} with the value \texttt{-1}.

\texttt{last@language}\texttt{m@ne}

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.

\loop
\read1 to \bbl@line
\endlinechar`^^M
If the file has reached its end, exit from the loop here. If not, empty lines are skipped. Add 3 space characters to the end of \bbl@line. This is needed to be able to recognize the arguments of \texttt{process@line} later on. The default language should be the very first one.

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

Check for the end of the file. We must reverse the test for \texttt{ifeof} without \texttt{else}. Then reactivate the default patterns, and close the configuration file.

\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
\closein1

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

\if\the\toks@
\errhelp{language.dat loads no language, only synonyms}
\errmessage{Orphan language synonym}
\fi

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

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

## 8 Font handling with fontspec

Add the bidi handler just before luautolatex, 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
\chardef\bbl@bidimode\z@
\DeclareOption{bidi=default}{\chardef\bbl@bidimode=\@ne}
\DeclareOption{bidi=basic}{\chardef\bbl@bidimode=101}
\DeclareOption{bidi=basic-r}{\chardef\bbl@bidimode=102}
\DeclareOption{bidi=bidi}{\chardef\bbl@bidimode=201}
\DeclareOption{bidi=bidi-r}{\chardef\bbl@bidimode=202}
\DeclareOption{bidi=bidi-l}{\chardef\bbl@bidimode=203}
```

With explicit languages, we could define the font at once, but we don’t. Just wait and see if the language is actually activated. \texttt{bbl@font} replaces hardcoded font names inside \..{}family by the corresponding macro \..{}default.

At the time of this writing, fontspec shows a warning about there are languages not available, which some people think refers to babel, even if there is nothing wrong. Here is hack to patch fontspec to avoid the misleading (and mostly unuseful) message.

```latex
\bbl@trace{Font handling with fontspec}
\ifx\ExplSyntaxOn\@undefined\else
\def\bbl@fs@warn@nx#1#2{\texttt{bbl@tempfs is the original macro}}
\def\bbl@fs@warn@nxx#1#2#3{\texttt{bbl@tempfs@nx[#1]{#2}})
\def\bbl@loadfontspec{\texttt{\usepackage{fontspec}}}
\fi
\@onlypreamble\babelfont
\newcommand\babelfont[2]{\texttt{\bbl@foreach{#1}{\expandafter\ifx\csname date##1\endcsname\relax\IfFileExists{babel-##1.tex}{\babelprovide{##1}}}}
\edef\bbl@tempa{#1}
\def\bbl@tempb{#2}
\bbl@loadfontspec
\EnableBabelHook{babel-fontspec}
\bbl@bblfont
```

```latex
\bbl@bblfont[2]{\texttt{\bbl@ifunset{\bbl@tempb family}{}}}
\edef\bbl@tempa{#1}
\def\bbl@tempb{#2}
\texttt{\usedef{\bbl@bblfont}{\texttt{\bbl@ifunset{\bbl@tempb family}{}}}}
```

```latex
\newcommand{\bbl@bblfont}[2]{\texttt{\bbl@ifunset{\bbl@tempb family}{}}}
\texttt{\newcommand{\bbl@bblfont}[2]{\texttt{\edef{\bbl@tempa{#1}{}}}}
\edef{\bbl@tempa{#1}{}}
```

```latex
\edef{\bbl@tempa{#1}{}}
```

```latex
\edef{\bbl@tempa{#1}{}}
```
If the family in the previous command does not exist, it must be defined. Here is how:

```latex
\def\bbl@providefam#1{% 
  % For the default font, just in case:
  \if@ишь \bbl@tempb \else \bbl@tempa \fi 
  % save bbl@rmdflt@
  \let<bbl@famrst@#1>\relax
  % order is relevant. TODO: but sometimes wrong!
}
```

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{% 
  \if@ишь \bbl@tempa \else \bbl@tempb \fi 
  % Flag, to avoid dupl warns
  \let<#1>\relax
  % 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 not set Script and Language for them, so you may consider defining a new family with \verbatim\texttt{babelfont}. See the manual for further details about \verbatim\texttt{babelfont}. Reported\}
}
```

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

```latex
\def\bbl@providefam#1{% 
  \if@ишь \bbl@tempb \else \bbl@tempa \fi 
  % save bbl@rmdflt@
  \let<bbl@famrst@#1>\relax
  % order is relevant. TODO: but sometimes wrong!
}
```
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.
For historical reasons, \LaTeX\ can select two different series (bx and b), for what is conceptually a single one. This can lead to problems when a single family requires several fonts, depending on the language, mainly because ‘substitutions’ with some combinations are not done consistently – sometimes bx/sc is the correct font, but sometimes points to b/n, even if b/sc exists. So, some substitutions are redefined (in a somewhat hackish way, by inspecting if the variant declaration contains >ssub*).

\def\bbl@font@set#1#2{% eg \bbl@mdflt@lang \rmdefault \rmfamily
\bbl@xint{<>}{#1}
\ifin@% 102
\bbl@exp{\\bbl@fontspec@set\#1\expandafter\@gobbletwo#1\#3}
\fi% 'Unprotected' macros return prev values
\def\#2{% eg, \rmdefault{\bbl@mdflt@lang}
\bbl@exp{\\bbl@fontspec@set\#1\expandafter\gobbletwo\#1\#3}}
\fi% TODO - next should be global?, but even local does its job. I'm still not sure -- must investigate:
\def\bbl@fontspec@set#1#2#3#4{% eg \bbl@rmfltr@lang \fnt@opt \fnt@nme \xxfamily
\let\bbl@tempe@bbl@mapselect
\edef\bbl@tempb{\bbl@stripslash#4/}% Catcodes hack (better pass it).
\bbl@exp{\\bbl@replace\\bbl@tempb{\bbl@stripslash\family/}{}}%
\let\bbl@mapselect\relax
\let#4\@empty % Make sure \renewfontfamily is valid
\bbl@exp{%
\let\bbl@temp@fam\relax% eg, 'rmfamily', to be restored below
\let#4\@empty % Make sure \renewfontfamily is valid
\bbl@exp{%
\let\bbl@temp@pfam\relax% eg, 'rmfamily '
\let\bbl@tempfs@nx\relax
\let\bbl@tempfs@nxx\relax
\renewfontfamily\#4%}\enewfontfamily\#4%
\begin{group}
\#4%
\edef\#1{%\f@family}% eg, \bbl@rmfltr@lang{FreeSerif(0)}
\endgroup % TODO. Find better tests:
\bbl@exp{\let\bbl@tempfam\relax}
\let\bbl@mapselect\bbl@tempe}

% font@rst and famrst are only used when there is no global settings, to save and restore de previous families. Not really necessary, but done for optimization.
\def\bbl@font@rst#1#2#3#4{%
\bbl@exp{\let\bbl@tempfam\relax}{\bbl@mapselect\bbl@tempe}%
\except{\font@rst}{\famrst@#4}{\bbl@font@set{#1}{#2}{#3}}
\bbl@csarg\def{famrst@#4}{\bbl@font@set{#1}{#2}{#3}}}

\begin{Font selection}

9 Hooks for XeTeX and LuaTeX
9.1 XeTeX

Unfortunately, the current encoding cannot be retrieved and therefore it is reset always to utf8, which seems a sensible default.
\def\bbl@font@rst#1#2#3#4{%
\bbl@csarg\def{famrst@#4}{\bbl@font@set{#1}{#2}{#3}}
\end{Font selection}

The default font families. They are eurocentric, but the list can be expanded easily with \babelfont.
\long\def\bbl@footnote@o{#1}{#2}{#3}{{\bbl@footnote@o{#1}{#2}{#3}}}\hspace{1em}

Now, the code.

\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\if\@ifnextchar[\bl@footnotetext@o{#1}{#2}{#3}{#4}\@namedef{\bbl@stripslash#1text}\bbl@footnotetext\else\bbl@exp{#1\bbl@footnotetext\bbl@exp{#1\foreignlanguage{#2}}}{#3}{#4}\@namedef{\bbl@stripslash#1text}\bbl@exp{#1\bbl@footnotetext\bbl@exp{#1\foreignlanguage{#2}}}{#3}{#4}\fi}
10 Support for interchar

xetex reserves some values for CJK (although they are not set in \XeLaTeX), so we make sure they are skipped. Define some user names for the global classes, too.

\ifnum xe@alloc@intercharclass<\thr@@\xetable@alloc@intercharclass\thr@@\fi
\chardef \xe@intercharclassdefault@=\z@
\chardef \xe@intercharclasscjkideogram@=\@ne
\chardef \xe@intercharclasscjkleftpunctuation@=\tw@
\chardef \xe@intercharclasscjkrightpunctuation@=\thr@@
\chardef \xe@intercharclassboundary@=4095
\chardef \xe@intercharclassignore@=4096

The machinery is activated with a hook (enabled only if actually used). Here \xe@tempc is pre-set with \xe@usingxeclass, defined below. The standard mechanism based on \originalTeX to save,
set and restore values is used. \count@ stores the previous char to be set, except at the beginning (0) and after \bbl@upto, which is the previous char negated, as a flag to mark a range.

\AddBabelHook{babel-interchar}{beforeextras}{%
@nameuse{bbl@xechars@\languagename}}
\DisableBabelHook{babel-interchar}
\protected\def\bbl@charclass#1{% 
  \ifnum\count@<\z@ \count@-\count@ \loop 
    \bbl@exp{% \\bbl@savevariable{\XeTeXcharclass`\Uchar\count@}}% 
    \XeTeXcharclass\count@ \bbl@tempc 
  \ifnum\count@<`#1\relax \advance\count@\@ne \repeat 
  \else \bbl@savevariable{\XeTeXcharclass`#1} \XeTeXcharclass`#1 \bbl@tempc \fi 
  \count@`#1\relax}

Now the two user macros. Char classes are declared implicitly, and then the macro to be executed at the babel-interchar hook is created. The list of chars to be handled by the hook defined above has internally the form \bbl@usingxeclass\bbl@xechars@punct@english\bbl@charclass{}, (etc.), where \bbl@usingxeclass stores the class to be applied to the subsequent characters. The \ifcat part deals with the alternative way to enter characters as macros (eg, \}). As a special case, hyphens are stored as \bbl@upto, to deal with ranges.

\newcommand\IfBabelIntercharT[1]{% 
% Assume to ignore 
\let\bbl@tempa\@gobble
% Assume to ignore 
\edef\bbl@tempb{\zap@space#1 \@empty} 
\ifx\bbl@KVP@interchar\@nnil
% Assume to ignore 
\bbl@replace\bbl@KVP@interchar{ }{,}
\bbl@foreach\bbl@tempb{% 
\bbl@xin@{,##1,}{,\bbl@KVP@interchar,} 
\ifin@
% Assume to ignore 
\let\bbl@tempa\@firstofone
\fi 
\fi}
\bbl@tempa}

\newcommand\babelcharclass[3]{% 
\EnableBabelHook{babel-interchar} 
\bbl@csarg\newXeTeXintercharclass{xeclass@#2@#1} 
\def\bbl@tempb##1{% 
  \ifx##1\@empty\else 
    \ifx##1-% 
    \bbl@upto 
  \else 
    \bbl@charclass{% 
      \ifcat\noexpand##1\relax\bbl@stripslash##1\else\string##1\fi} 
    \expandafter\bbl@tempb 
  \fi 
}\bbl@ifunset{bbl@xechars@#1}{}{\toks@{\bbl@savevariable{\XeTeXcharclass`\Uchar\count@}}} \toks@{\expandafter\expandafter\expandafter{\csname bbl@xechars@#1\endcsname}} \bbl@csarg\edef{xechars@#1}{\the\toks@ \bbl@usingxeclass\csname bbl@xeclass@#2@#1\endcsname \bbl@tempbw@empty}}
And finally, the command with the code to be inserted. If the language doesn't define a class, then use the global one, as defined above. For the definition there is an intermediate macro, which can be 'disabled' with \bbl@ic@<label>@<lang>.

\def\bbl@ignoreinterchar{\ifnum\language=\l@nohyphenation \expandafter\@gobble \else \expandafter\@firstofone \fi}
\newcommand\babelinterchar[5][]%  
\let\bbl@kv@label\@empty
\bbl@forkv{#1}{\bbl@csarg\edef{kv@##1}{##2}}%
@namedef{\zap@space bbl@xeinter@\bbl@kv@label @#3@#4@#2 \@empty}{\bbl@ignoreinterchar{#5}}%
\bbl@csarg\let{ic@\bbl@kv@label @#2}\@firstofone
\bbl@exp{\\\bbl@for\\bbl@tempa{\zap@space#3 \@empty}}{\bbl@exp{\\\bbl@for\\bbl@tempb{\zap@space#4 \@empty}}}{\XeTeXinterchartoks
\@nameuse{bbl@xeclass@\bbl@tempa @\bbl@ifunset{bbl@xeclass@\bbl@tempa @#2}{}{#2}} %
\@nameuse{bbl@xeclass@\bbl@tempb @\bbl@ifunset{bbl@xeclass@\bbl@tempb @#2}{}{#2}} %
= \expandafter{\csname bbl@ic@\bbl@kv@label @#2\expandafter\endcsname
\csname\zap@space bbl@xeinter@\bbl@kv@label @#3@\@empty\endcsname}}}%
\DeclareRobustCommand\enablelocaleinterchar[1]{\bbl@ifunset{bbl@ic@#1@\languagename}{\bbl@error{unknown-interchar-b}{#1}{}{}}\bbl@csarg\let{ic@#1@\languagename}\@gobble}%
\DeclareRobustCommand\disablelocaleinterchar[1]{\bbl@ifunset{bbl@ic@#1@\languagename}{\bbl@error{unknown-interchar-b}{#1}{}{}}\bbl@csarg\let{ic@#1@\languagename}\@firstofone}⟨\/*xetex|xetex⟩
\providecommand\bbl@provide@intraspace{}
\bbl@trace{Redefinitions for bidi layout}
\def\bbl@sspre@caption{\bbl@exp{\everyhbox{\\bbl@textdir\bbl@cs{wdir@\bbl@main@language}}}}
\ifx\bbl@opt@layout\@nnil\else % if layout=..
\def\bbl@startskip{\ifcase\bbl@thepardir\leftskip\else\rightskip\fi}
\def\bbl@endskip{\ifcase\bbl@thepardir\rightskip\else\leftskip\fi}
\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{txtbabel} as a shorthand for \texttt{tex–xet babel}, which is the bidi model in both \texttt{pdftex} and \texttt{xetex}.

10.1 Layout
Note elements like headlines and margins can be modified easily with packages like \texttt{fancyhdr}, \texttt{typearea} or \texttt{titleps}, and \texttt{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{txtbabel} as a shorthand for \texttt{tex–xet babel}, which is the bidi model in both \texttt{pdftex} and \texttt{xetex}.

⟨\/*xetex|texxet⟩
Implicitly reverses sectioning labels in bidi=basic, because the full stop is not in contact with L numbers any more. I think there must be a better way.
10.2 8-bit TeX

Which start just above, because some code is shared with xetex. Now, 8-bit specific stuff. If just one encoding has been declared, then assume no switching is necessary (1).
10.3 LuaTeX

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

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

The default setup preloads the first language into the format. This is intended mainly for ‘english’, so that it’s available without further intervention from the user. To avoid duplicating it, the following rule applies: if the “0th” language and the first language in language.dat have the same name then just ignore the latter. If there are new synonymous, the are added, but note if the language patterns have not been preloaded they won’t at run time.

Other preloaded languages could be read twice, if they have been preloaded into the format. This is not optimal, but it shouldn’t happen very often – with luatex patterns are best loaded when the document is typeset, and the “0th” language is preloaded just for backwards compatibility.

As of 1.1b, lua(e)tex is taken into account. Formerly, loading of patterns on the fly didn’t work in this format, but with the new loader it does. Unfortunately, the format is not based on babel, and data could be duplicated, because languages are reassigned above those in the format (nothing serious, anyway). Note even with this format language.dat is used (under the principle of a single source), instead of language.def.

Of course, there is room for improvements, like tools to read and 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). FIX - This isn’t true anymore. 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.

This files is read at three places: (1) when plain.def, babel.sty starts, to read the list of available languages from language.dat (for the base option); (2) at hyphen.cfg, to modify some macros; (3) in the middle of plain.def and babel.sty, by babel.def, with the commands and other definitions for luatex (eg, \babelpatterns).

\begin{verbatim}
110
\protect\ifx\AddBabelHook\@undefined % When plain.def, babel.sty starts
\bbl@trace{Read language.dat}
\protect\ifx\bbl@readstream\@undefined
\csname newread\endcsname\bbl@readstream
\fi
\begingroup
\toks@{}
\count@\z@ % 0=start, 1=0th, 2=normal
\def\bbl@process@line#1#2 #3 #4 {%
  \ifx=#1%
  \bbl@process@synonym{#2}%
  \else
  \bbl@process@language{#1#2}{#3}{#4}%
  \ignorespaces
}\def\bbl@manylang{%
  \ifnum\bbl@last>\@ne
  \bbl@info{Non-standard hyphenation setup}%
  \fi
  \bbl@process@language{#1#2}{#3}{#4}%
}\def\bbl@process@language#1#2#3 #4 {%
  \protect\ifx=#1%
  \m@process@synonym{#2}%
  \else
  \m@process@language{#1#2}{#3}{#4}%
  \fi
  \ignorespaces
}\def\m@process@language#1#2#3 #4 {%
  \@ifundefined{zth@#1}{\count@\tw@}{\count@\@ne}%
  \or
  \count@\tw@
  \fi
  \protect\ifnum\count@\=@\tw@
  \protect\expandafter\addlanguage\csname l@#1\endcsname
\endgroup
\end{verbatim}

110
\AddBabelHook{luatex}{loadexceptions}{%
  \input #1\relax
  \def\bbl@tempb##1##2{{##1}{#1}}%
  \expandafter\xdef\csname bbl@hyphendata@	he\language\endcsname
  \expandafter\expandafter\expandafter\bbl@tempb
  \csname bbl@hyphendata@	he\language\endcsname}}
\endinput\fi
%
% Here stops reading code for hyphen.cfg
% The following is read the 2nd time it’s loaded
\begingroup % TODO - to a lua file
\catcode`%=12
\catcode`\%=12
\catcode`\%=12
\catcode`\%=12
\directlua{
  Babel = Babel or {}
  function Babel.bytes(line)
    return line:gsub("(.)",
      function (chr) return unicode.utf8.char(string.byte(chr)) end)
  end
  function Babel.begin_process_input()
    if luatexbase and luatexbase.add_to_callback then
      luatexbase.add_to_callback('process_input_buffer',
        Babel.bytes,'Babel.bytes')
    else
      Babel.callback = callback.find('process_input_buffer')
      callback.register('process_input_buffer',Babel.bytes)
    end
  end
  function Babel.end_process_input ()
    if luatexbase and luatexbase.remove_from_callback then
      luatexbase.remove_from_callback('process_input_buffer','Babel.bytes')
    else
      callback.register('process_input_buffer',Babel.callback)
    end
  end
  function Babel.addpatterns(pp, lg)
    local lg = lang.new(lg)
    local pats = lang.patterns(lg) or ''
    lang.clear_patterns(lg)
    for p in pp:gmatch('[^%s]+') do
      ss = ''
      for i in string.utfcharacters(p:gsub('%d', '')) do
        ss = ss .. '%d?' .. i
      end
      ss = ss:gsub('^[%d%.]*%d?%', '%d?')
      ss = ss:gsub('%d?%.*', '%d?')
      pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
      if n == 0 then
        tex.sprint(
          \[
            [[\string\csname space bbl@info\endcsname{New pattern: }]]
            .. p .. [[]])
        pats = pats .. ' ' .. p
      else
        tex.sprint(
          \[
            [[\string\csname space bbl@info\endcsname{Renew pattern: }]]
            .. p .. [[]])
        end
      end
      lang.patterns(lg, pats)
    end
  end
  Babel.characters = Babel.characters or {}
Babel.ranges = Babel.ranges or {}

function Babel.hlist_has_bidi(head)
    local has_bidi = false
    local ranges = Babel.ranges
    for item in node.traverse(head) do
        if item.id == node.id'glyph' then
            local itemchar = item.char
            local chardata = Babel.characters[itemchar]
            local 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
                if dir and (dir == 'al' or dir == 'r') then
                    has_bidi = true
                end
            end
        end
    end
    return has_bidi
end

function Babel.set_chranges_b (script, chrng)
    if chrng == '' then return end
    texio.write('Replacing ' .. script .. ' script ranges')
    Babel.script_blocks[script] = {}
    for s, e in string.gmatch(chrng..' ', '(.-)%.%.(.-)%s') do
        table.insert(Babel.script_blocks[script], {tonumber(s,16), tonumber(e,16)})
    end
end

function Babel.discard_sublr(str)
    if str:find( \[
    ]indexentry\]) and
    str:find( \[
    ]babelsublr\]) then
        str = str:gsub( \[
    ]babelsublr%s*(%b{})\],
    function(m) return m:sub(2,-2) end )
    end
    return str
end

\AddBabelHook{luatex}{patterns}{%\ifundefined{bbl@hyphendata@the\language}{%\def\bbl@elt##1##2##3##4{%\ifnum##2=\csname l@#2\endcsname % #2=spanish, dutch:OT1...
\def\bbl@tempb(#3)%\expandafter\expandafter\expandafter\bbl@luapatterns\csname bbl@hyphendata@the\language\endcsname}%\ifx\bbl@tempb\@empty%\else % if not a synonymous\def\bbl@tempc{{##3}{##4}}%\fi\bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%\fi}\bbl@languages%\@ifundefined{bbl@hyphendata@the\language}{%\{\bbl@info{No hyphenation patterns were set for\% language '##2'. Reported}}%\expandafter\expandafter\expandafter\bbl@luapatterns\csname bbl@hyphendata@the\language\endcsname}%\ifx\bbl@patterns@\@empty%\else%\let\bbl@patterns@\@empty%\fi\@ifundefined{bbl@patterns@#1}%\@empty%\{\directlua{ Babel.addpatterns(\csname bbl@patterns@#1\endcsname), \number\language} %\xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%%\endgroup%}%\bbl@exp{%\bbl@ifunset{bbl@prehc@\languagename}{}%\bbl@ifblank{\bbl@cs{prehc@\languagename}}{}%\prehyphenchar=\bbl@cl{prehc}\relax}}}%This macro adds patterns. Two macros are used to store them: \bbl@patterns@ for the global ones and \bbl@patterns@<lang> for language ones. We make sure there is a space between words when multiple commands are used.%\onlypreamble\bbl@patterns%\AtEndOfPackage{%\newcommand\bbl@patterns[2]{%\protect\def\bbl@patterns@{\bbl@patterns@\space%\ifx\@empty#2%\empty%\else%\edef\bbl@tempb{\zap@space#1 \@empty}%\bbl@for\bbl@tempa\bbl@tempb{%\bbl@fixname\bbl@tempa
\bbl@iflanguage\bbl@tempa{%\protectededef\patterns@\bbl@tempa{%\ifundefined{\bbl@patterns@\bbl@tempa}{%\@empty%\{\csname bbl@patterns@\bbl@tempa@endcsname\space}%}}}%\bbl@patterns@\space%\empty%\bbl@warning{%\string\selectlanguage\space and \string\bbl@patterns\space or some patterns will not\% be taken into account. Reported}%}\fi}}%\ifx\@empty#1%\protectededef\bbl@patterns@{\bbl@patterns@\space2}%\else%\edef\bbl@tempb{\zap@space1 \@empty}%\bbl@for\bbl@tempa\bbl@tempb{%\bbl@fixname\bbl@tempa
\bbl@iflanguage\bbl@tempa{%\bbl@csarg\protectededef\patterns@\bbl@tempa{%\ifundefined{\bbl@patterns@\bbl@tempa}{%\@empty}}}%}}%\csname bbl@patterns@\bbl@tempa@endcsname\space%}}}
10.4 Southeast Asian scripts

First, some general code for line breaking, used by \bipelipathyphenation. Replace regular (ie, implicit) discretionary by spaceskips, based on the previous glyph (which I think makes sense, because the hyphen and the previous char go always together). Other discretionary are not touched. See Unicode UAX 14.

\% TODO - to a lua file
\directlua{
Babel = Babel or {}
Babel.linebreaking = Babel.linebreaking or {}
Babel.linebreaking.before = {}
Babel.linebreaking.after = {}
Babel.locale = {} \% Free to use, indexed by \localeid
function Babel.linebreaking.add_before(func, pos)
tex.print((\[\noexpand\csname bbl@luahyphenate\endcsname\]))
if pos == nil then
  table.insert(Babel.linebreaking.before, func)
else
  table.insert(Babel.linebreaking.before, pos, func)
end
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.localeProps[\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.localeProps[\the\localeid].intrapenalty = #1
}
\begingroup
\catcode`%=12
\catcode`^=14
\catcode`\%=12
\catcode`-=12
\edef\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 * 6536
  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 > 0x0C99 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
}
\bbl@luahyphenate

10.5 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.

\catcode`%=14
\edef\bbl@cjkintraspacerelax
\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 * 6536
    local last_class = nil
    local last_lang = nil
    for item in node.traverse(head) do
      if item.id == GLYPH then
        local lang = item.lang
        -- Insert glue, spaceskip, penalty here
        node.remove(head, item)
      end
    end
  end
}
local LOCALE = node.get_attribute(item, Babel.attr_locale)
local props = Babel.locale_props[LOCALE]

local class = Babel.cjk_class[item.char].c

if props.cjk_quotes and props.cjk_quotes[item.char] then
class = props.cjk_quotes[item.char]
end

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)
ode.insert_before(head, item, n)
end

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

lang.hyphenate(head)

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

}}}
for k, func in ipairs(Babel.linebreaking.after) do
    func(head)
end
if Babel.sea_enabled then
    Babel.sea_disc_to_space(head)
end
\endgroup
\def\bbl@provide@intraspace{%
\bbl@ifunset{bbl@intsp@\languagename}{}%
\if\expandafter{\ifx\csname bbl@intsp@\languagename\endcsname@empty\else\bbl@xintp{c}/\bbl@cl\{lnbrk}\fi}\fi%
\ifin@ % cjk
    \bbl@cjkintraspacet\directlua{
        \ifnum\bbl@bidimode>100 \ifnum\bbl@bidimode<200
            \def\bblar@chars{0628,0629,062A,062B,062C,062D,062E,062F,0630,0631,0632,0633,0634,0635,0636,0637,0638,0639,063A,063B,063C,063D,063E,063F,0640,0641,0642,0643,0644,0645,0646,0647,0649}
            \def\bblar@elongated{0626,0628,062A,062B,0633,0634,0635,0636,0637,0638,0639,063A,063B,063C,063D,063E,063F,0641,0642,0643,0644,0645,0646,0647,0649}
        \fi
    \fi}
\bbl@cjkintraspacet\directlua{
    \if\expandafter{\bbl@KVP@intrapenalty}@nil
        \bbl@intrapenalty0@nil
    \fi
    \if\expandafter{\bbl@KVP@intrapenalty}@nil
        \bbl@intrapenalty0@nil
    \fi
\endgroup
\def\bblar@chars{0628,0629,062A,062B,062C,062D,062E,062F,0630,0631,0632,0633,0634,0635,0636,0637,0638,0639,063A,063B,063C,063D,063E,063F,0640,0641,0642,0643,0644,0645,0646,0647,0649}
\def\bblar@elongated{0626,0628,062A,062B,0633,0634,0635,0636,0637,0638,0639,063A,063B,063C,063D,063E,063F,0641,0642,0643,0644,0645,0646,0647,0649}
\begin{group}
\catcode`_=11 \catcode`:=11
\gdef\bblar@nofswarn{\gdef\msg_warning:nnx##1##2##3{}}
\endgroup
\gdef\bbl@arabicjust{% TODO. Allow for several locales.
10.6 Arabic justification
WIP. \bbl@arabicjust is executed with both elongated an kashida. This must be fine tuned. The attribute kashida is set by transforms with kashida-
\let\bbl@arabicjust\relax
\newattribute\bbl@kashida
\directlua{ Babel.attr_kashida = luatexbase.registernumber'\bbl@kashida' }%
\bbl@kashida=\z@
\bbl@patchfont{{\bbl@parsejalt}}%
\directlua{Babel.arabic.elong_map = Babel.arabic.elong_map or {}
Babel.arabic.elong_map[\the\localeid] = {};
\luatexbase.add_to_callback('post_linebreak_filter',
\Babel.arabic.justify, 'Babel.arabic.justify')
\luatexbase.add_to_callback('hpack_filter',
\Babel.arabic.justify_hbox, 'Babel.arabic.justify_hbox')}
%
Save both node lists to make replacement. TODO. Save also widths to make computations.
\def\bblar@fetchjalt#1#2#3#4{%
\bbl@exp{\bbl@foreach{#1}}{%
\bbl@ifunset{bblar@JE@##1}%
\{\setbox\z@\hbox{\textdir TRT ^^^^200d\char"##1#2}}%
\{\setbox\z@\hbox{\textdir TRT ^^^^200d\char"\@nameuse{bblar@JE@##1}#2}}%
\directlua{%
local last = nil
for item in node.traverse(tex.box[0].head) do
if item.id == node.id'glyph' and item.char > 0x600 and
not (item.char == 0x200D) then
last = item
end
end
Babel.arabic.#3['##1#4'] = last.char
}}}
%
Elongated forms. Brut force. No rules at all, yet. The ideal: look at jalt table. And perhaps other
tables (falt?, cswh?). Whataboutkaf? And diacritic positioning?
\gdef\bblar@parsejalt{%
\ifx\addfontfeature\@undefined\else
\bbl@xin{/e}{/\bbl@cl{lnbrk}}%
\ifin@
\directlua{%
if Babel.arabic.elong_map[\the\localeid][\fontid\font] == nil then
Babel.arabic.elong_map[\the\localeid][\fontid\font] = {}
tex.print([\{\string\csname space bbl@parsejalti\endcsname\}])
end
\fi
%}
\fi
\gdef\bblar@parsejalti{%
\begingroup
\let\bbl@parsejalt\relax % To avoid infinite loop
\edef\bbl@tempb{\fontid\font}
\bblar@nofswarn
\bblar@fetchjalt\bblar@elongated{}{from}{}
\bblar@fetchjalt\bblar@chars{^^^^064a}{from}{a}% Alef maksura
\bblar@fetchjalt\bblar@chars{^^^^0649}{from}{y}% Yeh
\addfontfeature{RawFeature=+jalt}%
\@namedef{bblar@JE@0643}{06AA}% todo: catch medial kaf
\bblar@fetchjalt\bblar@elongated{}{dest}{}
\bblar@fetchjalt\bblar@chars{^^^^064a}{dest}{a}%
\bblar@fetchjalt\bblar@chars{^^^^0649}{dest}{y}%
\directlua{%
for k, v in pairs(Babel.arabic.from) do
if Babel.arabic.dest[k] and
not (Babel.arabic.from[k] == Babel.arabic.dest[k]) then
Babel.arabic.elong_map[\the\localeid][\bbl@tempb] = Babel.arabic.dest[k]
end
end
\endgroup
}
The actual justification (inspired by \texttt{chickenize}).

\begingroup
\catcode`#=11
\catcode`~=11
\directlua{
Babel.arabic = Babel.arabic or {}
Babel.arabic.from = {}
Babel.arabic.dest = {}
Babel.arabic.justify_factor = 0.95
Babel.arabic.justify_enabled = true
Babel.arabic.kashida_limit = -1

function Babel.arabic.justify(head)
if not Babel.arabic.justify_enabled then return head end
for line in node.traverse_id(node.id'\hlist', head) do
  Babel.arabic.justify_hlist(head, line)
end
return head
end

function Babel.arabic.justify_hbox(head, gc, size, pack)
local has_inf = false
if Babel.arabic.justify_enabled and pack == 'exactly' then
  for n in node.traverse_id(12, head) do
    if n.stretch_order > 0 then has_inf = true end
  end
  if not has_inf then
    Babel.arabic.justify_hlist(head, nil, gc, size, pack)
  end
end
return head
end

function Babel.arabic.justify_hlist(head, line, gc, size, pack)
local d, new
local k_list, k_item, pos_inline
local width, width_new, full, k_curr, wt_pos, goal, shift
local subst_done = false
local elong_map = Babel.arabic.elong_map
local cnt
local last_line
local GLYPH = node.id'\glyph'
local KASHIDA = Babel.attr_kashida
local LOCALE = Babel.attr_locale
if line == nil then
  line = {}
  line.glue_sign = 1
  line.glue_order = 0
  line.head = head
  line.shift = 0
  line.width = size
end
if line == nil then
  line = {
  line.glue_sign = 1
  line.glue_order = 0
  line.head = head
  line.shift = 0
  line.width = size
end
% Exclude last line. todo. But-- it discards one-word lines, too!
% ? Look for glue = 12:15
if (line.glue_sign == 1 and line.glue_order == 0) then
elongs = {}  % Stores elongated candidates of each line
k_list = {}  % And all letters with kashida
pos_inline = 0  % Not yet used

for n in node.traverse_id(GLYPH, line.head) do
    pos_inline = pos_inline + 1  % To find where it is. Not used.

    % Elongated glyphs
    if elong_map then
        local locale = node.get_attribute(n, LOCALE)
        if elong_map[locale] and elong_map[locale][n.font] and
            elong_map[locale][n.font][n.char] then
            table.insert(elongs, {node = n, locale = locale})
            node.set_attribute(n.prev, KASHIDA, 0)
        end
    end

    % Tatwil
    if Babel.kashida_wts then
        local k_wt = node.get_attribute(n, KASHIDA)
        if k_wt > 0 then  % todo. parameter for multi inserts
            table.insert(k_list, {node = n, weight = k_wt, pos = pos_inline})
        end
    end
end  % of node.traverse_id

if #elongs == 0 and #k_list == 0 then goto next_line end
full = line.width
shift = line.shift
goal = full * Babel.arabic.justify_factor % A bit crude
width = node.dimensions(line.head)  % The 'natural' width

% == Elongated ==
% Original idea taken from 'chikenize'
while (#elongs > 0 and width < goal) do
    subst_done = true
    local x = #elongs
    local curr = elongs[x].node
    local oldchar = curr.char
    curr.char = elong_map[elongs[x].locale][curr.font][curr.char]
    width = node.dimensions(line.head)  % Check if the line is too wide
    if width > goal then
        curr.char = oldchar
        break
    end
    % If continue, pop the just substituted node from the list:
    table.remove(elongs, x)
end

% == Tatwil ==
if #k_list == 0 then goto next_line end
width = node.dimensions(line.head)  % The 'natural' width
k_curr = #k_list  % Traverse backwards, from the end
wt_pos = 1
while width < goal do
    subst_done = true
    k_item = k_list[k_curr].node
    if k_list[k_curr].weight == Babel.kashida_wts[wt_pos] then
        d = node.copy(k_item)
    end
    subst_done = subst_done and d == k_item
    wt_pos = wt_pos + 1
    k_curr = k_curr - 1
    if subst_done then break end
end
\begin{verbatim}
\  d.char = 0x0640
\  d.yoffset = 0 % TODO. From the prev char. But 0 seems safe.
\  d.xoffset = 0
\end{verbatim}

\begin{verbatim}
\  line.head, new = node.insert_after(line.head, k_item, d)
\  width_new = node.dimensions(line.head)
\  if width > goal or width == width_new then
\    node.remove(line.head, new) % Better compute before
\    break
\  end
\  if Babel.fix_diacr then
\    Babel.fix_diacr(k_item.next)
\  end
\  width = width_new
\end{verbatim}

\begin{verbatim}
\  if k_curr == 1 then
\    k_curr = #k_list
\    wt_pos = (wt_pos >= table.getn(Babel.kashida_wts)) and 1 or wt_pos+1
\  else
\    k_curr = k_curr - 1
\  end
\end{verbatim}

\begin{verbatim}
\  ::next_line::<
\end{verbatim}

\begin{verbatim}
\  if subst_done and not gc then
\    d = node.hpack(line.head, full, 'exactly')
\    d.shift = shift
\    node.insert_before(head, line, d)
\    node.remove(head, line)
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\  \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
\  \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ckeckstdfonts}
\  \DisableBabelHook{babel-fontspec}
\end{verbatim}

\section{Common stuff}

\AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
\AddBabelHook{babel-fontspec}{beforestart}{\bbl@ckeckstdfonts}
\DisableBabelHook{babel-fontspec}

\end{document}
10.8 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 function `Babel.locale_map` which just traverses the node list to carry out the replacements. The table `loc_to_scr` stores the script range for each locale (whose id is the key), copied from this table (so that it can be modified on a locale basis); there is an intermediate table named `chr_to_loc` built on the fly for optimization, which maps a chart to the locale. This locale is then used to get the `language` 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 discretionarys are handled in a special way.

```lua
Babel.script_blocks = {
    ['dflt'] = {},
    ['Arab'] = {0x0600, 0x06FF, 0x08A0, 0x08FF, 0x0750, 0x077F},
    ['Armn'] = {0x0530, 0x058F},
    ['Beng'] = {0x0980, 0x09FF},
    ['Copt'] = {0x03E2, 0x03EF, 0x2C80, 0x2CFF, 0x0182E, 0x0182F},
    ['Cyril'] = {0x0400, 0x04FF, 0x0550, 0x055F, 0x01C0, 0x01C8F},
    ['Devan'] = {0x0900, 0x09FF},
    ['Ethi'] = {0x0c00, 0x0c7f, 0x01ff, 0x0200, 0x02ff},
    ['Geor'] = {0x0b00, 0x0b7f},
    ['Grek'] = {0x0370, 0x03e1, 0x03f0, 0x03ff},
    ['Hans'] = {0x2e80, 0x2e9f, 0x3000, 0x303f, 0x31c0, 0x31ff},
    ['Hebr'] = {0x0590, 0x05ff},
    ['Hebrew'] = {0x03f0, 0x03ff},
    ['Hind'] = {0x0c00, 0x0c7f},
    ['Hira'] = {0x0980, 0x09ff},
    ['Knda'] = {0x0b00, 0x0b7f},
    ['Khmr'] = {0x0c00, 0x0c7f},
    ['Korean'] = {0x1100, 0x11ff, 0x3000, 0x303f, 0x31c0, 0x31ff},
    ['Laoo'] = {0x0c00, 0x0c7f},
    ['Latein'] = {0x0c00, 0x0c7f},
    ['Telugu'] = {0x0c00, 0x0c7f},
    ['Thai'] = {0x0c00, 0x0c7f},
    ['Telugu'] = {0x0c00, 0x0c7f},
    ['Vaii'] = {0x0c00, 0x0c7f},
    ['Yi'] = {0x0c00, 0x0c7f},
    ['Zwt'] = {0x0c00, 0x0c7f},
}

Babel.script_blocks.Cyrs = Babel.script_blocks.Cyril
```
function Babel.locale_map(head)
  if not Babel.locale_mapped then return head end
  local LOCALE = Babel.attr_locale
  local GLYPH = node.id('glyph')
  local inmath = false
  local toloc_save
  for item in node.traverse(head) do
    local toloc
    if not inmath and item.id == GLYPH then
      % Optimization: build a table with the chars found
      if Babel.chr_to_loc[item.char] then
        toloc = Babel.chr_to_loc[item.char]
      else
        for lc, maps in pairs(Babel.loc_to_scr) do
          for _, rg in pairs(maps) do
            if item.char >= rg[1] and item.char <= rg[2] then
              Babel.chr_to_loc[item.char] = lc
              toloc = lc
              break
            end
          end
        end
        % Treat composite chars in a different fashion, because they
        % 'inherit' the previous locale.
        if (item.char >= 0x0300 and item.char <= 0x036F) or
          (item.char >= 0x1AB0 and item.char <= 0x1AFF) or
          (item.char >= 0x1DC0 and item.char <= 0x1DFF) then
          Babel.chr_to_loc[item.char] = -2000
          toloc = -2000
        end
        if not toloc then
          Babel.chr_to_loc[item.char] = -1000
        end
      end
    end
    if toloc == -2000 then
      toloc = toloc_save
    elseif toloc == -1000 then
      toloc = nil
    end
    if toloc and Babel.locale_props[toloc] and
      Babel.locale_props[toloc].letters and
      tex.getcatcode(item.char) \string== 11 then
      toloc = nil
    end
    if toloc and Babel.locale_props[toloc].script
      and Babel.locale_props[node.get_attribute(item, LOCALE)].script ==
      Babel.locale_props[node.get_attribute(item, LOCALE)].script then
      toloc = nil
    end
    if toloc then
      if Babel.locale_props[toloc].lg then
        item.lang = Babel.locale_props[toloc].lg
        node.set_attribute(item, LOCALE, toloc)
      end
      if Babel.locale_props[toloc]['/..item.font'] then
        item.font = Babel.locale_props[toloc]['/..item.font']
      end
    end
  end
end
The code for \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{charproperty-only-vertical}{}{}{}\%}
  \fi}
\newcommand{\bbl@chprop}[3]{{\the\count@}{\count@\relax
  \ifnum\count@<#1\relax
    \advance\count@\@ne
  \repeat}
\def{\bbl@chprop@direction#1}{{\directlua{\Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
      Babel.characters[\the\count@][d'] = '#1'
    }}}
\let{\bbl@chprop@bc}{\bbl@chprop@direction}
\def{\bbl@chprop@mirror#1}{{\directlua{\Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
      Babel.characters[\the\count@][m'] = '\number#1'
    }}}
\let{\bbl@chprop@bmg}{\bbl@chprop@mirror}
\def{\bbl@chprop@linebreak#1}{{\directlua{\Babel.cjk_characters[\the\count@] = Babel.cjk_characters[\the\count@] or {}
      Babel.cjk_characters[\the\count@][c'] = '#1'
    }}}
\let{\bbl@chprop@lb}{\bbl@chprop@linebreak}
\def{\bbl@chprop@locale#1}{{\directlua{\Babel.chr_to_loc = Babel.chr_to_loc or {}
      Babel.chr_to_loc[\the\count@] = \bbl@ifblank{#1}{-1000}{\the\bbl@cs{id@@#1}}\space
    }}}
\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). The Lua code is below.

\begin{verbatim}
\directlua{
  Babel.nohyphenation = \the\l@nohyphenation
}
\end{verbatim}

Now the \TeX high level interface, which requires the function defined above for converting strings to functions returning a string. These functions handle the \verb{\n} syntax. For example, \verb{pre=(1){1}-}
becomes
\begin{verbatim}
function(m) return m[1]..m[1]..'-' end
\end{verbatim}
, where \( m \) are the matches returned after applying the pattern. With a mapped capture, the functions are similar to
\begin{verbatim}
function(m) return Babel.capt_map(m[1], 1) end
\end{verbatim}
, where the last argument identifies the mapping to be applied to \( m[1] \). The way it is carried out is somewhat tricky, but the effect in not
dissimilar to Lua \texttt{load} – save the code as string in a \TeX{} macro, and expand this macro at the
appropriate place. As \texttt{\textbackslash{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}
\\begin{group}
\catcode``\~\=12
\catcode``\%\=12
\catcode``\&\=14
\catcode``\\=12
\\end{group}
\ifx\font@\relax
\font@\begin{group}
\font\string\set@math@font
\font\character\end{group}
\fi
\begin{verbatim}
\gdef\babelprehyphenation{&\@ifnextchar[{bl@settransform{0}}{bl@settransform{0}\[]}}
\gdef\babelposthyphenation{&\@ifnextchar[{bl@settransform{1}}{bl@settransform{1}\[]}}
\gdef\bbl@settransform#1\[#2\]#3#4#5{
  &\ifcase#1
    \bbl@activateprehyphen
  \or
    \bbl@activateposthyphen
  \fi
  \begingroup
    \def\babeltempa{\bbl@add@list\babeltempb}&\%
    \let\babeltempb\@empty
    \def\bbl@tempa{#5}&\%
    \bbl@replace\bbl@tempa{,}{ ,}&\%
    \expandafter\bbl@foreach\expandafter{\bbl@tempa}{&\%
      \bbl@ifsamestring{##1}{remove}{\bbl@add@list\babeltempb{nil}}&\%
      {\directlua{
        local rep = \[=\[##1\]=\]
        rep = rep:gsub('^%s*(remove)%s*$', 'remove = true')
        rep = rep:gsub('^%s*(insert)%s*$', 'insert = true, ')'
        rep = rep:gsub('[^%s,]*', Babel.capture_func)
        if #1 == 0 or #1 == 2 then
          rep = rep:gsub( '(space)%s*=%s*([%d%.-]+)%s+([%d%.-]+)%s+([%d%.-]+)',
            'space = {' .. '%2, %3, %4' .. '}')
          rep = rep:gsub( '(spacefactor)%s*=%s*([%d%.-]+)%s+([%d%.-]+)%s+([%d%.-]+)',
            'spacefactor = {' .. '%2, %3, %4' .. '}')
          rep = rep:gsub( '(kashida)%s*=%s*([^%s,]*)', Babel.capture_kashida)
          else
            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)
          end
          tex.print([[\string\babeltempa{]] .. rep .. [[}}])}
        }}&\%
      }&\%
    \bbl@foreach\babeltempb{&\%
      \bbl@forkv{{##1}}{&\%
        \in@{,####1,}{,nil,step,data,remove,insert,string,no,pre,,}{no,post,penalty,kashida,space,spacefactor,}{&\%
      }&\%
    }&\%
    \let\bbl@kv@attribute\relax
    \let\bbl@kv@label\relax
    \let\bbl@kv@fonts\@empty
    \bbl@forkv{#2}{\bbl@csarg\edef{kv@##1}{##2}}&\%
    \ifx\bbl@kv@fonts\@empty\else\bbl@settransfont\fi
    \ifx\bbl@kv@attribute\relax
      \ifx\bbl@kv@label\relax
        \bbl@exp{\bl@trim@def\bl@kv@fonts{\bbl@kv@fonts}}&\%
      \fi
    \fi
  \endgroup
\end{verbatim}
\end{verbatim}
\end{verbatim}

127
\bbl@replace\bbl@kv@fonts{ }{,}&%
\edef\bbl@kv@attribute{bbl@ATR@\bbl@kv@label @#3@\bbl@kv@fonts}&%
\count@\z@
\def\bbl@elt##1##2##3{&%
  \bbl@ifsamestring{#3,\bbl@kv@label}{##1,##2}&%
  {\bbl@ifsamestring{\bbl@kv@fonts}{##3}&%
    {\count@\@ne}&%
    {\bbl@error{font-conflict-transforms}{}{}{}}}&%
  }&%
\bbl@transfont@list
\ifnum\count@=\z@
\bbl@exp{\global\\bbl@add\\bbl@transfont@list}{\bbl@elt{#3}{\bbl@kv@label}{\bbl@kv@fonts}}&%
\fi
\bbl@ifunset{\bbl@kv@attribute}&%
{\global\bbl@carg\newattribute{\bbl@kv@attribute}}&%
{}&%
\global\bbl@carg\setattribute{\bbl@kv@attribute}@ne
\else
\edef\bbl@kv@attribute{\expandafter\bbl@stripslash\bbl@kv@attribute}&%
\fi
\directlua{
local lbkr = Babel.linebreaking.replacements[#1]
local u = unicode.utf8
local id, attr, label
if #1 == 0 then
  id = \the\csname bbl@id@@#3\endcsname\space
else
  id = \the\csname l@#3\endcsname\space
end
\if\bbl@kv@attribute\relax
  attr = -1
\else
  attr = luatexbase.registernumber'\bbl@kv@attribute'
\fi
\if\bbl@kv@label\relax
 &% Same refs:
  label = [==\[bl@kv@label\]==]
\fi
&% Convert pattern:
local patt = string.gsub([==[#4]==], '%s', '')
if #1 == 0 then
  patt = string.gsub(patt, '[]', ' ')
end
if not u.find(patt, '()', nil, true) then
  patt = '()' .. patt .. '()'
end
if #1 == 1 then
  patt = string.gsub(patt, '%(%)%^', '^()')
  patt = string.gsub(patt, '%$%(%)', '()$')
end
patt = u.gsub(patt, '{(.)}', function (n)
  return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
end)
patt = u.gsub(patt, '{(%x%x%x%x+)}', function (n)
  return u.gsub(u.char(tonumber(n, 16)), '(%p)', '%%%1')
end)
lbkr[id] = lbkr[id] or {}
table.insert(lbkr[id],
  { label=label, attr=attr, pattern=patt, replace={\babeltempb}})
}&%
The following experimental (and unfinished) macro applies the prehyphenation transforms for the current locale to a string (characters and spaces) and processes it in a fully expandable way (among other limitations, the string can’t contain \[==\]). The way it operates is admittedly rather cumbersome: it converts the string to a node list, processes it, and converts it back to a string. The lua code is in the lua file below.

\newcommand{\localeprehyphenation}[1]{
\directlua{ Babel.string_prehyphenation([==#1]==), \the\localeid} }

The following experimental (and unfinished) macro applies the prehyphenation transforms for the current locale to a string (characters and spaces) and processes it in a fully expandable way (among other limitations, the string can’t contain \[==\]). The way it operates is admittedly rather cumbersome: it converts the string to a node list, processes it, and converts it back to a string. The lua code is in the lua file below.

\newcommand{\localeprehyphenation}[1]{
\directlua{ Babel.string_prehyphenation([==#1]==), \the\localeid} }
10.9 Bidi

As a first step, add a handler for bidi and digits (and potentially other processes) just before \luatexload is applied, which is loaded by default by \BfX. Just in case, consider the possibility it has not been loaded.

\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',
    'luaotfload.node_processor') or nil)
%
luatexbase.add_to_callback('hpack_filter',
  Babel.pre_otfload_h,
  'Babel.pre_otfload_h',
  luatexbase.priority_in_callback('hpack_filter',
    'luaotfload.node_processor') or nil)
}

The basic setup. The output is modified at a very low level to set the \bodydir to the \pagedir. Sadly, we have to deal with boxes in math with basic, so the \bbl@mathboxdir hack is activated every math with the package option bidi=.

\ifnum\bbl@bidimode=1
  \let\bbl@beforeforeign\leavevmode
  \AtEndOfPackage{\EnableBabelHook{babel-bidi}}
  \RequirePackage{luatexbase}
  \bbl@activate@preotf
  \directlua{
    require('babel-data-bidi.lua')
    \ifcase\expandafter\@gobbletwo\the\bbl@bidimode\or
      require('babel-bidi-basic.lua')
    \or
      require('babel-bidi-basic-r.lua')
    \fi}
  \newattribute\bbl@attr@dir
  \directlua{ \bbl@attr_dir = luatexbase.registernumber'\bbl@attr@dir' }
  \bbl@exp{'\output{\bodydir\pagedir\the\output}'}
\fi
RTL text inside math needs special attention. It affects not only to actual math stuff, but also to `tabular`, which is based on a fake math.

```latex
\ifnum\bbl@bidimode>\z@ % Any bidi=
  \def\bbl@inside_math{0}%
  \def\bbl@every_math{\def\bbl@inside_math{1}}
  \def\bbl@every_display{\def\bbl@inside_math{2}}
  \frozen@every_math\expandafter{\def\bbl@inside_math{1}}
  \frozen@every_display\expandafter{\def\bbl@inside_math{1}}
  \AtBeginDocument{
    \directlua{
      \function Babel.math_box_dir(head)
        if not (token.get_macro('bbl@inside_math') == '0') then
          if Babel.hlist.has_bidi(head) then
            local d = node.new(node.id'dir')
            d.dir = '+TRT'
            node.insert_before(head, node.has_glyph(head), d)
            local inmath = false
            for item in node.traverse(head) do
              if item.id == 11 then
                inmath = (item.subtype == 0)
              elseif not inmath then
                node.set_attribute(item,
                  Babel.attr_dir, token.get_macro('bbl@thedir'))
              end
            end
          end
        end
      return head
    }
```

131
10.10 Layout

Unlike xetex, luatex requires only minimal changes for right-to-left layouts, particularly in monolingual documents (the engine itself reverses boxes – including column order or headings –, margins, etc.) with bidi=’basic’, without having to patch almost any macro where text direction is relevant.

Still, there are three areas deserving special attention, namely, tabular, math, and graphics, text and intrinsically left-to-right elements are intermingled. I’ve made some progress in graphics, but they’re essentially hacks; I’ve also made some progress in ’tabular’, but when I decided to tackle math (both standard math and ’amsmath’) the nightmare began. I’m still not sure how ’amsmath’ should be modified, but the main problem is that, boxes are “generic” containers that can hold text, math, and graphics (even at the same time; remember that inline math is included in the list of text nodes marked with ’math’ (11) nodes too). 

\@hangfrom is useful in many contexts and it is redefined always with the layout option.

There are, however, a number of issues when the text direction is not the same as the box direction (as set by \bodydir), and when \parbox and \hangindent are involved. Fortunately, latest releases of luatex simplify a lot the solution with \shapemode.

With the issue #15 I realized commands are best patched, instead of redefined. With a few lines, a modification could be applied to several classes and packages. Now, tabular seems to work (at least in simple cases) with array, tabularx, hhline, colortbl, longtable, booktabs, etc. However, dcolumn still fails.
Babel.digits[\the\localeid] =

if not Babel.numbers then
    function Babel.numbers(head)
        local LOCALE = Babel.attr_locale
        local GLYPH = node.id'glyph'
        local inmath = false
        for item in node.traverse(head) do
            if not inmath and item.id == GLYPH then
                local temp = node.get_attribute(item, LOCALE)
                if Babel.digits[temp] then
                    local chr = item.char
                    if chr > 47 and chr < 58 then
                        item.char = Babel.digits[temp][chr-47]
                    end
                elseif item.id == node.id'math' then
                    inmath = (item.subtype == 0)
                end
            end
        end
        return head
    end
end

% == transforms ==
\ifx\bbl@KVP@transforms@nnil\else
    \def\bbl@elt##1##2##3{%
        \in@{$transforms.}{$##1}%
        \ifin@
            \def\bbl@tempa{##1}%
            \bbl@replace\bbl@tempa{transforms.}{}%
            \bbl@carg\bbl@transforms{babel\bbl@tempa}{##2}{##3}%
        \fi}%
    \csname bbl@inidata@\languagename\endcsname
    \bbl@release@transforms% \relax closes the last item.
\fi

% Start tabular here:
\def\localerestoredirs{%
    \ifcase\bbl@thetextdir
        \ifnum\textdirection=\z@\else\textdir TLT\fi
    \else
        \ifnum\textdirection=\@ne\else\textdir TRT\fi
    \fi
    \ifcase\bbl@thepardir
        \ifnum\pardirection=\z@\else\pardir TLT\bodydir TLT\fi
    \else
        \ifnum\pardirection=\@ne\else\pardir TRT\bodydir TRT\fi
    \fi
}%
\IfBabelLayout{tabular}{
    \chardef\bbl@tabular@mode\tw@}% All RTL
    \IfBabelLayout{notabular}{
        \chardef\bbl@tabular@mode\z@}%
        \chardef\bbl@tabular@mode\@ne}% Mixed, with LTR cols
     \ifnum\bbl@bidimode=\@ne % Any lua bidi= except default=1
        \ifcase\bbl@tabular@mode\or \@one % 1
            \let\bbl@parabefore\relax
            \AddToHook{para/before}{\bbl@parabefore}
        \AtBeginDocument{%
            \bbl@replace\bbl@tabular{($)\${%}
            \def\bbl@insidemath{0}%
            \def\bbl@parabefore{\localerestoredirs}%
        \fi
    \fi
}
Very likely the output routine must be patched in a quite general way to make sure the bodydir is set to pagedir. Note outside output they can be different (and often are). For the moment, two ad hoc changes.

Omega provided a companion to \mathdir (\nextfakemath) for those cases where we did not want it to be applied, so that the writing direction of the main text was left unchanged. \bbllnextfake is an attempt to emulate it, because luatex has removed it without an alternative. Also, \hangindent does not honour direction changes by default, so we need to redefine \@hangfrom.

\ifnum\bbl@bidimode>\z@ % Any bidi=
\def\bbllnextfake##1{% non-local changes, use always inside a group!
\bbllexp{%  
\def\\bbll@insidemath(0)%  \mathdir\the\bodydir
  #1% Once entered in math, set boxes to restore values
  \ifmmode%    \the\everyvbox%    \bodydir\the\bodydir
  \mathdir\the\mathdir
  \everyvbox{\the\everyvbox}%
  \everyvbox{\the\everyvbox}%
  \everyvbox{\the\everyvbox}%
\else%  \fi%  \fi%
Implicitly reverses sectioning labels in bidi=\texttt{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=\texttt{basic}, but there are some additional readjustments for bidi=\texttt{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}
\IfBabelLayout{extras}{\bbl@ncarg\let\bbl@OL@underline{underline }\bbl@carg\bbl@sreplace{underline }{$\@@underline}{\bgroup\bbl@nextfake$\@@underline}$\bbl@carg\bbl@sreplace{underline }{$\m@th$}{\m@th$\egroup$}\let\bbl@OL@LaTeXe\LaTeXe\DeclareRobustCommand{\LaTeXe}{\mbox{$\if b\expandafter\@car\f@series\@nil\boldmath\fi$}\babelsublr{\LaTeX\kern.15em2\bbl@nextfake$_{\textstyle\varepsilon}$}}}
\end{verbatim}

\section{Lua: transforms}

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 discretionary, 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{lang\_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 \texttt{word\_head} points to the starting node of the text to be matched.

\begin{verbatim}
\directlua{
luatexbase.add_to_callback("process_output_buffer", Babel.discard_sublr , "Babel.discard_sublr")
}
\end{verbatim}
function Babel.str_to_nodes(fn, matches, base)
local n, head, last
if fn == nil then return nil end
for s in string.utfvalues(fn(matches)) do
  base = base.replace end
n = node.copy(base)
if not head then
  head = n
else
  last.next = n
end
last = n
return head
end
Babel.fetch_subtext = {}
Babel.ignore_pre_char = function(node)
  return (node.lang == Babel.nohyphenation)
end
-- Merging both functions doesn't seems feasible, because there are too
-- many differences.
Babel.fetch_subtext[0] = function(head)
local word_string = ''
local word_nodes = {}
local lang
local item = head
local inmath = false
while item do
  if item.id == 11 then
    inmath = (item.subtype == 0)
  end
  if inmath then
    -- pass
  elseif item.id == 29 then
    local locale = node.get_attribute(item, Babel.attr_locale)
    if lang == locale or lang == nil then
      lang = lang or locale
      if Babel.ignore_pre_char(item) then
        word_string = word_string .. Babel.us_char
      else
        word_string = word_string .. unicode.utf8.char(item.char)
      end
      word_nodes[#word_nodes+1] = item
    else
      break
    end
  elseif item.id == 12 and item.subtype == 13 then
    word_string = word_string .. ' '
    word_nodes[#word_nodes+1] = item
  end
  if item.id == 11 then
    inmath = (item.subtype == 0)
  end
end
return word_string, word_nodes
end
-- Ignore leading unrecognized nodes, too.
elseif word_string == '' then
    word_string = word_string .. Babel.us_char
    word_nodes[#word_nodes+1] = item -- Will be ignored
end

item = item.next
end

-- Here and above we remove some trailing chars but not the
-- corresponding nodes. But they aren't accessed.
if word_string:sub(-1) == ' ' then
    word_string = word_string:sub(1,-2)
end
word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
return word_string, word_nodes, item, lang
end

Babel.fetch_subtext[1] = function(head)
    local word_string = ''
    local word_nodes = {}
    local lang
    local item = head
    local inmath = false
    while item do
        if item.id == 11 then
            inmath = (item.subtype == 0)
        end
        if inmath then
            -- pass
        elseif item.id == 29 then
            if item.lang == lang or lang == nil then
                if (item.char ~= 124) and (item.char ~= 61) then -- not =, not |
                    lang = lang or item.lang
                    word_string = word_string .. unicode.utf8.char(item.char)
                    word_nodes[#word_nodes+1] = item
                end
            else
                break
            end
        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
        -- (1) Go to next word if nothing was found, and (2) implicitly
        -- remove leading USs.
        elseif word_string == '' then
            -- pass
        -- This is the responsible for splitting by words.
        elseif (item.id == 12 and item.subtype == 13) then
            break
        else
            break
        end
        end
    end

    return word_string, word_nodes, item, lang
end
word_string = word_string .. Babel.us_char
word_nodes[#word_nodes+1] = item -- Will be ignored
dend

dim = item.next
dend

word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
return word_string, word_nodes, item, lang
dend

define function Babel.pre_hyphenate_replace(head)
Babel.hyphenate_replace(head, 0)
dend

function Babel.post_hyphenate_replace(head)
Babel.hyphenate_replace(head, 1)
dend

Babel.us_char = string.char(31)

function Babel.hyphenate_replace(head, mode)
local u = unicode.utf8
local lbkr = Babel.linebreaking.replacements[mode]
local word_head = head
while true do -- for each subtext block
local w, w_nodes, nw, lang = Babel.fetch_subtext[mode](word_head)
if Babel.debug then
print()
print((mode == 0) and '@@@@<' or '@@@@>', w)
end
if nw == nil and w == '' then break end
if not lang then goto next end
if not lbkr[lang] then goto next end

-- For each saved (pre|post)hyphenation. TODO. Reconsider how
-- loops are nested.
for k=1, #lbkr[lang] do
local p = lbkr[lang][k].pattern
local r = lbkr[lang][k].replace
local attr = lbkr[lang][k].attr or -1
if Babel.debug then
print('*****', p, mode)
end

-- This variable is set in some cases below to the first *byte*
-- after the match, either as found by u.match (faster) or the
-- computed position based on sc if w has changed.
local last_match = 0
local step = 0

-- For every match.
while true do
if Babel.debug then
print('*****', p, mode)
end
while true do
if Babel.debug then
print('++++')
end
local new -- used when inserting and removing nodes
local matches = { u.match(w, p, last_match) }

if #matches < 2 then break end

-- Get and remove empty captures (with ()'s, which return a
-- number with the position), and keep actual captures
-- (from (...)), if any, in matches.
local first = table.remove(matches, 1)
local last = table.remove(matches, #matches)
-- Non re-fetched substrings may contain \31, which separates
-- substrings.
if string.find(w:sub(first, last-1), Babel.us_char) then break end

local save_last = last -- with A()BC()D, points to D

-- Fix offsets, from bytes to unicode. Explained above.
first = u.len(w:sub(1, first-1)) + 1
last = u.len(w:sub(1, last-1)) -- now last points to C

-- This loop stores in a small table the nodes
-- corresponding to the pattern. Used by 'data' to provide a
-- predictable behavior with 'insert' (w_nodes is modified on
-- the fly), and also access to 'remove'd nodes.
local sc = first-1 -- Used below, too
local data_nodes = {}

local enabled = true
for q = 1, last-first+1 do
  data_nodes[q] = w_nodes[sc+q]
  if enabled
    and attr > -1
    and not node.has_attribute(data_nodes[q], attr)
  then
    enabled = false
  end
end

-- This loop traverses the matched substring and takes the
-- corresponding action stored in the replacement list.
-- sc = the position in substr nodes / string
-- rc = the replacement table index
local rc = 0

while rc < last-first+1 do -- for each replacement
  if Babel.debug then
    print('.....', rc + 1)
  end
  sc = sc + 1
  rc = rc + 1

  if Babel.debug then
    Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
    local ss = ''
    for itt in node.traverse(head) do
      if itt.id == 29 then
        ss = ss .. unicode.utf8.char(itt.char)
      else
        ss = ss .. '{' .. itt.id .. '}'
      end
    end
    print('***************', ss)
  end
local crep = r[rc]
local item = w_nodes[sc]
local item_base = item
local placeholder = Babel.us_char
local d

if crep and crep.data then
    item_base = data_nodes[crep.data]
end

if crep then
    step = crep.step or 0
end

if (not enabled) or (crep and next(crep) == nil) then -- = {}
    last_match = save_last -- Optimization
    goto next
else crep == nil or crep.remove then
    node.remove(head, item)
    table.remove(w_nodes, sc)
    w = u.sub(w, 1, sc-1) .. u.sub(w, sc+1)
    sc = sc - 1 -- Nothing has been inserted.
    last_match = utf8.offset(w, sc+1+step)
    goto next
elseif crep and crep.kashida then -- Experimental
    node.set_attribute(item, Babel.attr_kashida, crep.kashida)
    last_match = utf8.offset(w, sc+1+step)
    goto next
elseif crep and crep.string then
    local str = crep.string(matches)
    if str == '' then -- Gather with nil
        node.remove(head, item)
        table.remove(w_nodes, sc)
        w = u.sub(w, 1, sc-1) .. u.sub(w, sc+1)
        sc = sc - 1 -- Nothing has been inserted.
    else
        local loop_first = true
        for s in string.utfvalues(str) do
            d = node.copy(item_base)
            d.char = s
            if loop_first then
                loop_first = false
                head, new = node.insert_before(head, item, d)
                if sc == 1 then
                    word_head = head
                end
                w_nodes[sc] = d
                w = u.sub(w, 1, sc-1) .. u.char(s) .. u.sub(w, sc+1)
            else
                sc = sc + 1
                head, new = node.insert_before(head, item, d)
                table.insert(w_nodes, sc, new)
                w = u.sub(w, 1, sc-1) .. u.char(s) .. u.sub(w, sc)
            end
            if Babel.debug then
                end
print('.....', 'str')
Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
end
end -- for
node.remove(head, item)
end -- if ''
last_match = utf8.offset(w, sc+1+step)
goto next

elseif mode == 1 and crep and (crep.pre or crep.no or crep.post) then
d = node.new(7, 3) -- (disc, regular)
d.pre = Babel.str_to_nodes(crep.pre, matches, item_base)
d.post = Babel.str_to_nodes(crep.post, matches, item_base)
d.replace = Babel.str_to_nodes(crep.no, matches, item_base)
d.attr = item_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
placeholder = '|'  
head, new = node.insert_before(head, item, d)

elseif mode == 0 and crep and (crep.pre or crep.no or crep.post) then
  -- ERROR

elseif crep and crep.penalty then
d = node.new(14, 0) -- (penalty, userpenalty)
d.attr = item_base.attr
d.penalty = crep.penalty
head, new = node.insert_before(head, item, d)

elseif crep and crep.space then
  -- 655360 = 10 pt = 10 * 65536 sp
  d = node.new(12, 13) -- (glue, spaceskip)
  local quad = font.getfont(item_base.font).size or 655360
  if mode == 0 then
    placeholder = ' '
  end
  head, new = node.insert_before(head, item, d)

elseif crep and crep.spacefactor then
d = node.new(12, 13) -- (glue, spaceskip)
local base_font = font.getfont(item_base.font)
if mode == 0 then
  placeholder = ' '
end
head, new = node.insert_before(head, item, d)

elseif mode == 0 and crep and crep.space then
  -- ERROR

end -- ie replacement cases

-- Shared by disc, space and penalty.
if sc == 1 then
word_head = head
if crep.insert then
  w = u.sub(w, 1, sc-1) .. placeholder .. u.sub(w, sc)
  table.insert(w_nodes, sc, new)
  last = last + 1
else
  w_nodes[sc] = d
  node.remove(head, item)
  w = u.sub(w, 1, sc-1) .. placeholder .. u.sub(w, sc+1)
end
last_match = utf8.offset(w, sc+1+step)
::next::
end -- for each replacement
if Babel.debug then
  print('.....', '/')
  Babel.debug_hyph(w, w_nodes, sc, first, last, last_match)
end
end -- for match
end -- for patterns
::next::
word_head = nw
end -- for substring
return head
-- This table stores capture maps, numbered consecutively
Babel.capture_maps = {}

function Babel.capture_func(key, cap)
  local ret = "[* .. cap:gsub('{{[0-9]}}', "]%m[%1]%"] .. "]"
  local cnt
  local u = unicode.utf8
  ret, cnt = ret:gsub('{([0-9]|[^|]+)|(-.)}', Babel.capture_func_map)
  if cnt == 0 then
    ret = u.gsub(ret, '{(%x%x%x%x+)}', function (n)
      return u.char(tonumber(n, 16))
    end)
  end
  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

function Babel.capture_func_map(capno, from, to)
  local u = unicode.utf8
  from = u.gsub(from, '{{%x%x%x%x+}}', function (n)
    return u.char(tonumber(n, 16))
  end)
  return u.char(tonumber(n, 16))
7383   end)
7384   to = u.gsub(to, '{(%x%x%x%x%+)}','
7385       function (n)
7386           return u.char(tonumber(n, 16))
7387       end)
7388   local froms = {}
7389   for s in string.utfcharacters(from) do
7390       table.insert(froms, s)
7391   end
7392   local cnt = 1
7393   table.insert(Babel.capture_maps, {})
7394   local mlen = table.getn(Babel.capture_maps)
7395   for s in string.utfcharacters(to) do
7396       Babel.capture_maps[mlen][froms[cnt]] = s
7397       cnt = cnt + 1
7398   end
7399   return "]}.Babel.capt_map(m[" .. capno .. "]," ..
7400       (mlen) .. ").").. "["
7401   end
7402   -- Create/Extend reversed sorted list of kashida weights:
7403   function Babel.capture_kashida(key, wt)
7404       wt = tonumber(wt)
7405       if Babel.kashida_wts then
7406           for p, q in ipairs(Babel.kashida_wts) do
7407               if wt == q then
7408                   break
7409               elseif wt > q then
7410                   table.insert(Babel.kashida_wts, p, wt)
7411                   break
7412               elseif table.getn(Babel.kashida_wts) == p then
7413                   table.insert(Babel.kashida_wts, wt)
7414               end
7415           end
7416       end
7417       else
7418           Babel.kashida_wts = { wt }
7419       end
7420       return 'kashida = ' .. wt
7421   end
7422   -- Experimental: applies prehyphenation transforms to a string (letters
7423       and spaces).
7424   function Babel.string_prehyphenation(str, locale)
7425       local n, head, last, res
7426       head = node.new(8, 0) -- dummy (hack just to start)
7427       last = head
7428       for s in string.utfvalues(str) do
7429           if s == 20 then
7430               n = node.new(12, 0)
7431           else
7432               n = node.new(29, 0)
7433               n.char = s
7434           end
7435           node.set_attribute(n, Babel.attr_locale, locale)
7436           last.next = n
7437           last = n
7438       end
7439       head = Babel.hyphenate_replace(head, 0)
7440       res = ''
7441       for n in node.traverse(head) do
7442           if n.id == 12 then
7443               res = res .. ' '
7444           elseif n.id == 29 then
7445               res = res .. '
7446               last = last.next
7447           end
7448       end
7449       return res
10.12 Lua: 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:

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

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 (<d>, <r> or <al>).

From UAX#9: “Where available, markup should be used instead of the explicit formatting characters”. So, this simple version just ignores formatting characters. Actually, most of that annex is devoted to how to handle them.

BD14-BD16 are not implemented. Unicode (and the W3C) are making a great effort to deal with some special problematic cases in “streamed” plain text. I don’t think this is the way to go – particular issues should be fixed by a high level interface taking into account the needs of the document. And here is where luatex excels, because everything related to bidi writing is under our control.
d.dir = '-' .. dir
node.insert_after(head, to, d)
end

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

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

    local strong = ('TRT' == tex.pardir) and 'r' or 'l'
    local strong_lr = (strong == 'l') and 'l' or 'r'
    local outer = strong
    local new_dir = false
    local first_dir = false
    local inmath = false
    local last_lr
    local type_n = ''

    for item in node.traverse(head) do
      -- three cases: glyph, dir, otherwise
      if item.id == node.id'glyph'
        or (item.id == 7 and item.subtype == 2) then
        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'
        if inmath then dir = ('TRT' == tex.mathdir) and 'r' or 'l' end
      end
      attr_dir = 0
      for at in node.traverse(item.attr) do
        if at.number == Babel.attr_dir then
          attr_dir = at.value & 0x3
        end
      end
    end
  end
if attr_dir == 1 then
    strong = 'r'
elseif attr_dir == 2 then
    strong = 'al'
else
    strong = 'l'
end

strong_lr = (strong == 'l') and 'l' or 'r'
outer = strong_lr

new_dir = false
end

if dir == 'nsm' then dir = strong end -- W1

Numbers. The dual <al>/<r> system for R is somewhat cumbersome.

dir_real = dir -- We need dir_real to set strong below
if dir == 'al' then dir = 'r' end -- W3

By W2, there are no <en> <et> <es> if strong == <al>, only <an>. Therefore, there are not <et en> nor <en et>, W5 can be ignored, and W6 applied:

if strong == 'al' then
    if dir == 'en' then dir = 'an' end -- W2
    if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
    strong_lr = 'r' -- W3
end

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

elseif item.id == node.id'dir' and not inmath then
    new_dir = true
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 <en>, <et>, <an>, <es> and <cs> is typeset (with some rules) in L mode. We store the starting and ending points, and only when anything different is found (including nil, i.e., a non-char), the textdir is set. This means you cannot insert, say, a whatsit, but this is what I would expect (with luacolor you may colorize some digits). Anyway, this behavior could be changed with a switch in the future. Note in the first branch only <an> is relevant if <al>.

if dir == 'en' or dir == 'an' or dir == 'et' then
    if dir ~= 'et' then
        type_n = dir
    end
    first_n = first_n or item
    last_n = last_es or item
    last_es = nil
elseif dir == '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')
        else
        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
            else
                strong_lr == 'l' and type_n ~= '' then
            last_d = last_n
        end
        type_n = ''
        first_n, last_n = nil, nil
    end

end
R text in L, or L text in R. Order of dir* mark's are relevant: d goes outside n, and therefore it's emitted after. See dir* mark to understand why (but is the nesting actually necessary or is a flat dir structure enough?). Only L, R (and AL) chars are taken into account – everything else, including spaces, whatstis, 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
    dir_mark(head, first_d, last_d, outer)
    first_d, last_d = nil, nil
  end
end
```

**Mirroring.** Each chunk of text in a certain language is considered a “closed” sequence. If <r on r> and <l on l>, it's clearly <r> and <l>, respity, but with other combinations depends on outer. From all these, we select only those resolving <on> → <r>. At the beginning (when last_lr is nil) of an R text, they are mirrored directly.

**TODO** -numbers in R mode are processed. It doesn't hurt, but should not be done.

```plaintext
if dir and not last_lr and dir ~= 'l' and outer == 'r' then
  item.char = characters[item.char] 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
```

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).

```plaintext
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
```

Mirror the last chars if they are no directed. And make sure any open block is closed, too.

```plaintext
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
```

In boxes, the dir node could be added before the original head, so the actual head is the previous node.

```plaintext
return node.prev(head) or head
```
And here the Lua code for bidi=basic:

```lua
Babel = Babel or {}

-- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
Babel.fontmap = Babel.fontmap or {}

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

-- To cancel mirroring. Also OML, OMS, U?
Babel.symbol_fonts = Babel.symbol_fonts or {}
Babel.symbol_fonts[font.id('tenln')] = true
Babel.symbol_fonts[font.id('tenlmw')] = true
Babel.symbol_fonts[font.id('tencirc')] = true
Babel.symbol_fonts[font.id('tencircw')] = true

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)
        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
        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
```

152
local function glyph_not_symbol_font(node)
    if node.id == GLYPH then
        return not Babel.symbol_fonts[node.font]
    else
        return false
    end
end

-- TODO - \hbox with an explicit dir can lead to wrong results
-- \<R \hbox dir TLT\(<R>\)> and \<L \hbox dir TRT\(<L>\)>. A small attempt
-- was made to improve the situation, but the problem is the 3-dir
-- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
-- well.

function Babel.bidi(head, ispar, hdir)
    local d = -- d is used mainly for computations in a loop
    local prev_d = ''
    local new_d = false
    local nodes = {}
    local outer_first = nil
    local inmath = false
    local glue_d = nil
    local glue_i = nil
    local has_en = false
    local first_et = nil
    local has_hyperlink = false
    local ATDIR = Babel.attr_dir
    local save_outer
    local temp = node.get_attribute(head, ATDIR)
    if temp then
        temp = temp & 0x3
        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
    local outer = save_outer
    local last = outer
    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 glyph_not_symbol_font(item)
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]
if not d or d == 'nsm' then
    for nn, et in ipairs(ranges) do
        if item_r.char < et[1] then
            break
        elseif item_r.char <= et[2] then
            if not d then d = et[3]
            elseif d == 'nsm' then d_font = et[3]
            end
            break
        end
    end
end
end

if new_d then
    table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
end

-- A short 'pause' in bidi for mapfont
d_font = d_font or d
_d_font = (d_font == 'l' and 0) or
    (d_font == 'nsm' and 0) or
    (d_font == 'r' and 1) or
    (d_font == 'al' and 2) or
    (d_font == 'an' and 2) or nil
if d_font and fontmap and fontmap[d_font][item_r.font] then
    item_r.font = fontmap[d_font][item_r.font]
end

if glue_d then
    if inmath then
        attr_d = 0
    else
        attr_d = node.get_attribute(item, ATDIR)
        attr_d = attr_d & 0x3
    end
    if attr_d == 1 then
        outer_first = 'r'
        last = 'r'
    elseif attr_d == 2 then
        outer_first = 'r'
        last = 'al'
    else
        outer_first = 'l'
        last = 'l'
    end
    outer = last
    has_en = false
    first_et = nil
    new_d = false
end
if glue_d then
if (d == 'l' and 'l' or 'r') ~= glue_d then
    table.insert(nodes, {glue_i, 'on', nil})
end

if head ~= item then new_d = true end

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)

elseif item.id == 8 and item.subtype == 19 then
    has_hyperlink = true

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)
elsefirst et then -- d may be nil here!
    if has_en then
        if last == 'l' then
            temp = 'l' -- W7
        else
            temp = 'en' -- W5
        end
    end
end
else
    temp = 'on' -- W6
end
for e = first_et, #nodes do
    if glyph_not_symbol_font(nodes[e][1]) then nodes[e][2] = temp end
end
first_et = nil
has_en = false
end

-- Force mathdir in math if ON (currently works as expected only
-- with 'l')
if inmath and d == 'on' then
    d = ('TRT' == tex.mathdir) and 'r' or 'l'
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
end
for e = first_et, #nodes do
    if glyph_not_symbol_font(nodes[e][1]) then nodes[e][2] = temp 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
end
last = outer_first or last

local d = nodes[q][2]
if d == 'an' or d == 'en' then d = 'r' end
if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
if d == 'on' then
  first_on = first_on or q
elseif first_on then
  if last == d then
    temp = d
  else
    temp = outer
  end
  for r = first_on, q - 1 do
    nodes[r][2] = temp
    item = nodes[r][1] -- MIRRORING
    if Babel.mirroring_enabled and glyph_not_symbol_font(item)
      and temp == 'r' and characters[item.char] then
      local font_mode = ''
      if item.font > 0 and font.fonts[item.font].properties then
        font_mode = font.fonts[item.font].properties.mode
      end
      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

-------- IMPLICIT, REORDER --------

outer = save_outer
last = outer

local state = {}
state.has_r = false
for q = 1, #nodes do
  local item = nodes[q][1]
  outer = nodes[q][3] or outer
  local d = nodes[q][2]
  if d == 'nsm' then d = last end -- W1
  if d == 'en' then d = 'an' end
  local isdir = (d == 'r' or d == 'l')
  if outer == 'l' and d == 'an' then
    state.san = state.san or item
    state.ean = item
  elseif state.san then
    head, state = insert_numeric(head, state)
  end
  if outer == 'l' then
    if d == 'an' or d == 'r' then -- im -> implicit

157
if d == 'r' then state.has_r = true end
state.sim = state.sim or item
state.eim = item
elseif d == 'l' and state.sim and state.has_r then
    head, state = insert_implicit(head, state, outer)
elseif d == 'l' then
    state.sim, state.eim, state.has_r = nil, nil, false
end
else
    if d == 'an' or d == 'l' then
        if nodes[q][3] then -- nil except after an explicit dir
            state.sim = item -- so we move sim 'inside' the group
        else
            state.sim = state.sim or item
        end
        state.eim = item
    elseif d == 'r' and state.sim then
        head, state = insert_implicit(head, state, outer)
    elseif d == 'r' then
        state.sim, state.eim = nil, nil
    end
end

if isdir then
    last = d -- Don't search back - best save now
else d == 'on' and state.san then
    state.san = state.san or item
    state.ean = item
end
end
end
head = node.prev(head) or head

---------------- FIX HYPERLINKS ----------------
if has_hyperlink then
    local flag, linking = 0, 0
    for item in node.traverse(head) do
        if item.id == DIR then
            if item.dir == '+TRT' or item.dir == '+TLT' then
                flag = flag + 1
            elseif item.dir == '-TRT' or item.dir == '-TLT' then
                flag = flag - 1
            end
        elseif item.id == 8 and item.subtype == 19 then
            linking = flag
        elseif item.id == 8 and item.subtype == 20 then
            if linking > 0 then
                if item.prev.id == DIR and
                    (item.prev.dir == '-TRT' or item.prev.dir == '-TLT') then
                    d = node.new(DIR)
                    d.dir = item.prev.dir
                    node.remove(head, item.prev)
                    node.insert_after(head, item, d)
                end
                linking = 0
            end
        end
    end
return head
11 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.

12 The ‘nil’ language

This ‘language’ does nothing, except setting the hyphenation patterns to nohyphenation. For this language currently no special definitions are needed or available. The macro \LaTeXInit 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]] v[[version]]] Nil language
\LaTeXInit{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.

\begin{verbatim}
\ifx\l@nil\@undefined
\newlanguage\l@nil
@namedef{bbl@hyphendata@\the\l@nil}{{}{}}% Remove warning
\let\bbl@elt\relax
\edef\bbl@languages{% Add it to the list of languages
\bbl@languages\bbl@elt{nil}{\the\l@nil}{}{}}
\fi
\end{verbatim}

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

\begin{verbatim}
\providehyphenmins{\CurrentOption}{\m@ne\m@ne}
\end{verbatim}

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

\begin{verbatim}
\captionnil
\datenil
\let\captionsnil\@empty
\let\datenil\@empty
\end{verbatim}

There is no locale file for this pseudo-language, so the corresponding fields are defined here.

\begin{verbatim}
\def\bbl@inidata@nil{%
\bbl@elt{identification}{tag.ini}{und}%
\bbl@elt{identification}{load.level}{0}%
\bbl@elt{identification}{charset}{utf8}%
\bbl@elt{identification}{version}{1.0}%
\bbl@elt{identification}{date}{2022-05-16}%
\bbl@elt{identification}{name.local}{nil}%
\bbl@elt{identification}{name.english}{nil}%
\bbl@elt{identification}{name.babel}{nil}%
\bbl@elt{identification}{tag.bcp47}{und}%
\bbl@elt{identification}{language.tag.bcp47}{und}%
\bbl@elt{identification}{tag.opentype}{dflt}%
\bbl@elt{identification}{script.name}{Latin}%
\bbl@elt{identification}{script.tag.bcp47}{Latin}%
\end{verbatim}
The macro \texttt{\ldf@finish} takes care of looking for a configuration file, setting the main language to be switched on at \texttt{\begin{document}} and resetting the category code of @ to its original value.

13 Calendars

The code for specific calendars are placed in the specific files, loaded when requested by an ini file in the identification section with require.calendars.

Start with function to compute the Julian day. It's based on the little library calendar.js, by John Walker, in the public domain.

\begin{verbatim}
\def\bbl@fpmod#1#2{(#1-#2*floor(#1/#2))}
\def\bbl@cs@gregleap#1{\((\bbl@fpmod{#1}{4} == 0) && (!((\bbl@fpmod{#1}{100} == 0) && (\bbl@fpmod{#1}{400} != 0)))}\)
\def\bbl@cs@jd#1#2#3{\fp_eval:n{ 1721424.5 + (365 * (#1 - 1)) + floor((#1 - 1) / 4) + (-floor((#1 - 1) / 100)) + floor(((367 * #2) - 362) / 12) + ((#2 <= 2) ? 0 : (\bbl@cs@gregleap{#1} ? -1 : -2)) + #3} }
\end{verbatim}

13.1 Islamic

The code for the Civil calendar is based on it, too.

\begin{verbatim}
\ExplSyntaxOn
\def\bbl@cs@isltojd#1#2#3{ (% year, month, day
\begin{verbatim}
\edef\tempa{\fp_eval:n{ floor(\bbl@cs@jd{#2}{#3}{#4})+0.5 #1}}
\edef#5{\fp_eval:n{ floor(\bbl@cs@jd{#2}{#3}{#4})+0.5 #1}}
\end{verbatim}
\end{verbatim}

The Civil calendar.

\begin{verbatim}
\def\bbl@cs@isltojd#1#2#3{ (% year, month, day
\begin{verbatim}
\edef\tempa{\fp_eval:n{ floor(\bbl@cs@jd{#2}{#3}{#4})+0.5 #1}}
\edef#5{\fp_eval:n{ floor(\bbl@cs@jd{#2}{#3}{#4})+0.5 #1}}
\end{verbatim}
\end{verbatim}
The Umm al-Qura calendar, used mainly in Saudi Arabia, is based on moment-hijri, by Abdullah Alsigar (license MIT).

Since the main aim is to provide a suitable \today, and maybe some closed dates, data just covers Hijri \texttt{∼}1435/\texttt{∼}1460 (Gregorian \texttt{∼}2014/\texttt{∼}2038).

\begin{verbatim}
\ExplSyntaxOff
161
\end{verbatim}
This is basically the set of macros written by Michail Rozman in 1991, with corrections and adoptions by Rama Porrat, Misha, Dan Haran and Boris Lavva. This must be eventually replaced by computations with l3fp. An explanation of what's going on can be found in hebcal.sty

\def\bbl@remainder#1#2#3{\relax#3=#1\relax\divide #3 by #2\relax\multiply #3 by -#2\relax\advance #3 by #1\relax}
\newif\ifbbl@divisible
\def\bbl@checkifdivisible#1#2{\countdef\tmp=0\bbl@remainder{#1}{#2}{\tmp}\ifnum \tmp=0\global\bbl@divisibletrue\else\global\bbl@divisiblefalse\fi}
\def\bbl@gregleap{\bbl@checkifdivisible{#1}{4}\ifbbl@divisible\bbl@checkifdivisible{#1}{100}\ifbbl@divisible\bbl@gregleaptrue\else\bbl@gregleapfalse\fi\else\bbl@gregleaptrue\fi\else\bbl@gregleapfalse\fi
\def\bbl@gregdayspriormonths#1#2#3{#3=\ifcase #1 0 \or 0 \or 31 \or 31 \or 59 \or 90 \or 120 \or 151 \or 181 \or 212 \or 243 \or 273 \or 304 \or 334 \fi\bbl@ifgregleap{#2}\ifnum #1 > 2\advance #3 by 1\fi\fi\bbl@gregdaysprioryears#1#2{#3=\ifcase #1 0 \or 0 \or 31 \or 31 \or 59 \or 90 \or 120 \or 151 \or 181 \or 212 \or 243 \or 273 \or 304 \or 334 \fi\bbl@ifgregleap{#2}\ifnum #1 > 2\advance #3 by 1\fi\fi\global\bbl@cntcommon=#3}
\multiply \tmpc by 365
#2=\tmpc
\divide \tmpc by 4
\advance #2 by \tmpc
\tmpc=\tmpb
\divide \tmpc by 100
\divide \tmpc by 400
\global\bbl@cntcommon=#2\relax}
#2=\bbl@cntcommon}
\def\bbl@absfromgreg#1#2#3#4{%
{\countdef\tmpd=0
#4=#1\relax
\bbl@gregdayspriormonths{#2}{#3}{\tmpd}%
\advance #4 by \tmpd
\bbl@gregdaysprioryears{#3}{\tmpd}%
\advance #4 by \tmpd
\global\bbl@cntcommon=#4\relax}%
#4=\bbl@cntcommon}
\newif\ifbbl@hebrleap
\def\bbl@checkleaphebryear#1{%
{\countdef\tmpa=0
\countdef\tmpb=1
\tmpa=#1\relax
\multiply \tmpa by 7
\advance \tmpa by 1
\bbl@remainder{\tmpa}{19}{\tmpb}%
\ifnum \tmpb < 7
\global\bbl@hebrleaptrue
\else
\global\bbl@hebrleapfalse
\fi}%
\def\bbl@hebrelishapsedmonths#1#2{%
{\countdef\tmpa=0
\countdef\tmpb=1
\countdef\tmpc=2
\bbl@hebrelishapsedmonths{#1}{#2}%
\tmpa=#2\relax
\multiply \tmpa by 13753
\bbl@remainder{\tmpa}{25920}{\tmpc}%
\tmpc=ConjunctionParts
\global\bbl@cntcommon=#2}%
#2=\bbl@cntcommon}
\def\bbl@hebrelishapseddays#1#2{%
{\countdef\tmpa=0
\countdef\tmpb=1
\countdef\tmpc=2
\bbl@hebrelishapsedmonths(#1)\tmpb}%
\tmpa=years\%19-years this cycle
\tmpc=\tmpb
\multiply \tmpc by 12
\advance \tmpc by 2
\tmpc=\tmpb
\multiply \tmpc by 7
\advance \tmpc by 1
\divide \tmpc by 19
\bbl@remainder{\tmpa}{19}{\tmpb}%
\tmpa=years\%19-years this cycle
\tmpc=\tmpb
\multiply \tmpc by 235
\bbl@remainder{\tmpa}{19}{\tmpb}%
\tmpa=years\%19-years this cycle
\tmpc=\tmpb
\multiply \tmpc by 12
\advance \tmpc by 2
\tmpc=\tmpb
\multiply \tmpc by 7
\advance \tmpc by 1
\divide \tmpc by 19
\bbl@remainder{\tmpa}{19}{\tmpb}%
\tmpa=years\%19-years this cycle
\tmpc=\tmpb
\multiply \tmpc by 13753
\bbl@remainder{\tmpa}{25920}{\tmpc}%
\tmpc=ConjunctionParts
\divide \tmpa by 25920
\multiply #2 by 29
\advance #2 by 1
\advance #2 by \tmpa
\bbl@remainder{#2}{7}{\tmpa}%
\ifnum \tmpc < 19440
\ifnum \tmpc < 9924
\else
\ifnum \tmpa=2
\bbl@checkleaphebryear{#1}% of a common year
\ifbbl@hebrleap
\else
\advance #2 by 1
\fi
\fi
\fi
\ifnum \tmpc < 16789
\else
\ifnum \tmpa=1
\advance #1 by -1
\bbl@checkleaphebryear{#1}% at the end of leap year
\ifbbl@hebrleap
\advance #2 by 1
\fi
\fi
\else
\advance #2 by 1
\fi
\bbl@remainder{#2}{7}{\tmpa}%
\ifnum \tmpa=0
\advance #2 by 1
\else
\ifnum \tmpa=3
\advance #2 by 1
\fi
\fi
\global\bbl@cntcommon=#2\relax%
#2=\bbl@cntcommon}
def\bbl@daysinhebryear#1#2{%
#2=\bbl@hebrelapseddays(#1){\tmpe}%
\advance #1 by 1
\bbl@hebrelapseddays(#1){\tmpe}%
\advance #2 by -\tmpe
\global\bbl@cntcommon=#2%
#2=\bbl@cntcommon}
def\bbl@hebrdayspriormonths#1#2#3{%
#3=\ifcase #1\relax
0 \or
0 \or
30 \or
59 \or
89 \or
118 \or
148 \or
148 \or
177 \or
164
13.3 Persian

There is an algorithm written in TeX by Jabri, Abolhassani, Pournader and Esfahbod, created for the first versions of the Farsi\TeX{} system (no longer available), but the original license is GPL, so its use with LPPL is problematic. The code here follows loosely that by John Walker, which is free and accurate, but sadly very complex, so the relevant data for the years 2013-2050 have been pre-calculated and stored. Actually, all we need is the first day (either March 20 or March 21).

13.4 Coptic and Ethiopic

Adapted from\{jquery.calendars.package-1.1.4\}, written by Keith Wood, 2010. Dual license: GPL and MIT. The only difference is the epoch.
\edef\bbl@tempd{\fp_eval:n{\bbl@cs@jd{#1}{#2}{#3} - 2457072.5}}
\count@\z@\@tempcnta=2015
\bbl@foreach\bbl@cs@chinese@data{\ifnum##1>\bbl@tempd\else
\advance\count@\@ne\ifnum\count@>12\count@\@ne\advance\@tempcnta\@ne\fi
\bbl@xin{,##1,}{,\bbl@cs@chinese@leap,}\ifin@
\advance\count@\m@ne\edef\bbl@tempe{\the\numexpr\count@+12\relax}\else
\edef\bbl@tempe{\the\count@}\fi
\edef\bbl@tempb{##1}\fi}
13.5 Buddhist

That's very simple.
\edef\bbl@ca@buddhist#1-#2-#3\@@#4#5#6{\edef#4{\number\numexpr#1+543\relax} \edef#5{#2} \edef#6{#3}}
\subsection{Chinese}

Brute force, with the Julian day of first day of each month. The table has been computed with the help of `\textsf{python-lunardate}` by Ricky Yeung, GPLv2 (but the code itself has not been used). The range is 2015-2044.
14 Support for Plain \TeX\ (plain.def)

14.1 Not renaming hyphen.tex

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

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

People can have a file locallyhyphen.tex or whatever they like, but they mustn't diddle with hyphen.tex (or plain.tex except to preload additional fonts).

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

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

\catcode\{=1 % left brace is begin-group character
If a file called `hyphen.cfg` can be found, we make sure that it will be read instead of the file `hyphen.tex`. 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).

```
\openin 0 hyphen.cfg
\ifeof0
\else
\let\input\a
\fi
```

Then \input is defined to forget about its argument and load `hyphen.cfg` instead. Once that's done the original meaning of \input can be restored and the definition of \a can be forgotten.

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

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

```
\let\plain\a plain.tex
\let\lplain\a lplain.tex
```

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

```
\def\fmtname{babel-plain}
\def\fmtname{babel-lplain}
```

When you are using a different format, based on plain.tex you can make a copy of `blplain.tex`, rename it and replace `plain.tex` with the name of your format file.

### 14.2 Emulating some \LaTeX features

The file `babel.def` expects some definitions made in the \LaTeX2ε style file. So, in 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).

```
\def\@empty{}
\def\loadlocalcfg#1{%
\openin0#1.cfg
\ifeof0
\closein0
\else
\closein0
\immediate\write16{*************************************}%
\immediate\write16{* Local config file #1.cfg used}%
\immediate\write16{*}%
\input #1.cfg\relax%
\fi
\@endofldf}
```

### 14.3 General tools

A number of \BibTeX macro's that are needed later on.

```
\long\def@firstofone#1{#1}
\long\def@firstoftwo#1#2{#1}
\long\def@secondoftwo#1#2{#2}
```
\LaTeX\ has the command \texttt{\@onlypreamble} which adds commands to a list of commands that are no longer needed after \texttt{\begin{document}}.

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

\begin{verbatim}
\def\begindocument{%
  \@begindocumenthook
  \global\let\@begindocumenthook\@undefined
  \def\do##1\{\global\let##1\@undefined\}
  \@preamblecmds
  \global\let\do\noexpand}
\ifx\@begindocumenthook\@undefined
  \def\@begindocumenthook{}
\fi
\@onlypreamble\@begindocumenthook
\def\AtBeginDocument{\g@addto@macro\@begindocumenthook}
\end{verbatim}

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

\begin{verbatim}
\def\AtEndOfPackage#1{\g@addto@macro\@endofldf{#1}}
\end{verbatim}

\let\bbl@afterlang\@empty
\chardef\bbl@opt@hyphenmap\z@

\LaTeX needs to be able to switch off writing to its auxiliary files; plain doesn't have them by default. There is a trick to hide some conditional commands from the outer \texttt{\ifx}. The same trick is applied below.

\begin{verbatim}
\catcode\&=\z@
\ifx&if@filesw\@undefined
  \expandafter\let\csname if@filesw\expandafter\endcsname\csname iffalse\endcsname
\fi
\catcode\&=4
\end{verbatim}

\LaTeX commands to define control sequences.

\begin{verbatim}
\def\newcommand{\@star@or@long\new@command}
\def\new@command#1{\@testopt{\@newcommand#1}0}
\def\@newcommand#1[#2]{\@ifnextchar [{{\@xargdef#1[#2]}}{\@argdef#1[#2]}}
\long\def\@argdef#1[#2]#3{\@yargdef#1\@ne{#2}{#3}}
\long\def\@xargdef#1[#2][#3]#4{\expandafter\def\expandafter#1\expandafter{\expandafter\@protected@testopt\expandafter #1\@string#1\expandafter{#3}}\expandafter\@yargdef \csname\string#1\endcsname\tw@{#2}{#4}}
\long\def\@yargdef#1#2#3{\@tempcnta#3\relax\advance\@tempcnta\@ne\let\@hash@\relax\edef\reserved@a{\ifx#2\tw@ \@hash@1\@hash@\fi}\@tempcntb\@whilenum\@tempcntb<\@tempcnta\do{\edef\reserved@a{\reserved@a\@hash@\the\@tempcntb}}\let\@hash@\reserved@a\@whilenum\tempcntb<\tempcnta\do{\edef\reserved@a{\reserved@a\@hash@\the\tempcntb}}\advance\tempcntb\@ne\let\@hash@\reserved@a\@whilenum\tempcntb<\tempcnta\do{\edef\reserved@a{\reserved@a\@hash@\the\tempcntb}}\advance\tempcntb\@ne\let\@hash@\reserved@a
\end{verbatim}

\providecommand{\@star@or@long\provide@command}
\def\provide@command#1{\begingroup\escapechar\m@ne\xdef\@gtempa{{\string#1}}\endgroup}
\end{verbatim}
The following little macro is taken from latex.ltx; it checks whether its first argument is part of its second argument. It uses the boolean \in@; allocating a new boolean inside conditionally executed code is not possible, hence the construct with the temporary definition of \bbl@tempa.

\def\bbl@tempa{\csname newif\endcsname\&ifin@}
\catcode\&=4
\ifx\in@\@undefined
\def\in@#1#2{\def\in@@##1#1##2##3\in@@{\ifx\in@##2\in@false\else\in@true\fi}\in@@#2#1\in@\in@@}
\else\let\bbl@tempa\@empty\fi
\bbl@tempa

LATEX has a macro to check whether a certain package was loaded with specific options. The command has two extra arguments which are code to be executed in either the true or false case. This is used to detect whether the document needs one of the accents to be activated (activegrave and activeacute). For plain \TeX we assume that the user wants them to be active by default. Therefore the only thing we do is execute the third argument (the code for the true case).

\def\@ifpackagewith#1#2#3#4{#3}

The \LaTeX{} macro \@ifloaded 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.

\def\@ifloaded#1#2#3#4{}

For the following code we need to make sure that the commands \newcommand and \providecommand exist with some sensible definition. They are not fully equivalent to their \LaTeX{} versions; just enough to make things work in plain \TeX{} environments.

\ifx@tempcnta\@undefined
\csname newcounter@endcsname@tempcnta\relax
\fi
To prevent wasting two counters in \TeX{} (because counters with the same name are allocated later by it) we reset the counter that holds the next free counter (~\texttt{\count10}).

\texttt{\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}%\futurelet\@let@token\@ifnch}\def\@ifnch{%\ifx\@let@token\@sptoken\let\reserved@c@xifnch\else\ifx\@let@token\reserved@d\let\reserved@c\reserved@a\else\let\reserved@c\reserved@b\fi\fi\reserved@c}
def\:{\let\@sptoken=} \: % this makes \@sptoken a space token
\def\:{\@xifnch} \expandafter\def\: {\futurelet\@let@token\@ifnch}\fi
\def\@testopt#1#2{\@ifnextchar\[{#1}{#1[#2]}}
def\@protected@testopt#1{\ifx\protect\@typeset@protect\expandafter\@testopt\else\@x@protect#1%\fi}
\long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax#2\relax}\fi}\long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum\else\expandafter\@gobble\fi{#1}}

14.4 Encoding related macros

Code from \texttt{\textbackslash outsoutenc.dtx}, adapted for use in the plain \TeX{} environment.

\def\DeclareTextCommand{%\dec@text@cmd\providecommand\fi}
def\ProvideTextCommand{%\dec@text@cmd\providecommand\fi}
def\DeclareTextSymbol#1#2#3{%\chardef#1{#2}#3\relax\fi}
def\@dec@text@cmd#1#2#3{%\expandafter\def\expandafter#2%\expandafter{\csname#3-cmd\expandafter\endcsname#2%\csname#3\string#2\endcsname}%\let\@ifdefinable\@rc@ifdefinable\expandafter#1\csname#3\string#2\endcsname\fi}
Currently we only use the \texttt{\TeXZapf} 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 \texttt{plain \TeX}.

For a couple of languages we need the \texttt{\TeXZapf}-control sequence \texttt{\scriptsize} to be available. Because plain \TeX doesn't have such a sophisticated font mechanism as \TeXZapf has, we just \texttt{\let it to \sevenrm}.

And a few more “dummy” definitions.
15 Acknowledgements

I would like to thank all who volunteered as \beta-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 Spith 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.

There are also many contributors for specific languages, which are mentioned in the respective files. Without them, babel just wouldn’t exist.

References