

Babel

Version 3.38
2020/01/15

Original author
Johannes L. Braams

Current maintainer
Javier Bezos

Localization and
internationalization

TeX

pdfTeX

LuaTeX

LuaHBTeX

XeTeX

Contents

I	User guide	4
1	The user interface	4
1.1	Monolingual documents	4
1.2	Multilingual documents	6
1.3	Modifiers	7
1.4	Troubleshooting	8
1.5	Plain	8
1.6	Basic language selectors	8
1.7	Auxiliary language selectors	9
1.8	More on selection	10
1.9	Shorthands	11
1.10	Package options	15
1.11	The base option	17
1.12	ini files	17
1.13	Selecting fonts	24
1.14	Modifying a language	26
1.15	Creating a language	27
1.16	Digits	30
1.17	Accessing language info	31
1.18	Hyphenation and line breaking	31
1.19	Selecting scripts	33
1.20	Selecting directions	34
1.21	Language attributes	38
1.22	Hooks	38
1.23	Languages supported by babel with ldf files	39
1.24	Unicode character properties in luatex	40
1.25	Tweaking some features	41
1.26	Tips, workarounds, known issues and notes	41
1.27	Current and future work	42
1.28	Tentative and experimental code	43
2	Loading languages with language.dat	44
2.1	Format	44
3	The interface between the core of babel and the language definition files	45
3.1	Guidelines for contributed languages	46
3.2	Basic macros	46
3.3	Skeleton	48
3.4	Support for active characters	49
3.5	Support for saving macro definitions	49
3.6	Support for extending macros	49
3.7	Macros common to a number of languages	50
3.8	Encoding-dependent strings	50
4	Changes	54
4.1	Changes in babel version 3.9	54
II	Source code	54
5	Identification and loading of required files	54

6	locale directory	55
7	Tools	55
7.1	Multiple languages	59
8	The Package File (\LaTeX, babel.sty)	60
8.1	base	61
8.2	key=value options and other general option	63
8.3	Conditional loading of shorthands	64
8.4	Language options	66
9	The kernel of Babel (babel.def, common)	68
9.1	Tools	68
9.2	Hooks	71
9.3	Setting up language files	73
9.4	Shorthands	75
9.5	Language attributes	84
9.6	Support for saving macro definitions	87
9.7	Short tags	87
9.8	Hyphens	88
9.9	Multiencoding strings	89
9.10	Macros common to a number of languages	95
9.11	Making glyphs available	95
9.11.1	Quotation marks	96
9.11.2	Letters	97
9.11.3	Shorthands for quotation marks	98
9.11.4	Umlauts and tremas	99
9.12	Layout	100
9.13	Load engine specific macros	101
9.14	Creating languages	101
10	Adjusting the Babel behavior	112
11	The kernel of Babel (babel.def for \LaTeXonly)	113
11.1	The redefinition of the style commands	113
11.2	Cross referencing macros	114
11.3	Marks	117
11.4	Preventing clashes with other packages	118
11.4.1	ifthen	118
11.4.2	varioref	119
11.4.3	hhline	119
11.4.4	hyperref	120
11.4.5	fancyhdr	120
11.5	Encoding and fonts	121
11.6	Basic bidi support	122
11.7	Local Language Configuration	125
12	Multiple languages (switch.def)	126
12.1	Selecting the language	127
12.2	Errors	136
13	Loading hyphenation patterns	138
14	Font handling with fontspec	142

15	Hooks for XeTeX and LuaTeX	147
15.1	XeTeX	147
15.2	Layout	149
15.3	LuaTeX	151
15.4	Southeast Asian scripts	157
15.5	CJK line breaking	160
15.6	Automatic fonts and ids switching	161
15.7	Layout	167
15.8	Auto bidi with basic and basic-r	169
16	Data for CJK	180
17	The ‘nil’ language	181
18	Support for Plain TeX (plain.def)	181
18.1	Not renaming hyphen.tex	181
18.2	Emulating some L ^A T _E X features	183
18.3	General tools	183
18.4	Encoding related macros	187
19	Acknowledgements	189

Troubleshooting

Paragraph ended before \UTFviii@three@octets was complete	5
No hyphenation patterns were preloaded for (babel) the language ‘LANG’ into the format	5
You are loading directly a language style	8
Unknown language ‘LANG’	8
Argument of \language@active@arg” has an extra }	12
Package fontspec Warning: ‘Language ‘LANG’ not available for font ‘FONT’ with script ‘SCRIPT’ ‘Default’ language used instead’	26
Package babel Info: The following fonts are not babel standard families	26

Part I

User guide

- This user guide focuses on \LaTeX . There are also some notes on its use with Plain \TeX .
- Changes and new features with relation to version 3.8 are highlighted with **New X.XX**. The most recent features could be still unstable. Please, report any issues you find in <https://github.com/latex3/babel/issues>, which is better than just complaining on an e-mail list or a web forum.
- If you are interested in the \TeX multilingual support, please join the kadingira list on <http://tug.org/mailman/listinfo/kadingira>. You can follow the development of babel in <https://github.com/latex3/babel> (which provides some sample files, too).
- See section 3.1 for contributing a language.
- The first sections describe the traditional way of loading a language (with `ldf` files). The alternative way based on `ini` files, which complements the previous one (it does *not* replace it), is described below.

1 The user interface

1.1 Monolingual documents

In most cases, a single language is required, and then all you need in \LaTeX is to load the package using its standard mechanism for this purpose, namely, passing that language as an optional argument. In addition, you may want to set the font and input encodings. Many languages are compatible with `xetex` and `luatex`. With them you can use babel to localize the documents. When these engines are used, the Latin script is covered by default in current \LaTeX (provided the document encoding is UTF-8), because the font loader is preloaded and the font is switched to `lmroman`. Other scripts require loading `fontspec`. You may want to set the font attributes with `fontspec`, too.

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

PDFTEX

```
\documentclass{article}

\usepackage[T1]{fontenc}
% \usepackage[utf8]{inputenc} % Uncomment if LaTeX < 2018-04-01

\usepackage[french]{babel}

\begin{document}

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

\end{document}
```

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

```

\documentclass{article}

\usepackage[russian]{babel}

\babelfont{rm}{DejaVu Serif}

\begin{document}

Россия, находящаяся на пересечении множества культур, а также
с учётом многонационального характера её населения, – отличается
высокой степенью этнокультурного многообразия и способностью к
межкультурному диалогу.

\end{document}

```

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

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

Or the more explanatory:

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

Make sure you set the encoding actually used by your editor.

Another approach is making the language (french in the example) a global option in order to let other packages detect and use it:

```

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

```

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

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

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

```

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

```

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

1.2 Multilingual documents

In multilingual documents, just use a list of the required languages as package or class options. The last language is considered the main one, activated by default. Sometimes, the main language changes the document layout (eg, spanish and french).

EXAMPLE In \LaTeX , the preamble of the document:

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

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

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

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

Examples of cases where main is useful are the following.

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

```
\PassOptionsToPackage{main=english}{babel}
```

WARNING Languages may be set as global and as package option at the same time, but in such a case you should set explicitly the main language with the package option main:

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

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

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

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

PDFTEX

```
\documentclass{article}

\usepackage[T1]{fontenc}
\usepackage[utf8]{inputenc}

\usepackage[english,french]{babel}

\begin{document}

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

\selectlanguage{english}

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

\end{document}
```

EXAMPLE With xetex and luatex, the following bilingual, single script document in UTF-8 encoding just prints a couple of ‘captions’ and \today in Danish and Vietnamese. No additional packages are required.

LUATEX/XETEX

```
\documentclass{article}

\usepackage[vietnamese,danish]{babel}

\begin{document}

\prefacename{} -- \alsoname{} -- \today

\selectlanguage{vietnamese}

\prefacename{} -- \alsoname{} -- \today

\end{document}
```

1.3 Modifiers

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

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

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

¹No predefined “axis” for modifiers are provided because languages and their scripts have quite different needs.

1.4 Troubleshooting

- Loading directly sty files in L^AT_EX (ie, `\usepackage{⟨language⟩}`) is deprecated and you will get the error:²

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

- Another typical error when using babel is the following:³

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

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

1.5 Plain

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

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

WARNING Not all languages provide a sty file and some of them are not compatible with Plain.⁴

1.6 Basic language selectors

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

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

`\selectlanguage` `{⟨language⟩}`

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

²In old versions the error read “You have used an old interface to call babel”, not very helpful.

³In old versions the error read “You haven’t loaded the language LANG yet”.

⁴Even in the babel kernel there were some macros not compatible with plain. Hopefully these issues have been fixed.

```
\selectlanguage{german}
```

This command can be used as environment, too.

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

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

```
{\selectlanguage{<inner-language>} ...}\selectlanguage{<outer-language>}
```

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

\foreignlanguage `{\language}{\text}`

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

1.7 Auxiliary language selectors

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

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

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

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

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

Same as `\foreignlanguage` but as environment. Spaces after the environment are *not* ignored.

This environment was originally intended for intermixing left-to-right typesetting with right-to-left typesetting in engines not supporting a change in the writing direction inside a

line. However, by default it never complied with the documented behavior and it is just a version as environment of `\foreignlanguage`, except when the option `bidi` is set – in this case, `\foreignlanguage` emits a `\leavevmode`, while `otherlanguage*` does not.

`\begin{hyphenrules}` $\langle language \rangle$... `\end{hyphenrules}`

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

1.8 More on selection

`\babeltags` $\langle tag1 \rangle = \langle language1 \rangle, \langle tag2 \rangle = \langle language2 \rangle, \dots$

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

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

EXAMPLE With

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

you can write

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

and

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

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

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

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

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

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

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

```
\babelensure{polish}
```

By default only the basic captions and `\today` are redefined, but you can add further macros with the key `include` in the optional argument (without commas). Macros not to be modified are listed in `exclude`. You can also enforce a font encoding with `fontenc`.⁵ A couple of examples:

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

They are activated when the language is selected (at the `afterextras` event), and it makes some assumptions which could not be fulfilled in some languages. Note also you should include only macros defined by the language, not global macros (eg, \TeX or `\dag`). With ini files (see below), captions are ensured by default.

1.9 Shorthands

A *shorthand* is a sequence of one or two characters that expands to arbitrary \TeX code. Shorthands can be used for different kinds of things, for example: (1) in some languages shorthands such as "a are defined to be able to hyphenate the word if the encoding is OT1; (2) in some languages shorthands such as ! are used to insert the right amount of white space; (3) several kinds of discretionary and breaks can be inserted easily with "-", "=", etc. The package `inputenc` as well as `xetex` and `luatex` have alleviated entering non-ASCII characters, but minority languages and some kinds of text can still require characters not directly available on the keyboards (and sometimes not even as separated or precomposed Unicode characters). As to the point 2, now `pdfTeX` provides `\knbcode`, and `luatex` can manipulate the glyph list. Tools for point 3 can be still very useful in general. There are three levels of shorthands: *user*, *language*, and *system* (by order of precedence). Version 3.9 introduces the *language user* level on top of the user level, as described below. In most cases, you will use only shorthands provided by languages.

NOTE Note the following:

1. Activated chars used for two-char shorthands cannot be followed by a closing brace } and the spaces following are gobbled. With one-char shorthands (eg, :), they are preserved.
2. If on a certain level (system, language, user) there is a one-char shorthand, two-char ones starting with that char and on the same level are ignored.
3. Since they are active, a shorthand cannot contain the same character in its definition (except if it is deactivated with, eg, `\string`).

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

TROUBLESHOOTING A typical error when using shorthands is the following:

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

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

\shorthandon $\{\langle shorthands-list \rangle\}$
\shorthandoff $\ast\{\langle shorthands-list \rangle\}$

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

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

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

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

If you do not need shorthands, or prefer an alternative approach of your own, you may want to switch them off with the package option `shorthands=off`, as described below.

\useshorthands $\ast\{\langle char \rangle\}$

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

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

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

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

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

New 3.9a An optional argument allows to (re)define language and system shorthands (some languages do not activate shorthands, so you may want to add `\languageshorthands{\langle lang \rangle}` to the corresponding `\extras{\langle lang \rangle}`, as explained below). By default, user shorthands are (re)defined.

User shorthands override language ones, which in turn override system shorthands. Language-dependent user shorthands (new in 3.9) take precedence over “normal” user shorthands.

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

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

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

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

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

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

`\languageshorthands` $\{\langle language \rangle\}$

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

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

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

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

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

`\babelshorthand` $\{\langle shorthand \rangle\}$

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

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

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

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

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

Languages with no shorthands Croatian, English (any variety), Indonesian, Hebrew, Interlingua, Irish, Lower Sorbian, Malaysian, North Sami, Romanian, Scottish, Welsh
Languages with only " as defined shorthand character Albanian, Bulgarian, Danish, Dutch, Finnish, German (old and new orthography, also Austrian), Icelandic, Italian, Norwegian, Polish, Portuguese (also Brazilian), Russian, Serbian (with Latin script), Slovene, Swedish, Ukrainian, Upper Sorbian

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

In addition, the babel core declares ~ as a one-char shorthand which is let, like the standard ~, to a non breaking space.⁸

`\ifbabelshorthand` {<character>}{<true>}{<false>}

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

`\aliasshorthand` {<original>}{<alias>}

The command `\aliasshorthand` can be used to let another character perform the same functions as the default shorthand character. If one prefers for example to use the character / over " in typing Polish texts, this can be achieved by entering `\aliasshorthand{"}{/}`. For the reasons in the warning below, usage of this macro is not recommended.

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

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

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

⁷Thanks to Enrico Gregorio

⁸This declaration serves to nothing, but it is preserved for backward compatibility.

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

1.10 Package options

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

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

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

activegrave Same for ```.

shorthands= $\langle char \rangle \langle char \rangle \dots$ | off

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

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

If `'` is included, `activeacute` is set; if ``` is included, `activegrave` is set. Active characters (like `~`) should be preceded by `\string` (otherwise they will be expanded by \TeX before they are passed to the package and therefore they will not be recognized); however, `t` is provided for the common case of `~` (as well as `c` for not so common case of the comma). With `shorthands=off` no language shorthands are defined. As some languages use this mechanism for tools not available otherwise, a macro `\babelshorthand` is defined, which allows using them; see above.

safe= none | ref | bib

Some \TeX macros are redefined so that using shorthands is safe. With `safe=bib` only `\nocite`, `\bibcite` and `\bibitem` are redefined. With `safe=ref` only `\newlabel`, `\ref` and `\pageref` are redefined (as well as a few macros from `varioref` and `ifthen`).

With `safe=none` no macro is redefined. This option is strongly recommended, because a good deal of incompatibilities and errors are related to these redefinitions. As of

New 3.34, in $\epsilon\TeX$ based engines (ie, almost every engine except the oldest ones) shorthands can be used in these macros (formerly you could not).

math= active | normal

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

config= $\langle file \rangle$

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

main=	<code><language></code> Sets the main language, as explained above, ie, this language is always loaded last. If it is not given as package or global option, it is added to the list of requested languages.
headfoot=	<code><language></code> By default, headlines and footlines are not touched (only marks), and if they contain language-dependent macros (which is not usual) there may be unexpected results. With this option you may set the language in heads and foots.
noconfigs	Global and language default config files are not loaded, so you can make sure your document is not spoilt by an unexpected .cfg file. However, if the key config is set, this file is loaded.
showlanguages	Prints to the log the list of languages loaded when the format was created: number (remember dialects can share it), name, hyphenation file and exceptions file.
nocase	New 3.9l Language settings for uppercase and lowercase mapping (as set by <code>\SetCase</code>) are ignored. Use only if there are incompatibilities with other packages.
silent	New 3.9l No warnings and no <i>infos</i> are written to the log file. ⁹
strings=	<code>generic unicode encoded <label> </code> Selects the encoding of strings in languages supporting this feature. Predefined labels are <code>generic</code> (for traditional \TeX , LICR and ASCII strings), <code>unicode</code> (for engines like xetex and luatex) and <code>encoded</code> (for special cases requiring mixed encodings). Other allowed values are font encoding codes (T1, T2A, LGR, L7X...), but only in languages supporting them. Be aware with encoded captions are protected, but they work in <code>\MakeUppercase</code> and the like (this feature misuses some internal \LaTeX tools, so use it only as a last resort).
hyphenmap=	<code>off main select other other*</code> New 3.9g Sets the behavior of case mapping for hyphenation, provided the language defines it. ¹⁰ It can take the following values: off deactivates this feature and no case mapping is applied; first sets it at the first switching commands in the current or parent scope (typically, when the aux file is first read and at <code>\begin{document}</code>), but also the first <code>\selectlanguage</code> in the preamble, and it's the default if a single language option has been stated; ¹¹ select sets it only at <code>\selectlanguage</code> ; other also sets it at <code>otherlanguage</code> ; other* also sets it at <code>otherlanguage*</code> as well as in heads and foots (if the option <code>headfoot</code> is used) and in auxiliary files (ie, at <code>\select@language</code>), and it's the default if several language options have been stated. The option <code>first</code> can be regarded as an optimized version of <code>other*</code> for monolingual documents. ¹²
bidi=	<code>default basic basic-r bidi-l bidi-r</code>

⁹You can use alternatively the package `silence`.

¹⁰Turned off in plain.

¹¹Duplicated options count as several ones.

¹²Providing `foreign` is pointless, because the case mapping applied is that at the end of the paragraph, but if either xetex or luatex change this behavior it might be added. On the other hand, `other` is provided even if I [JBL]

New 3.14 Selects the bidi algorithm to be used in luatex and xetex. See sec. 1.20.

layout=

New 3.16 Selects which layout elements are adapted in bidi documents. See sec. 1.20.

1.11 The base option

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

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

This command is currently the only provided by base. Executes $\langle code \rangle$ when the file loaded by the corresponding package option is finished (at `\ldf@finish`). The setting is global. So

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

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

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

```
\usepackage[base]{babel}
\AfterBabelLanguage{foo}{%
  \let\macroFoo\macro
  \let\macro\relax}
\usepackage[foo,bar]{babel}
```

1.12 ini files

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

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

think it isn't really useful, but who knows.

```

\documentclass{book}

\usepackage{babel}
\babelprovide[import, main]{georgian}

\babelfont{rm}{DejaVu Sans}

\begin{document}

\tableofcontents

\chapter{სამმარეულო და სუფრის ტრადიციები}

ქართული ტრადიციული სამმარეულო ერთ-ერთი უმდიდრესია მთელ მსოფლიოში.

\end{document}

```

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

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

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

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

```
\newfontscript{Devanagari}{deva}
```

Other Indic scripts are still under development in luatex. On the other hand, xetex is better. The upcoming lualatex will be based on lua_Hbtex, so Indic scripts will be rendered correctly with the option `Renderer=Harfbuzz` in `FONTSPEC`.

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

```

\babelprovide[import,hyphenrules=+]{lao}
\babelpatterns[lao]{1ᦺ 1ᦴ 1ᦵ 1ᦶ 1ᦷ 1ᦸ} % Random

```

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

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

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

af	Afrikaans ^{ul}	en-CA	English ^{ul}
agq	Aghem	en-GB	English ^{ul}
ak	Akan	en-NZ	English ^{ul}
am	Amharic ^{ul}	en-US	English ^{ul}
ar	Arabic ^{ul}	eo	Esperanto ^{ul}
ar-DZ	Arabic ^{ul}	es-MX	Spanish ^{ul}
ar-MA	Arabic ^{ul}	es	Spanish ^{ul}
ar-SY	Arabic ^{ul}	et	Estonian ^{ul}
as	Assamese	eu	Basque ^{ul}
asa	Asu	ewo	Ewondo
ast	Asturian ^{ul}	fa	Persian ^{ul}
az-Cyrl	Azerbaijani	ff	Fulah
az-Latn	Azerbaijani	fi	Finnish ^{ul}
az	Azerbaijani ^{ul}	fil	Filipino
bas	Basaa	fo	Faroese
be	Belarusian ^{ul}	fr	French ^{ul}
bem	Bemba	fr-BE	French ^{ul}
bez	Bena	fr-CA	French ^{ul}
bg	Bulgarian ^{ul}	fr-CH	French ^{ul}
bm	Bambara	fr-LU	French ^{ul}
bn	Bangla ^{ul}	fur	Friulian ^{ul}
bo	Tibetan ^u	fy	Western Frisian
brx	Bodo	ga	Irish ^{ul}
bs-Cyrl	Bosnian	gd	Scottish Gaelic ^{ul}
bs-Latn	Bosnian ^{ul}	gl	Galician ^{ul}
bs	Bosnian ^{ul}	gsw	Swiss German
ca	Catalan ^{ul}	gu	Gujarati
ce	Chechen	guz	Gusii
cgg	Chiga	gv	Manx
chr	Cherokee	ha-GH	Hausa
ckb	Central Kurdish	ha-NE	Hausa ^l
cs	Czech ^{ul}	ha	Hausa
cy	Welsh ^{ul}	haw	Hawaiian
da	Danish ^{ul}	he	Hebrew ^{ul}
dav	Taita	hi	Hindi ^u
de-AT	German ^{ul}	hr	Croatian ^{ul}
de-CH	German ^{ul}	hsb	Upper Sorbian ^{ul}
de	German ^{ul}	hu	Hungarian ^{ul}
dje	Zarma	hy	Armenian
dsb	Lower Sorbian ^{ul}	ia	Interlingua ^{ul}
dua	Duala	id	Indonesian ^{ul}
dyo	Jola-Fonyi	ig	Igbo
dz	Dzongkha	ii	Sichuan Yi
ebu	Embu	is	Icelandic ^{ul}
ee	Ewe	it	Italian ^{ul}
el	Greek ^{ul}	ja	Japanese
en-AU	English ^{ul}		

jgo	Ngomba	nd	North Ndebele
jmc	Machame	ne	Nepali
ka	Georgian ^{ul}	nl	Dutch ^{ul}
kab	Kabyle	nmg	Kwasio
kam	Kamba	nn	Norwegian Nynorsk ^{ul}
kde	Makonde	nnh	Ngiemboon
kea	Kabuverdianu	nus	Nuer
khq	Koyra Chiini	nyn	Nyankole
ki	Kikuyu	om	Oromo
kk	Kazakh	or	Odia
kkj	Kako	os	Ossetic
kl	Kalaallisut	pa-Arab	Punjabi
kln	Kalenjin	pa-Guru	Punjabi
km	Khmer	pa	Punjabi
kn	Kannada ^{ul}	pl	Polish ^{ul}
ko	Korean	pms	Piedmontese ^{ul}
kok	Konkani	ps	Pashto
ks	Kashmiri	pt-BR	Portuguese ^{ul}
ksb	Shambala	pt-PT	Portuguese ^{ul}
ksf	Bafia	pt	Portuguese ^{ul}
ksh	Colognian	qu	Quechua
kw	Cornish	rm	Romansh ^{ul}
ky	Kyrgyz	rn	Rundi
lag	Langi	ro	Romanian ^{ul}
lb	Luxembourgish	rof	Rombo
lg	Ganda	ru	Russian ^{ul}
lkt	Lakota	rw	Kinyarwanda
ln	Lingala	rwk	Rwa
lo	Lao ^{ul}	sa-Beng	Sanskrit
lrc	Northern Luri	sa-Deva	Sanskrit
lt	Lithuanian ^{ul}	sa-Gujr	Sanskrit
lu	Luba-Katanga	sa-Knda	Sanskrit
luo	Luo	sa-Mlym	Sanskrit
luy	Luyia	sa-Telu	Sanskrit
lv	Latvian ^{ul}	sa	Sanskrit
mas	Masai	sah	Sakha
mer	Meru	saq	Samburu
mfe	Morisyen	sbp	Sangu
mg	Malagasy	se	Northern Sami ^{ul}
mgf	Makhuwa-Meetto	seh	Sena
mgo	Meta'	ses	Koyraboro Senni
mk	Macedonian ^{ul}	sg	Sango
ml	Malayalam ^{ul}	shi-Latn	Tachelhit
mn	Mongolian	shi-Tfng	Tachelhit
mr	Marathi ^{ul}	shi	Tachelhit
ms-BN	Malay ^l	si	Sinhala
ms-SG	Malay ^l	sk	Slovak ^{ul}
ms	Malay ^{ul}	sl	Slovenian ^{ul}
mt	Maltese	smn	Inari Sami
mua	Mundang	sn	Shona
my	Burmese	so	Somali
mzn	Mazanderani	sq	Albanian ^{ul}
naq	Nama	sr-Cyrl-BA	Serbian ^{ul}
nb	Norwegian Bokmål ^{ul}	sr-Cyrl-ME	Serbian ^{ul}

sr-Cyrl-XK	Serbian ^{ul}	uz-Latn	Uzbek
sr-Cyrl	Serbian ^{ul}	uz	Uzbek
sr-Latn-BA	Serbian ^{ul}	vai-Latn	Vai
sr-Latn-ME	Serbian ^{ul}	vai-Vaii	Vai
sr-Latn-XK	Serbian ^{ul}	vai	Vai
sr-Latn	Serbian ^{ul}	vi	Vietnamese ^{ul}
sr	Serbian ^{ul}	vun	Vunjo
sv	Swedish ^{ul}	wae	Walser
sw	Swahili	xog	Soga
ta	Tamil ^u	yav	Yangben
te	Telugu ^{ul}	yi	Yiddish
teo	Teso	yo	Yoruba
th	Thai ^{ul}	yue	Cantonese
ti	Tigrinya	zgh	Standard Moroccan Tamazight
tk	Turkmen ^{ul}	zh-Hans-HK	Chinese
to	Tongan	zh-Hans-MO	Chinese
tr	Turkish ^{ul}	zh-Hans-SG	Chinese
twq	Tasawaq	zh-Hans	Chinese
tzm	Central Atlas Tamazight	zh-Hant-HK	Chinese
ug	Uyghur	zh-Hant-MO	Chinese
uk	Ukrainian ^{ul}	zh-Hant	Chinese
ur	Urdu ^{ul}	zh	Chinese
uz-Arab	Uzbek	zu	Zulu
uz-Cyrl	Uzbek		

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

aghem	azerbaijani-latn
akan	azerbaijani
albanian	bafia
american	bambara
amharic	basaa
arabic	basque
arabic-algeria	belarusian
arabic-DZ	bemba
arabic-morocco	bena
arabic-MA	bengali
arabic-syria	bodo
arabic-SY	bosnian-cyrillic
armenian	bosnian-cyrl
assamese	bosnian-latin
asturian	bosnian-latn
asu	bosnian
australian	brazilian
austrian	breton
azerbaijani-cyrillic	british
azerbaijani-cyrl	bulgarian
azerbaijani-latin	burmese

canadian	french-ch
cantonese	french-lu
catalan	french-luxembourg
centralatlastamazight	french-switzerland
centralkurdish	french
chechen	friulian
cherokee	fulah
chiga	galician
chinese-hans-hk	ganda
chinese-hans-mo	georgian
chinese-hans-sg	german-at
chinese-hans	german-austria
chinese-hant-hk	german-ch
chinese-hant-mo	german-switzerland
chinese-hant	german
chinese-simplified-hongkongsarchina	greek
chinese-simplified-macausarchina	gujarati
chinese-simplified-singapore	gusii
chinese-simplified	hausa-gh
chinese-traditional-hongkongsarchina	hausa-ghana
chinese-traditional-macausarchina	hausa-ne
chinese-traditional	hausa-niger
chinese	hausa
cognian	hawaiian
cornish	hebrew
croatian	hindi
czech	hungarian
danish	icelandic
duala	igbo
dutch	inarisami
dzongkha	indonesian
embu	interlingua
english-au	irish
english-australia	italian
english-ca	japanese
english-canada	jolafoyi
english-gb	kabuverdianu
english-newzealand	kabyle
english-nz	kako
english-unitedkingdom	kalaallisut
english-unitedstates	kalenjin
english-us	kamba
english	kannada
esperanto	kashmiri
estonian	kazakh
ewe	khmer
ewondo	kikuyu
faroeese	kinyarwanda
filipino	konkani
finnish	korean
french-be	koyraborosenni
french-belgium	koyrachiini
french-ca	kwasio
french-canada	kyrgyz

lakota	persian
langi	piedmontese
lao	polish
latvian	portuguese-br
lingala	portuguese-brazil
lithuanian	portuguese-portugal
lowersorbian	portuguese-pt
lsorbian	portuguese
lubakatanga	punjabi-arab
luo	punjabi-arabic
luxembourgish	punjabi-gurmukhi
luyia	punjabi-guru
macedonian	punjabi
machame	quechua
makhuwameetto	romanian
makonde	romansh
malagasy	rombo
malay-bn	rundi
malay-brunei	russian
malay-sg	rwa
malay-singapore	sakha
malay	samburu
malayalam	samin
maltese	sango
manx	sangu
marathi	sanskrit-beng
masai	sanskrit-bengali
mazanderani	sanskrit-deva
meru	sanskrit-devanagari
meta	sanskrit-gujarati
mexican	sanskrit-gujr
mongolian	sanskrit-kannada
morisyen	sanskrit-knda
mundang	sanskrit-malayalam
nama	sanskrit-mlym
nepali	sanskrit-telu
newzealand	sanskrit-telugu
ngiemboon	sanskrit
ngomba	scottishgaelic
norsk	sena
northernluri	serbian-cyrillic-bosniaherzegovina
northernsami	serbian-cyrillic-kosovo
northndebele	serbian-cyrillic-montenegro
norwegianbokmal	serbian-cyrillic
norwegiannynorsk	serbian-cyrl-ba
nswissgerman	serbian-cyrl-me
nuer	serbian-cyrl-xk
nyankole	serbian-cyrl
nynorsk	serbian-latin-bosniaherzegovina
occitan	serbian-latin-kosovo
oriya	serbian-latin-montenegro
oromo	serbian-latin
ossetic	serbian-latn-ba
pashto	serbian-latn-me

serbian-latn-xk	turkish
serbian-latn	turkmen
serbian	ukenglish
shambala	ukrainian
shona	upporsorbian
sichuanyi	urdu
sinhala	usenglish
slovak	usorbian
slovene	uyghur
slovenian	uzbek-arab
soga	uzbek-arabic
somali	uzbek-cyrillic
spanish-mexico	uzbek-cyrl
spanish-mx	uzbek-latin
spanish	uzbek-latn
standardmoroccantamazight	uzbek
swahili	vai-latin
swedish	vai-latn
swissgerman	vai-vai
tachelhit-latin	vai-vaii
tachelhit-latn	vai
tachelhit-tfng	vietnam
tachelhit-tifinagh	vietnamese
tachelhit	vunjo
taita	walser
tamil	welsh
tasawaq	westernfrisian
telugu	yangben
teso	yiddish
thai	yoruba
tibetan	zarma
tigrinya	zulu afrikaans
tongan	

1.13 Selecting fonts

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

`\babelfont` [*<language-list>*] {*<font-family>*} [*<font-options>*] {*<font-name>*}

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

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

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

¹³See also the package `combofont` for a complementary approach.

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

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

LUATEX/XETEX

```
\documentclass{article}

\usepackage[swedish, bidi=default]{babel}

\babelprovide[import]{hebrew}

\babelfont{rm}{FreeSerif}

\begin{document}

Svenska \foreignlanguage{hebrew}{עֵבְרִית} svenska.

\end{document}
```

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

LUATEX/XETEX

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

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

EXAMPLE Here is how to do it:

LUATEX/XETEX

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

Now, `\kaifamily` and `\kaidefault`, as well as `\textkai` are at your disposal.

NOTE You may load fontspec explicitly. For example:

LUATEX/XETEX

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

This makes sure the OpenType script for Devanagari is `deva` and not `dev2` (luatex does not detect automatically the correct script¹⁴). You may also pass some options to fontspec: with `silent`, the warnings about unavailable scripts or languages are not shown (they are only really useful when the document format is being set up).

¹⁴And even with the correct code some fonts could be rendered incorrectly by fontspec, so double-check the results. xetex fares better, but some fonts are still problematic.

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

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

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

WARNING Do not use `\setxxxxfont` and `\babelfont` at the same time. `\babelfont` follows the standard \TeX conventions to set the basic families – define `\xxdefault`, and activate it with `\xxfamily`. On the other hand, `\setxxxxfont` in `fontspec` takes a different approach, because `\xxfamily` is redefined with the family name hardcoded (so that `\xxdefault` becomes no-op). Of course, both methods are incompatible, and if you use `\setxxxxfont`, font switching with `\babelfont` just does *not* work (nor the standard `\xxdefault`, for that matter). As of [New 3.34](#) there is an attempt to make them compatible, but the language system will not be set by `babel` and should be set with `fontspec` if necessary.

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

TROUBLESHOOTING *Package babel Info: The following fonts are not babel standard families.*

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

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

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

1.14 Modifying a language

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

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

```
\addto\captionenglish{%
\renewcommand\contentsname{Foo}%
}
```

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

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

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

- Macros to be run when a language is selected can be add to `\extras<lang>`:

```
\addto\extrasrussian{\mymacro}
```

There is a counterpart for code to be run when a language is unselected: `\noextras<lang>`.

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

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

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

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

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

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

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

1.15 Creating a language

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

`\babelprovide` [*options*]{*language-name*}

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

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

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

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

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

EXAMPLE If you need a language named arhinish:

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

The main language is not changed (danish in this example). So, you must add

`\selectlanguage{arhinish}` or other selectors where necessary.

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

`import=` *language-tag*

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

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

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

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

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

There are about 200 ini files, with data taken from the ldf files and the CLDR provided by Unicode. Not all languages in the latter are complete, and therefore neither are the ini files. A few languages will show a warning about the current lack of suitability of the date format (hindi, french, breton, and occitan).

Besides `\today`, this option defines an additional command for dates: `\<language>date`, which takes three arguments, namely, year, month and day numbers. In fact, `\today` calls `\<language>today`, which in turn calls `\<language>date{\the\year}{\the\month}{\the\day}`.

captions= *<language-tag>*

Loads only the strings. For example:

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

hyphenrules= *<language-list>*

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

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

If none of the listed hyphenrules exist, the default behavior applies. Note in this example we set `chavacano` as first option – without it, it would select `spanish` even if `chavacano` exists.

A special value is `+`, which allocates a new language (in the \TeX sense). It only makes sense as the last value (or the only one; the subsequent ones are silently ignored). It is mostly useful with `luatex`, because you can add some patterns with `\babelpatterns`, as for example:

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

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

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

script= *<script-name>*

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

language= *<language-name>*

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

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

mapfont= direction

Assigns the font for the writing direction of this language (only with `bidi=basic`). More precisely, what `mapfont=direction` means is, ‘when a character has the same direction as the script for the “provided” language, then change its font to that set for this language’. There are 3 directions, following the bidi Unicode algorithm, namely, Arabic-like, Hebrew-like and left to right. So, there should be at most 3 directives of this kind.

onchar= ids | fonts

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

intraspace= $\langle base \rangle$ $\langle shrink \rangle$ $\langle stretch \rangle$

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

intrapenalty= $\langle penalty \rangle$

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

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

1.16 Digits

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

For example:

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

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

New 3.30 With luatex there is an alternative approach for mapping digits, namely, `mapdigits`. Conversion is based on the language and it is applied to the typeset text (not

math, PDF bookmarks, etc.) before bidi and fonts are processed (ie, to the node list as generated by the \TeX code). This means the local digits have the correct bidirectional behavior (unlike `Numbers=Arabic` in `fontspec`, which is not recommended).

1.17 Accessing language info

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

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

`\iflanguage` $\{\langle\textit{language}\rangle\}\{\langle\textit{true}\rangle\}\{\langle\textit{false}\rangle\}$

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

`\localeinfo` $\{\langle\textit{field}\rangle\}$

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

`name.english` as provided by the Unicode CLDR.

`tag.ini` is the tag of the ini file (the way this file is identified in its name).

`tag.bcp47` is the BCP 47 language tag.

`tag.opentype` is the tag used by OpenType (usually, but not always, the same as BCP 47).

`script.name` as provided by the Unicode CLDR.

`script.tag.bcp47` is the BCP 47 language tag of the script used by this locale.

`script.tag.opentype` is the tag used by OpenType (usually, but not always, the same as BCP 47).

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

1.18 Hyphenation and line breaking

`\babelhyphen` $\ast\{\langle\textit{type}\rangle\}$

`\babelhyphen` $\ast\{\langle\textit{text}\rangle\}$

New 3.9a It is customary to classify hyphens in two types: (1) *explicit* or *hard hyphens*, which in \TeX are entered as `-`, and (2) *optional* or *soft hyphens*, which are entered as `\-`. Strictly, a *soft hyphen* is not a hyphen, but just a breaking opportunity or, in \TeX terms, a “discretionary”; a *hard hyphen* is a hyphen with a breaking opportunity after it. A further type is a *non-breaking hyphen*, a hyphen without a breaking opportunity. In \TeX , `-` and `\-` forbid further breaking opportunities in the word. This is the desired behavior very often, but not always, and therefore many languages provide shorthands for these cases. Unfortunately, this has not been done consistently: for example, `-` in Dutch, Portuguese, Catalan or Danish is a hard hyphen, while in German, Spanish, Norwegian, Slovak or Russian is a soft hyphen. Furthermore, some of them even redefine `\-`, so that you cannot insert a soft hyphen without breaking opportunities in the rest of the word.

Therefore, some macros are provided with a set of basic “hyphens” which can be used by themselves, to define a user shorthand, or even in language files.

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

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

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

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

`\babelhyphenation` [`<language>`, `<language>`, ...]{`<exceptions>`}

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

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

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

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

`\babelpatterns` [`<language>`, `<language>`, ...]{`<patterns>`}

New 3.9m *In *luatex* only*,¹⁵ adds or replaces patterns for the languages given or, without the optional argument, for *all* languages. If a pattern for a certain combination already exists, it gets replaced by the new one.

¹⁵With *luatex* exceptions and patterns can be modified almost freely. However, this is very likely a task for a separate package and *babel* only provides the most basic tools.

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

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

New 3.31 (Only `luatex`.) With `\babelprovide` and imported CJK languages, a simple generic line breaking algorithm (push-out-first) is applied, based on a selection of the Unicode rules (**New 3.32** it is disabled in verbatim mode, or more precisely when the `hyphenrules` are set to `nohyphenation`). It can be activated alternatively by setting explicitly the `intraspace`.

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

1.19 Selecting scripts

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

Some languages sharing the same script define macros to switch it (eg, `\textcyrillic`), but be aware they may also set the language to a certain default. Even the babel core defined `\textlatin`, but it was somewhat buggy because in some cases it messed up encodings and fonts (for example, if the main Latin encoding was `LY1`), and therefore it has been deprecated.¹⁷

`\ensureascii` `{⟨text⟩}`

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

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

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

¹⁶The so-called Unicode fonts do not improve the situation either. So, a font suited for Vietnamese is not necessarily suited for, say, the romanization of Indic languages, and the fact it contains glyphs for Modern Greek does not mean it includes them for Classic Greek.

¹⁷But still defined for backwards compatibility.

1.20 Selecting directions

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

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

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

There are some package options controlling bidi writing.

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

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

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

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

New 3.37 There is some experimental support for luahtex (with lualatex-dev) and the latest releases of luaotfload (3.11), with `Renderer = Harfbuzz` in fontspec. Since it is based on luatex, the option basic mostly works (You may need deactivate the `rtlm` or the `rtla` font features, or alternatively deactivate mirroring in babel with `\babeladjust`.) There are samples on GitHub, under `/required/babel/samples`. See particularly `lua-bidibasic.tex` and `lua-secenum.tex`.

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

```
\documentclass{article}

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

\babelprovide[import, main]{arabic}

\babelfont{rm}{FreeSerif}

\begin{document}
```

وقد عرفت شبه جزيرة العرب طيلة العصر الهيليني (اللاغريقي) بـ
 Arabia أو Aravia (باللاغريقية Ἀραβία)، استخدم الرومان ثلاث
 بادئات بـ“Arabia” على ثلاث مناطق من شبه الجزيرة العربية، إلا أنها
 حقيقةً كانت أكبر مما تعرف عليه اليوم.

`\end{document}`

EXAMPLE With `bidi=basic` both L and R text can be mixed without explicit markup (the latter will be only necessary in some special cases where the Unicode algorithm fails). It is used much like `bidi=basic-r`, but with R text inside L text you may want to map the font so that the correct features are in force. This is accomplished with an option in `\babelprovide`, as illustrated:

```
\documentclass{book}

\usepackage[english, bidi=basic]{babel}

\babelprovide[mapfont=direction]{arabic}

\babelfont{rm}{Crimson}
\babelfont[*arabic]{rm}{FreeSerif}

\begin{document}

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

\end{document}
```

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

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

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

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

layout= sectioning | counters | lists | contents | footnotes | captions | columns | graphics | extras

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

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

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

With counters, `\arabic` is not only considered L text always (with `\babelsublr`, see below), but also an “isolated” block which does not interact with the surrounding chars. So, while `1.2` in R text is rendered in that order with `bidi=basic` (as a decimal number), in `\arabic{c1}.\arabic{c2}` the visual order is `c2.c1`. Of course, you may always adjust the order by changing the language, if necessary.¹⁸

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

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

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

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

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

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

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

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

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

EXAMPLE Typically, in an Arabic document you would need:

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

`\babelsublr` `{\lr-text}`

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

¹⁸Next on the roadmap are counters and numeral systems in general. Expect some minor readjustments.

if necessary. It's intended for what Unicode calls weak characters, because words are best set with the corresponding language. For this reason, there is no `rl` counterpart. Any `\babelsublr` in *explicit* L mode is ignored. However, with `bidi=basic` and *implicit* L, it first returns to R and then switches to explicit L. To clarify this point, consider, in an R context:

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

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

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

`\BabelPatchSection` `{<section-name>}`

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

`\BabelFootnote` `{<cmd>}{<local-language>}{<before>}{<after>}`

New 3.17 Something like:

```
\BabelFootnote{\parsfootnote}{\language\language}{\footnote}
```

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

```
\footnote{(\foreignlanguage{\language\language}{note})}
```

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

```
\BabelFootnote{\footnote}{\language\language}{\footnote}%
\BabelFootnote{\localfootnote}{\language\language}{\footnote}%
\BabelFootnote{\mainfootnote}{\footnote}
```

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

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

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

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

1.21 Language attributes

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

1.22 Hooks

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

\AddBabelHook [`<lang>`]{`<name>`}{`<event>`}{`<code>`}

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

Names containing the string `babel` are reserved (they are used, for example, by `\useshortands*` to add a hook for the event `afterextras`). **New 3.33** They may be also applied to a specific language with the optional argument; language-specific settings are executed after global ones.

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

addialect (language name, dialect name) Used by `luababel.def` to load the patterns if not preloaded.

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

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

defaultcommands Used (locally) in `\StartBabelCommands`.

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

stopcommands Used to reset the above, if necessary.

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

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

afterextras Just after executing `\extras<language>`. For example, the following deactivates shortands in all languages:


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

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

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

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

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

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

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

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

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

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

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

1.23 Languages supported by babel with ldf files

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

Afrikaans afrikaans

Azerbaijani azerbaijani

Basque basque

Breton breton

Bulgarian bulgarian

Catalan catalan

Croatian croatian

Czech czech

Danish danish

Dutch dutch

English english, USenglish, american, UKenglish, british, canadian, australian, newzealand

Esperanto esperanto

Estonian estonian

Finnish finnish

French french, francais, canadien, acadian

Galician galician

German austrian, german, germanb, ngerman, naustrian
Greek greek, polutonikogreek
Hebrew hebrew
Icelandic icelandic
Indonesian bahasa, indonesian, indon, bahasai
Interlingua interlingua
Irish Gaelic irish
Italian italian
Latin latin
Lower Sorbian lowersorbian
Malay bahasam, malay, melayu
North Sami samin
Norwegian norsk, nynorsk
Polish polish
Portuguese portuges, portuguese, brazilian, brazil
Romanian romanian
Russian russian
Scottish Gaelic scottish
Spanish spanish
Slovakian slovak
Slovenian slovene
Swedish swedish
Serbian serbian
Turkish turkish
Ukrainian ukrainian
Upper Sorbian uppersorbian
Welsh welsh

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

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

```

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

```

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

1.24 Unicode character properties in luatex

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

\backslash babelcharproperty $\{ \langle char-code \rangle \} [\langle to-char-code \rangle] \{ \langle property \rangle \} \{ \langle value \rangle \}$

New 3.32 Here, $\{ \langle char-code \rangle \}$ is a number (with \TeX syntax). With the optional argument,

you can set a range of values. There are three properties (with a short name, taken from Unicode): direction (bc), mirror (bmg), linebreak (lb). The settings are global. For example:

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

This command is allowed only in vertical mode (the preamble or between paragraphs).

1.25 Tweaking some features

`\babeladjust` *{(key-value-list)}*

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

1.26 Tips, workarounds, known issues and notes

- If you use the document class *book* and you use `\ref` inside the argument of `\chapter` (or just use `\ref` inside `\MakeUppercase`), \LaTeX will keep complaining about an undefined label. To prevent such problems, you could revert to using uppercase labels, you can use `\lowercase{\ref{foo}}` inside the argument of `\chapter`, or, if you will not use shorthands in labels, set the `safe` option to `none` or `bib`.
- Both `ltxdoc` and `babel` use `\AtBeginDocument` to change some catcodes, and `babel` reloads `hline` to make sure `:` has the right one, so if you want to change the catcode of `|` it has to be done using the same method at the proper place, with

```
\AtBeginDocument{\DeleteShortVerb{\|}}
```

before loading `babel`. This way, when the document begins the sequence is (1) make `|` active (`ltxdoc`); (2) make it unactive (your settings); (3) make `babel` shorthands active (`babel`); (4) reload `hline` (`babel`, now with the correct catcodes for `|` and `:`).

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

```
\addto\extrasfrench{\inputencoding{latin1}}
\addto\extrasrussian{\inputencoding{koi8-r}}
```

(A recent version of `inputenc` is required.)

- For the hyphenation to work correctly, `lccodes` cannot change, because \TeX only takes into account the values when the paragraph is hyphenated, i.e., when it has been

finished.¹⁹ So, if you write a chunk of French text with `\foreignlanguage`, the apostrophes might not be taken into account. This is a limitation of \TeX , not of babel. Alternatively, you may use `\usesorthands` to activate ' and `\definesorthand`, or redefine `\textquoteright` (the latter is called by the non-ASCII right quote).

- `\bibitem` is out of sync with `\selectlanguage` in the `.aux` file. The reason is `\bibitem` uses `\immediate` (and others, in fact), while `\selectlanguage` doesn't. There is no known workaround.
- Babel does not take into account `\normalsfcodes` and (non-)French spacing is not always properly (un)set by languages. However, problems are unlikely to happen and therefore this part remains untouched in version 3.9 (but it is in the 'to do' list).
- Using a character mathematically active (ie, with math code "8000) as a shorthand can make \TeX enter in an infinite loop in some rare cases. (Another issue in the 'to do' list, although there is a partial solution.)

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

csquotes Logical markup for quotes.

iflang Tests correctly the current language.

hyphsubst Selects a different set of patterns for a language.

translator An open platform for packages that need to be localized.

siunitx Typesetting of numbers and physical quantities.

biblatex Programmable bibliographies and citations.

bicaption Bilingual captions.

babelbib Multilingual bibliographies.

microtype Adjusts the typesetting according to some languages (kerning and spacing).

Ligatures can be disabled.

substitutefont Combines fonts in several encodings.

mkpattern Generates hyphenation patterns.

tracklang Tracks which languages have been requested.

ucharclasses (xetex) Switches fonts when you switch from one Unicode block to another.

zhspacing Spacing for CJK documents in xetex.

1.27 Current and future work

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

Useful additions would be, for example, time, currency, addresses and personal names.²⁰ But that is the easy part, because they don't require modifying the \LaTeX internals.

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

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

An option to manage bidirectional document layout in luatex (lists, footnotes, etc.) is almost finished, but xetex required more work. Unfortunately, proper support for xetex requires patching somehow lots of macros and packages (and some issues related to

¹⁹This explains why \LaTeX assumes the lowercase mapping of T1 and does not provide a tool for multiple mappings. Unfortunately, `\savingsphcodes` is not a solution either, because `lccodes` for hyphenation are frozen in the format and cannot be changed.

²⁰See for example POSIX, ISO 14652 and the Unicode Common Locale Data Repository (CLDR). Those systems, however, have limited application to \TeX because their aim is just to display information and not fine typesetting.

\specials remain, like color and hyperlinks), so babel resorts to the bidi package (by Vafa Khalighi). See the babel repository for a small example (xe-bidi).

1.28 Tentative and experimental code

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

Modifying, and adding, values of ini files

New 3.37 There is a way to modify the values of ini files when they get loaded with \babelprovide. To set, say, digits.native in the numbers section, use something like numbers..digits.native=abcdefghijkl (note the double dot between the section and the key name). New keys may be added, too.

Non-standard hyphenation

New 3.37 With luatex it is now possible to define non-standard hyphenation rules, like f-f → ff-f. No rules are currently provided by default, but they can be defined as shown in the following example:

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

This feature must be explicitly activated with:

```
\babeladjust{ hyphenation.extra = on }
```

See the babel wiki for a more detailed description and some examples:

```
https://github.com/latex3/babel/wiki
```

Old and deprecated stuff

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

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

So, for example:

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

2 Loading languages with `language.dat`

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

New 3.9q With luatex, however, patterns are loaded on the fly when requested by the language (except the “0th” language, typically english, which is preloaded always).²¹ Until 3.9n, this task was delegated to the package `luatex-hyphen`, by Khaled Hosny, Élie Roux, and Manuel Pégourié-Gonnard, and required an extra file named `language.dat.lua`, but now a new mechanism has been devised based solely on `language.dat`. **You must rebuild the formats** if upgrading from a previous version. You may want to have a local `language.dat` for a particular project (for example, a book on Chemistry).²²

2.1 Format

In that file the person who maintains a TeX environment has to record for which languages he has hyphenation patterns *and* in which files these are stored²³. When hyphenation exceptions are stored in a separate file this can be indicated by naming that file *after* the file with the hyphenation patterns.

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

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

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

You may also set the font encoding the patterns are intended for by following the language name by a colon and the encoding code.²⁴ For example:

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

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

A typical error when using babel is the following:

```
No hyphenation patterns were preloaded for
the language '<lang>' into the format.
Please, configure your TeX system to add them and
```

²¹This feature was added to 3.9o, but it was buggy. Both 3.9o and 3.9p are deprecated.

²²The loader for lua(e)tex is slightly different as it's not based on babel but on `etex.src`. Until 3.9p it just didn't work, but thanks to the new code it works by reloading the data in the babel way, i.e., with `language.dat`.

²³This is because different operating systems sometimes use very different file-naming conventions.

²⁴This is not a new feature, but in former versions it didn't work correctly.

```
rebuild the format. Now I will use the patterns
preloaded for english instead}}
```

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

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

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

The following assumptions are made:

- Some of the language-specific definitions might be used by plain \TeX users, so the files have to be coded so that they can be read by both \LaTeX and plain \TeX . The current format can be checked by looking at the value of the macro `\fmtname`.
- The common part of the babel system redefines a number of macros and environments (defined previously in the document style) to put in the names of macros that replace the previously hard-wired texts. These macros have to be defined in the language definition files.
- The language definition files must define five macros, used to activate and deactivate the language-specific definitions. These macros are `\<lang>hyphenmins`, `\captions<lang>`, `\date<lang>`, `\extras<lang>` and `\noextras<lang>` (the last two may be left empty); where `<lang>` is either the name of the language definition file or the name of the \LaTeX option that is to be used. These macros and their functions are discussed below. You must define all or none for a language (or a dialect); defining, say, `\date<lang>` but not `\captions<lang>` does not raise an error but can lead to unexpected results.
- When a language definition file is loaded, it can define `\l@<lang>` to be a dialect of `\language0` when `\l@<lang>` is undefined.
- Language names must be all lowercase. If an unknown language is selected, babel will attempt setting it after lowercasing its name.
- The semantics of modifiers is not defined (on purpose). In most cases, they will just be simple separated options (eg, `spanish`), but a language might require, say, a set of options organized as a tree with suboptions (in such a case, the recommended separator is `/`).

Some recommendations:

- The preferred shorthand is `"`, which is not used in \LaTeX (quotes are entered as ``` and `'`). Other good choices are characters which are not used in a certain context (eg, `=` in an ancient language). Note however `=`, `<`, `>`, `:` and the like can be dangerous, because they may be used as part of the syntax of some elements (numeric expressions, key/value pairs, etc.).
- Captions should not contain shorthands or encoding-dependent commands (the latter is not always possible, but should be clearly documented). They should be defined using the LICR. You may also use the new tools for encoded strings, described below.

- Avoid adding things to `\noextras⟨lang⟩` except for `umlauthhigh` and friends, `\bbl@deactivate`, `\bbl@(non)frenchspacing`, and language-specific macros. Use always, if possible, `\bbl@save` and `\bbl@savevariable` (except if you still want to have access to the previous value). Do not reset a macro or a setting to a hardcoded value. Never. Instead save its value in `\extras⟨lang⟩`.
- Do not switch scripts. If you want to make sure a set of glyphs is used, switch either the font encoding (low-level) or the language (high-level, which in turn may switch the font encoding). Usage of things like `\latintext` is deprecated.²⁵
- Please, for “private” internal macros do not use the `\bbl@` prefix. It is used by `babel` and it can lead to incompatibilities.

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

3.1 Guidelines for contributed languages

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

- Do not hesitate stating on the file heads you are the author and the maintainer, if you actually are. There is no need to state the `babel` maintainer(s) as authors if they have not contributed significantly to your language files.
- Fonts are not strictly part of a language, so they are best placed in the corresponding TeX tree. This includes not only `tfm`, `vf`, `ps1`, `otf`, `mf` files and the like, but also `fd` ones.
- Font and input encodings are usually best placed in the corresponding tree, too, but sometimes they belong more naturally to the `babel` style. Note you may also need to define a LICR.
- `Babel` `ldf` files may just interface a framework, as it happens often with Oriental languages/scripts. This framework is best placed in its own directory.

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

3.2 Basic macros

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

- `\addlanguage` The macro `\addlanguage` is a non-outer version of the macro `\newlanguage`, defined in `plain.tex` version 3.x. For older versions of `plain.tex` and `lplain.tex` a substitute definition is used. Here “language” is used in the TeX sense of set of hyphenation patterns.
- `\adddialect` The macro `\adddialect` can be used when two languages can (or must) use the same hyphenation patterns. This can also be useful for languages for which no patterns are preloaded in the format. In such cases the default behavior of the `babel` system is to define

²⁵But not removed, for backward compatibility.

`\<lang>hyphenmins` this language as a ‘dialect’ of the language for which the patterns were loaded as `\language0`. Here “language” is used in the \TeX sense of set of hyphenation patterns. The macro `\<lang>hyphenmins` is used to store the values of the `\lefthyphenmin` and `\righthyphenmin`. Redefine this macro to set your own values, with two numbers corresponding to these two parameters. For example:

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

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

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

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

`\date<lang>` The macro `\date<lang>` defines `\today`.

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

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

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

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

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

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

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

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

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

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

3.3 Skeleton

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

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

\ifx\undefined\l@<language>
  \@nopatterns{<Language>}
  \adddialect\l@<language>0
\fi

\adddialect\l@<dialect>\l@<language>

\bbI@declare@ttribute{<language>}{<attrib>}{%
  \expandafter\addto\expandafter\extras<language>
  \expandafter{\extras<attrib><language>}%
  \let\captions<language>\captions<attrib><language>}

\providehyphenmins{<language>}{\tw@\thr@@}

\StartBabelCommands*{<language>}{captions}
\SetString\chaptername{<chapter name>}
% More strings

\StartBabelCommands*{<language>}{date}
\SetString\monthname{<name of first month>}
% More strings

\StartBabelCommands*{<dialect>}{captions}
\SetString\chaptername{<chapter name>}
% More strings

\StartBabelCommands*{<dialect>}{date}
\SetString\monthname{<name of first month>}
% More strings

\EndBabelCommands

\addto\extras<language>{}
\addto\noextras<language>{}
\let\extras<dialect>\extras<language>
\let\noextras<dialect>\noextras<language>

\ldf@finish{<language>}
```

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

```
\AtEndOfPackage{%
  \RequirePackage{dingbat}%      Delay package
```

<code>\savebox{\myeye}{\eye}%</code>	And direct usage
<code>\newsavebox{\myeye}</code>	
<code>\newcommand\myanchor{\anchor}%</code>	But OK inside command

3.4 Support for active characters

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

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

`\bbl@activate` The command `\bbl@activate` is used to change the way an active character expands.
`\bbl@deactivate` `\bbl@activate` ‘switches on’ the active behavior of the character. `\bbl@deactivate` lets the active character expand to its former (mostly) non-active self.

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

`\bbl@add@special` The \TeX book states: “Plain \TeX includes a macro called `\dospecials` that is essentially a set
`\bbl@remove@special` macro, representing the set of all characters that have a special category code.” [4, p. 380] It is used to set text ‘verbatim’. To make this work if more characters get a special category code, you have to add this character to the macro `\dospecial`. \TeX adds another macro called `\@sanitize` representing the same character set, but without the curly braces. The macros `\bbl@add@special⟨char⟩` and `\bbl@remove@special⟨char⟩` add and remove the character `⟨char⟩` to these two sets.

3.5 Support for saving macro definitions

Language definition files may want to *redefine* macros that already exist. Therefore a mechanism for saving (and restoring) the original definition of those macros is provided. We provide two macros for this²⁶.

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

`\babel@savevariable` A second macro is provided to save the current value of a variable. In this context, anything that is allowed after the `\the` primitive is considered to be a variable. The macro takes one argument, the `⟨variable⟩`.
The effect of the preceding macros is to append a piece of code to the current definition of `\originalTeX`. When `\originalTeX` is expanded, this code restores the previous definition of the control sequence or the previous value of the variable.

3.6 Support for extending macros

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

²⁶This mechanism was introduced by Bernd Raichle.

3.7 Macros common to a number of languages

<code>\bbl@allowhyphens</code>	In several languages compound words are used. This means that when \TeX has to hyphenate such a compound word, it only does so at the ‘-’ that is used in such words. To allow hyphenation in the rest of such a compound word, the macro <code>\bbl@allowhyphens</code> can be used.
<code>\allowhyphens</code>	Same as <code>\bbl@allowhyphens</code> , but does nothing if the encoding is T1. It is intended mainly for characters provided as real glyphs by this encoding but constructed with <code>\accent</code> in OT1. Note the previous command (<code>\bbl@allowhyphens</code>) has different applications (hyphens and discretionary) than this one (composite chars). Note also prior to version 3.7, <code>\allowhyphens</code> had the behavior of <code>\bbl@allowhyphens</code> .
<code>\set@low@box</code>	For some languages, quotes need to be lowered to the baseline. For this purpose the macro <code>\set@low@box</code> is available. It takes one argument and puts that argument in an <code>\hbox</code> , at the baseline. The result is available in <code>\box0</code> for further processing.
<code>\save@sf@q</code>	Sometimes it is necessary to preserve the <code>\spacefactor</code> . For this purpose the macro <code>\save@sf@q</code> is available. It takes one argument, saves the current <code>spacefactor</code> , executes the argument, and restores the <code>spacefactor</code> .
<code>\bbl@frenchspacing</code> <code>\bbl@nonfrenchspacing</code>	The commands <code>\bbl@frenchspacing</code> and <code>\bbl@nonfrenchspacing</code> can be used to properly switch French spacing on and off.

3.8 Encoding-dependent strings

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

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

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

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

`\StartBabelCommands` $\{ \langle \textit{language-list} \rangle \} \{ \langle \textit{category} \rangle \} [\langle \textit{selector} \rangle]$

The $\langle \textit{language-list} \rangle$ specifies which languages the block is intended for. A block is taken into account only if the `\CurrentOption` is listed here. Alternatively, you can define `\BabelLanguages` to a comma-separated list of languages to be defined (if undefined, `\StartBabelCommands` sets it to `\CurrentOption`). You may write `\CurrentOption` as the language, but this is discouraged – a explicit name (or names) is much better and clearer. A “selector” is a name to be used as value in package option strings, optionally followed by extra info about the encodings to be used. The name `unicode` must be used for `xetex` and `luatex` (the key strings has also other two special values: `generic` and `encoded`).

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

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

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

Blocks without a selector are read always if the key `strings` has been used. They provide fallback values, and therefore must be the last blocks; they should be provided always if possible and all strings should be defined somehow inside it; they can be the only blocks (mainly LGC scripts using the LICR). Blocks without a selector can be activated explicitly with `strings=generic` (no block is taken into account except those). With `strings=encoded`, strings in those blocks are set as default (internally, `?`). With `strings=encoded` strings are protected, but they are correctly expanded in `\MakeUppercase` and the like. If there is no key `strings`, string definitions are ignored, but `\SetCases` are still honored (in an encoded way).

The `<category>` is either `captions`, `date` or `extras`. You must stick to these three categories, even if no error is raised when using other name.²⁷ It may be empty, too, but in such a case using `\SetString` is an error (but not `\SetCase`).

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

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

\EndBabelCommands
```

A real example is:

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

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

\StartBabelCommands{austrian}{date}
\SetString\monthinname{J\"a\"nner}

\StartBabelCommands{german}{date}
\SetString\monthinname{Januar}

\StartBabelCommands{german,austrian}{date}
\SetString\monthiiname{Februar}
\SetString\monthiiiname{M\"a\"rz}
\SetString\monthivname{April}
\SetString\monthvname{Mai}
\SetString\monthviname{Juni}
\SetString\monthviiname{Juli}
\SetString\monthviiiname{August}
\SetString\monthixname{September}
\SetString\monthxname{Oktober}
\SetString\monthxiname{November}
\SetString\monthxiiname{Dezenber}
```

²⁷In future releases further categories may be added.

```

\SetString\today{\number\day.\sim%
\csname month\romannumeral\month name\endcsname\space
\number\year}

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

\EndBabelCommands

```

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

$\backslash StartBabelCommands$ $\star \{ \langle language-list \rangle \} \{ \langle category \rangle \} [\langle selector \rangle]$

The starred version just forces strings to take a value – if not set as package option, then the default for the engine is used. This is not done by default to prevent backward incompatibilities, but if you are creating a new language this version is better. It's up to the maintainers of the current languages to decide if using it is appropriate.²⁸

$\backslash EndBabelCommands$ Marks the end of the series of blocks.

$\backslash AfterBabelCommands$ $\{ \langle code \rangle \}$

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

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

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

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

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

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

```

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

```

#1 is replaced by the roman numeral.

$\backslash SetCase$ $[\langle map-list \rangle] \{ \langle toupper-code \rangle \} \{ \langle tolower-code \rangle \}$

Sets globally code to be executed at $\backslash MakeUppercase$ and $\backslash MakeLowercase$. The code would typically be things like $\backslash let \backslash BB \backslash bb$ and $\backslash uccode$ or $\backslash lccode$ (although for the reasons explained above, changes in lc/uc codes may not work). A $\langle map-list \rangle$ is a series of macros using the internal format of $\backslash @uclclist$ (eg, $\backslash bb \backslash BB \backslash cc \backslash CC$). The mandatory

²⁸This replaces in 3.9g a short-lived $\backslash UseStrings$ which has been removed because it did not work.

arguments take precedence over the optional one. This command, unlike `\SetString`, is executed always (even without strings), and it is intended for minor readjustments only. For example, as T1 is the default case mapping in \LaTeX , we could set for Turkish:

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

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

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

\EndBabelCommands
```

(Note the mapping for OT1 is not complete.)

`\SetHyphenMap` $\{ \langle to\text{-}lower\text{-}macros \rangle \}$

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

- `\BabelLower{<uccode>}{<lccode>}` is similar to `\lccode` but it's ignored if the char has been set and saves the original `lccode` to restore it when switching the language (except with `hyphenmap=first`).
- `\BabelLowerMM{<uccode-from>}{<uccode-to>}{<step>}{<lccode-from>}` loops though the given uppercase codes, using the step, and assigns them the `lccode`, which is also increased (MM stands for *many-to-many*).
- `\BabelLowerMO{<uccode-from>}{<uccode-to>}{<step>}{<lccode>}` loops though the given uppercase codes, using the step, and assigns them the `lccode`, which is fixed (MO stands for *many-to-one*).

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

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

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

4 Changes

4.1 Changes in babel version 3.9

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

- `\select@language` did not set `\language`. This meant the language in force when auxiliary files were loaded was the one used in, for example, shorthands – if the language was german, a `\select@language{spanish}` had no effect.
- `\foreignlanguage` and `otherlanguage*` messed up `\extras<language>`. Scripts, encodings and many other things were not switched correctly.
- The `:ENC` mechanism for hyphenation patterns used the encoding of the *previous* language, not that of the language being selected.
- `'` (with `activeacute`) had the original value when writing to an auxiliary file, and things like an infinite loop could happen. It worked incorrectly with `^` (if activated) and also if deactivated.
- Active chars were not reset at the end of language options, and that led to incompatibilities between languages.
- `\textormath` raised an error with a conditional.
- `\aliasshorthand` didn't work (or only in a few and very specific cases).
- `\l@english` was defined incorrectly (using `\let` instead of `\chardef`).
- `ldf` files not bundled with babel were not recognized when called as global options.

Part II

Source code

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

5 Identification and loading of required files

Code documentation is still under revision.

The babel package after unpacking consists of the following files:

switch.def defines macros to set and switch languages.

babel.def defines the rest of macros. It has two parts: a generic one and a second one only for LaTeX.

babel.sty is the \LaTeX package, which sets options and loads language styles.

plain.def defines some \LaTeX macros required by `babel.def` and provides a few tools for Plain.

hyphen.cfg is the file to be used when generating the formats to load hyphenation patterns. By default it also loads `switch.def`.

The babel installer extends docstrip with a few “pseudo-guards” to set “variables” used at installation time. They are used with `<@name@>` at the appropriated places in the source code and shown below with `<<name>>`. That brings a little bit of literate programming.

6 locale directory

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

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

This is a preliminary documentation.

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

Most keys are self-explanatory.

charset the encoding used in the ini file.

version of the ini file

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

encodings a descriptive list of font encodings.

[captions] section of captions in the file charset

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

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

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

7 Tools

```
1 <<version=3.38>>
2 <<date=2020/01/15>>
```

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

We define some basic macros which just make the code cleaner. `\bbl@add` is now used internally instead of `\addto` because of the unpredictable behavior of the latter. Used in `babel.def` and in `babel.sty`, which means in \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@stripslash{\expandafter\@gobble\string}
6 \def\bbl@add#1#2{%
```



```

7 \bbl@ifunset{\bbl@stripslash#1}%
8 {\def#1{#2}}%
9 {\expandafter\def\expandafter#1\expandafter{#1#2}}
10 \def\bbl@xin@{\@expandtwoargs\in@}
11 \def\bbl@csarg#1#2{\expandafter#1\csname bbl@#2\endcsname}%
12 \def\bbl@cs#1{\csname bbl@#1\endcsname}
13 \def\bbl@loop#1#2#3{\bbl@loop#1{#3}#2,\@nnil,}
14 \def\bbl@loopx#1#2{\expandafter\bbl@loop\expandafter#1\expandafter{#2}}
15 \def\bbl@loop#1#2#3,{%
16 \ifx\@nnil#3\relax\else
17 \def#1{#3}#2\bbl@afterfi\bbl@loop#1{#2}%
18 \fi}
19 \def\bbl@for#1#2#3{\bbl@loopx#1{#2}{\ifx#1\@empty\else#3\fi}}

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

20 \def\bbl@add@list#1#2{%
21 \edef#1{%
22 \bbl@ifunset{\bbl@stripslash#1}%
23 {}%
24 {\ifx#1\@empty\else#1,\fi}%
25 #2}}

\bbl@afterelse \bbl@afterfi Because the code that is used in the handling of active characters may need to look ahead,
we take extra care to ‘throw’ it over the \else and \fi parts of an \if-statement29. These
macros will break if another \if... \fi statement appears in one of the arguments and it
is not enclosed in braces.

26 \long\def\bbl@afterelse#1\else#2\fi{\fi#1}
27 \long\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 and \<. > for \noexpand applied to a built
macro name (the latter does not define the macro if undefined to \relax, because it is
created locally). The result may be followed by extra arguments, if necessary.

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

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

34 \def\bbl@tempa#1{%
35 \long\def\bbl@trim##1##2{%
36 \futurelet\bbl@trim@a\bbl@trim@c##2\@nil\@nil#1\@nil\relax{##1}}%
37 \def\bbl@trim@c{%
38 \ifx\bbl@trim@a\@sptoken
39 \expandafter\bbl@trim@b
40 \else

```

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

```

41 \expandafter\bb1@trim@b\expandafter#1%
42 \fi}%
43 \long\def\bb1@trim@b#1##1 \@nil{\bb1@trim@i##1}}
44 \bb1@tempa{ }
45 \long\def\bb1@trim@i#1\@nil#2\relax#3{#3{#1}}
46 \long\def\bb1@trim@def#1{\bb1@trim{\def#1}}

```

`\bb1@ifunset` To check if a macro is defined, we create a new macro, which does the same as `\@ifundefined`. However, in an ϵ -tex engine, it is based on `\ifcsname`, which is more efficient, and do not waste memory.

```

47 \begingroup
48 \gdef\bb1@ifunset#1{%
49 \expandafter\ifx\csname#1\endcsname\relax
50 \expandafter\@firstoftwo
51 \else
52 \expandafter\@secondoftwo
53 \fi}
54 \bb1@ifunset{ifcsname}%
55 {}%
56 {\gdef\bb1@ifunset#1{%
57 \ifcsname#1\endcsname
58 \expandafter\ifx\csname#1\endcsname\relax
59 \bb1@afterelse\expandafter\@firstoftwo
60 \else
61 \bb1@afterfi\expandafter\@secondoftwo
62 \fi
63 \else
64 \expandafter\@firstoftwo
65 \fi}}
66 \endgroup

```

`\bb1@ifblank` A tool from url, by Donald Arseneau, which tests if a string is empty or space.

```

67 \def\bb1@ifblank#1{%
68 \bb1@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
69 \long\def\bb1@ifblank@i#1#2\@nil#3#4#5\@nil{#4}

```

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

```

70 \def\bb1@forkv#1#2{%
71 \def\bb1@kvcmd##1##2##3{#2}%
72 \bb1@kvnext#1,\@nil,}
73 \def\bb1@kvnext#1,{%
74 \ifx\@nil#1\relax\else
75 \bb1@ifblank{#1}{\bb1@forkv@eq#1=\@empty=\@nil{#1}}%
76 \expandafter\bb1@kvnext
77 \fi}
78 \def\bb1@forkv@eq#1=#2=#3\@nil#4{%
79 \bb1@trim@def\bb1@forkv@a{#1}%
80 \bb1@trim{\expandafter\bb1@kvcmd\expandafter{\bb1@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).

```

81 \def\bb1@vforeach#1#2{%
82 \def\bb1@forcmd##1{#2}%
83 \bb1@fornext#1,\@nil,}
84 \def\bb1@fornext#1,{%
85 \ifx\@nil#1\relax\else

```

```

86 \bbl@ifblank{#1}{\bbl@trim\bbl@forcmd{#1}}%
87 \expandafter\bbl@fornext
88 \fi}
89 \def\bbl@foreach#1{\expandafter\bbl@vforeach\expandafter{#1}}

```

\bbl@replace

```

90 \def\bbl@replace#1#2#3{% in #1 -> repl #2 by #3
91 \toks@{}}%
92 \def\bbl@replace@aux##1#2##2#2{%
93 \ifx\bbl@nil##2%
94 \toks@\expandafter{\the\toks@##1}%
95 \else
96 \toks@\expandafter{\the\toks@##1#3}%
97 \bbl@afterfi
98 \bbl@replace@aux##2#2%
99 \fi}%
100 \expandafter\bbl@replace@aux#1#2\bbl@nil#2%
101 \edef#1{\the\toks@}}

```

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

```

102 \bbl@exp{\def\\bbl@parsedef##1\detokenize{macro:}}#2->#3\relax{%
103 \def\bbl@tempa{#1}%
104 \def\bbl@tempb{#2}%
105 \def\bbl@tempe{#3}}
106 \def\bbl@sreplace#1#2#3{%
107 \begingroup
108 \expandafter\bbl@parsedef\meaning#1\relax
109 \def\bbl@tempc{#2}%
110 \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
111 \def\bbl@tempd{#3}%
112 \edef\bbl@tempd{\expandafter\strip@prefix\meaning\bbl@tempd}%
113 \bbl@xin@{\bbl@tempc}{\bbl@tempe}% If not in macro, do nothing
114 \ifin@
115 \bbl@exp{\\bbl@replace\\bbl@tempe{\bbl@tempc}{\bbl@tempd}}%
116 \def\bbl@tempc{% Expanded an executed below as 'uplevel'
117 \\\makeatletter % "internal" macros with @ are assumed
118 \\\scantokens{%
119 \bbl@tempa\\@namedef{\bbl@stripslash#1}\bbl@tempb{\bbl@tempe}}%
120 \catcode64=\the\catcode64\relax}% Restore @
121 \else
122 \let\bbl@tempc@empty % Not \relax
123 \fi
124 \bbl@exp{% For the 'uplevel' assignments
125 \endgroup
126 \bbl@tempc}} % empty or expand to set #1 with changes

```

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

```

127 \def\bbl@ifsamestring#1#2{%
128 \begingroup
129 \protected@edef\bbl@tempb{#1}%

```

```

130 \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
131 \protected@edef\bbl@tempc{#2}%
132 \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
133 \ifx\bbl@tempb\bbl@tempc
134 \aftergroup\@firstoftwo
135 \else
136 \aftergroup\@secondoftwo
137 \fi
138 \endgroup}
139 \chardef\bbl@engine=%
140 \ifx\directlua\@undefined
141 \ifx\XeTeXinputencoding\@undefined
142 \z@
143 \else
144 \tw@
145 \fi
146 \else
147 \@ne
148 \fi
149 <</Basic macros>>

```

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

```

150 <<*Make sure ProvidesFile is defined>> ≡
151 \ifx\ProvidesFile\@undefined
152 \def\ProvidesFile#1[#2 #3 #4]{%
153 \wlog{File: #1 #4 #3 <#2>}%
154 \let\ProvidesFile\@undefined}
155 \fi
156 <</Make sure ProvidesFile is defined>>

```

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

```

157 <<*Load patterns in luatex>> ≡
158 \ifx\directlua\@undefined\else
159 \ifx\bbl@luapatterns\@undefined
160 \input luababel.def
161 \fi
162 \fi
163 <</Load patterns in luatex>>

```

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

```

164 <<*Load macros for plain if not LaTeX>> ≡
165 \ifx\AtBeginDocument\@undefined
166 \input plain.def\relax
167 \fi
168 <</Load macros for plain if not LaTeX>>

```

7.1 Multiple languages

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

```

169 <<*Define core switching macros>> ≡
170 \ifx\language\@undefined
171 \csname newcount\endcsname\language

```

```

172 \fi
173 <</Define core switching macros>>

```

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

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

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

```

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

```

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

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

8 The Package File (L^AT_EX, `babel.sty`)

In order to make use of the features of L^AT_EX 2_ε, the `babel` system contains a package file, `babel.sty`. This file is loaded by the `\usepackage` command and defines all the language options whose name is different from that of the `.ldf` file (like variant spellings). It also takes care of a number of compatibility issues with other packages and defines a few additional package options.

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

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

8.1 base

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

```
190 (*package)
191 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
192 \ProvidesPackage{babel}[\langle\langle date \rangle\rangle \langle\langle version \rangle\rangle The Babel package]
193 \@ifpackagewith{babel}{debug}
194   {\providecommand\bbl@trace[1]{\message{^^J[ #1 ]}}}%
195   \let\bbl@debug\@firstofone}
196   {\providecommand\bbl@trace[1]{}}%
197   \let\bbl@debug\@gobble}
198 \ifx\bbl@switchflag\undefined % Prevent double input
199   \let\bbl@switchflag\relax
200   \input switch.def\relax
201 \fi
202 \langle\langle Load patterns in luatex \rangle\rangle
203 \langle\langle Basic macros \rangle\rangle
204 \def\AfterBabelLanguage#1{%
205   \global\expandafter\bbl@add\csname#1.ldf-h@k\endcsname}%
```

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.

```
206 \ifx\bbl@languages\undefined\else
207   \begingroup
208     \catcode\^^I=12
209     \@ifpackagewith{babel}{showlanguages}{%
210       \begingroup
211         \def\bbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}}%
212         \wlog{<*languages>}%
213         \bbl@languages
214         \wlog{</languages>}%
215       \endgroup}{%
216     \endgroup
217     \def\bbl@elt#1#2#3#4{%
218       \ifnum#2=\z@
219         \gdef\bbl@nulllanguage{#1}%
220         \def\bbl@elt##1##2##3##4{}%
221       \fi}%
222     \bbl@languages
223 \fi
224 \ifodd\bbl@engine
225   % Harftex is evolving, so the callback is not hardcoded, just in case
226   \def\bbl@harfprefline{Harf pre_linebreak_filter callback}%
227   \def\bbl@activate@preotf{%
228     \let\bbl@activate@preotf\relax % only once
229     \directlua{
230       Babel = Babel or {}
231       %
232       function Babel.pre_otfload_v(head)
233         if Babel.numbers and Babel.digits_mapped then
234           head = Babel.numbers(head)
235         end
236         if Babel.bidi_enabled then
237           head = Babel.bidi(head, false, dir)
238         end
239         return head
240       end
```

```

241 %
242 function Babel.pre_otfload_h(head, gc, sz, pt, dir)
243   if Babel.numbers and Babel.digits_mapped then
244     head = Babel.numbers(head)
245   end
246   if Babel.bidi_enabled then
247     head = Babel.bidi(head, false, dir)
248   end
249   return head
250 end
251 %
252 luatexbase.add_to_callback('pre_linebreak_filter',
253   Babel.pre_otfload_v,
254   'Babel.pre_otfload_v',
255   luatexbase.priority_in_callback('pre_linebreak_filter',
256     '\bbl@harfpreamble')
257   or luatexbase.priority_in_callback('pre_linebreak_filter',
258     'luaotfload.node_processor')
259   or nil)
260 %
261 luatexbase.add_to_callback('hpack_filter',
262   Babel.pre_otfload_h,
263   'Babel.pre_otfload_h',
264   luatexbase.priority_in_callback('hpack_filter',
265     '\bbl@harfpreamble')
266   or luatexbase.priority_in_callback('hpack_filter',
267     'luaotfload.node_processor')
268   or nil)
269 }%
270 \@ifpackageloaded{harfload}%
271   {\directlua{ Babel.mirroring_enabled = false }}%
272   {}%
273 \let\bbl@tempa\relax
274 \@ifpackagewith{babel}{bidi=basic}%
275   {\def\bbl@tempa{basic}}%
276   {\@ifpackagewith{babel}{bidi=basic-r}%
277     {\def\bbl@tempa{basic-r}}%
278     {}%
279   \ifx\bbl@tempa\relax\else
280     \let\bbl@beforeforeign\leavevmode
281     \AtEndOfPackage{\EnableBabelHook{babel-bidi}}%
282     \RequirePackage{luatexbase}%
283     \directlua{
284       require('babel-data-bidi.lua')
285       require('babel-bidi-\bbl@tempa.lua')
286     }
287     \bbl@activate@preotf
288   \fi
289 \fi

```

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

```

290 \bbl@trace{Defining option 'base'}
291 \@ifpackagewith{babel}{base}{%
292   \ifx\directlua\undefined
293     \DeclareOption*{\bbl@patterns{\CurrentOption}}%
294   \else
295     \DeclareOption*{\bbl@patterns@lua{\CurrentOption}}%
296   \fi

```

```

297 \DeclareOption{base}{}%
298 \DeclareOption{showlanguages}{}%
299 \ProcessOptions
300 \global\expandafter\let\csname opt@babel.sty\endcsname\relax
301 \global\expandafter\let\csname ver@babel.sty\endcsname\relax
302 \global\let@ifl@ter@@\ifl@ter
303 \def@ifl@ter#1#2#3#4#5{\global\let@ifl@ter\ifl@ter@@}%
304 \endinput}{}%

```

8.2 key=value options and other general option

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

```

305 \bbl@trace{key=value and another general options}
306 \bbl@csarg\let\tempa\expandafter\csname opt@babel.sty\endcsname
307 \def\bbl@tempb#1.#2{%
308   #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%
309 \def\bbl@tempd#1.#2@nnil{%
310   \ifx\@empty#2%
311     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
312   \else
313     \in@{=}{#1}\ifin@
314     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
315   \else
316     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
317     \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
318   \fi
319 \fi}
320 \let\bbl@tempc\@empty
321 \bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty\@nnil}
322 \expandafter\let\csname opt@babel.sty\endcsname\bbl@tempc

```

The next option tells babel to leave shorthand characters active at the end of processing the package. This is *not* the default as it can cause problems with other packages, but for those who want to use the shorthand characters in the preamble of their documents this can help.

```

323 \DeclareOption{KeepShorthandsActive}{}
324 \DeclareOption{activeacute}{}
325 \DeclareOption{activegrave}{}
326 \DeclareOption{debug}{}
327 \DeclareOption{noconfigs}{}
328 \DeclareOption{showlanguages}{}
329 \DeclareOption{silent}{}
330 \DeclareOption{mono}{}
331 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
332 % Don't use. Experimental:
333 \newif\ifbbl@single
334 \DeclareOption{selectors=off}{\bbl@singletrue}
335 <<More package options>>

```

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.


```

336 \let\bbl@opt@shorthands\@nnil
337 \let\bbl@opt@config\@nnil
338 \let\bbl@opt@main\@nnil
339 \let\bbl@opt@headfoot\@nnil
340 \let\bbl@opt@layout\@nnil

```

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

```

341 \def\bbl@tempa#1=#2\bbl@tempa{%
342   \bbl@csarg\ifx{opt@#1}\@nnil
343     \bbl@csarg\edef{opt@#1}{#2}%
344   \else
345     \bbl@error{%
346       Bad option `#1=#2'. Either you have misspelled the\\%
347       key or there is a previous setting of `#1'}{%
348       Valid keys are `shorthands', `config', `strings', `main',\\%
349       `headfoot', `safe', `math', among others.}
350   \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.

```

351 \let\bbl@language@opts\@empty
352 \DeclareOption*{%
353   \bbl@xin@{\string=}{\CurrentOption}%
354   \ifin@
355     \expandafter\bbl@tempa\CurrentOption\bbl@tempa
356   \else
357     \bbl@add@list\bbl@language@opts{\CurrentOption}%
358   \fi}

```

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

```

359 \ProcessOptions*

```

8.3 Conditional loading of shorthands

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

```

360 \bbl@trace{Conditional loading of shorthands}
361 \def\bbl@sh@string#1{%
362   \ifx#1\@empty\else
363     \ifx#1t\string~%
364     \else\ifx#1c\string,%
365     \else\string#1%
366   \fi\fi
367   \expandafter\bbl@sh@string
368 \fi}
369 \ifx\bbl@opt@shorthands\@nnil
370   \def\bbl@ifshorthand#1#2#3{#2}%
371 \else\ifx\bbl@opt@shorthands\@empty
372   \def\bbl@ifshorthand#1#2#3{#3}%
373 \else

```

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

```

374 \def\bbl@ifshorthand#1{%
375   \bbl@xin@{\string#1}{\bbl@opt@shorthands}%
376   \ifin@
377     \expandafter\@firstoftwo
378   \else
379     \expandafter\@secondoftwo
380   \fi}

```

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

```

381 \edef\bbl@opt@shorthands{%
382   \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.

```

383 \bbl@ifshorthand{'}%
384   {\PassOptionsToPackage{activeacute}{babel}}{}
385 \bbl@ifshorthand{`}%
386   {\PassOptionsToPackage{activegrave}{babel}}{}
387 \fi\fi

```

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

```

388 \ifx\bbl@opt@headfoot\@nnil\else
389   \g@addto@macro\@resetactivechars{%
390     \set@typeset@protect
391     \expandafter\select@language@x\expandafter{\bbl@opt@headfoot}%
392     \let\protect\noexpand}
393 \fi

```

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

```

394 \ifx\bbl@opt@safe\@undefined
395   \def\bbl@opt@safe{BR}
396 \fi
397 \ifx\bbl@opt@main\@nnil\else
398   \edef\bbl@language@opts{%
399     \ifx\bbl@language@opts\@empty\else\bbl@language@opts,\fi
400     \bbl@opt@main}
401 \fi

```

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

```

402 \bbl@trace{Defining IfBabelLayout}
403 \ifx\bbl@opt@layout\@nnil
404   \newcommand\IfBabelLayout[3]{#3}%
405 \else
406   \newcommand\IfBabelLayout[1]{%
407     \@expandtwoargs\in@{.#1.}{.\bbl@opt@layout.}%
408     \ifin@
409       \expandafter\@firstoftwo
410     \else
411       \expandafter\@secondoftwo
412     \fi}
413 \fi

```

8.4 Language options

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

```
414 \bbl@trace{Language options}
415 \let\bbl@afterlang\relax
416 \let\BabelModifiers\relax
417 \let\bbl@loaded@empty
418 \def\bbl@load@language#1{%
419   \InputIfFileExists{#1.ldf}%
420   {\edef\bbl@loaded{\CurrentOption
421     \ifx\bbl@loaded@empty\else,\bbl@loaded\fi}%
422     \expandafter\let\expandafter\bbl@afterlang
423       \csname\CurrentOption.ldf-h@@k\endcsname
424     \expandafter\let\expandafter\BabelModifiers
425       \csname bbl@mod@\CurrentOption\endcsname}%
426   {\bbl@error{%
427     Unknown option '\CurrentOption'. Either you misspelled it\\%
428     or the language definition file \CurrentOption.ldf was not found}{%
429     Valid options are: shorthands=, KeepShorthandsActive,\\%
430     activeacute, activegrave, noconfigs, safe=, main=, math=\\%
431     headfoot=, strings=, config=, hyphenmap=, or a language name.}}}
```

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

```
432 \def\bbl@try@load@lang#1#2#3{%
433   \IfFileExists{\CurrentOption.ldf}%
434   {\bbl@load@language{\CurrentOption}}%
435   {#1\bbl@load@language{#2}#3}}
436 \DeclareOption{afrikaans}{\bbl@try@load@lang{}{dutch}{}}
437 \DeclareOption{brazil}{\bbl@try@load@lang{}{portuges}{}}
438 \DeclareOption{brazilian}{\bbl@try@load@lang{}{portuges}{}}
439 \DeclareOption{hebrew}{%
440   \input{rlbabel.def}%
441   \bbl@load@language{hebrew}}
442 \DeclareOption{hungarian}{\bbl@try@load@lang{}{magyar}{}}
443 \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}{}}
444 \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}{}}
445 \DeclareOption{polutonikogreek}{%
446   \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}}
447 \DeclareOption{portuguese}{\bbl@try@load@lang{}{portuges}{}}
448 \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}{}}
449 \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}{}}
450 \DeclareOption{uppersorbian}{\bbl@try@load@lang{}{usorbian}{}}
```

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

```
451 \ifx\bbl@opt@config\@nnil
452   \@ifpackagewith{babel}{noconfigs}{}%
453   {\InputIfFileExists{bblopts.cfg}%
454     {\typeout{*****^J%
455       * Local config file bblopts.cfg used^^J%
456       *}}%
457     {}}%
```

```

458 \else
459   \InputIfFileExists{\bbl@opt@config.cfg}%
460   {\typeout{*****^J%
461             * Local config file \bbl@opt@config.cfg used^^J%
462             *}}%
463   {\bbl@error{%
464     Local config file '\bbl@opt@config.cfg' not found}{%
465     Perhaps you misspelled it.}}%
466 \fi

```

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

```

467 \bbl@for\bbl@tempa\bbl@language@opts{%
468   \bbl@ifunset{ds@\bbl@tempa}%
469   {\edef\bbl@tempb{%
470     \noexpand\DeclareOption
471     {\bbl@tempa}%
472     {\noexpand\bbl@load@language{\bbl@tempa}}}%
473   \bbl@tempb}%
474   \@empty}

```

Now, we make sure an option is explicitly declared for any language set as global option, by checking if an `ldf` exists. The previous step was, in fact, somewhat redundant, but that way we minimize accessing the file system just to see if the option could be a language.

```

475 \bbl@foreach\@classoptionslist{%
476   \bbl@ifunset{ds@#1}%
477   {\IfFileExists{#1.ldf}%
478    {\DeclareOption{#1}{\bbl@load@language{#1}}}%
479    {}}%
480   {}}

```

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

```

481 \ifx\bbl@opt@main\@nnil\else
482   \expandafter
483   \let\expandafter\bbl@loadmain\csname ds@\bbl@opt@main\endcsname
484   \DeclareOption{\bbl@opt@main}{}
485 \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 (except, of course, global options, which \LaTeX processes before):

```

486 \def\AfterBabelLanguage#1{%
487   \bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang}{}}
488 \DeclareOption*{}
489 \ProcessOptions*

```

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

```

490 \ifx\bbl@opt@main\@nnil
491   \edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
492   \let\bbl@tempc\@empty

```

```

493 \bbl@for\bbl@tempb\bbl@tempa{%
494   \bbl@xin@{,\bbl@tempb,}{,\bbl@loaded,}%
495   \ifin@edef\bbl@tempc{\bbl@tempb}\fi}
496 \def\bbl@tempa#1,#2\@nnil{\def\bbl@tempb{#1}}
497 \expandafter\bbl@tempa\bbl@loaded,\@nnil
498 \ifx\bbl@tempb\bbl@tempc\else
499   \bbl@warning{%
500     Last declared language option is '\bbl@tempc',\%
501     but the last processed one was '\bbl@tempb'.\%
502     The main language cannot be set as both a global\%
503     and a package option. Use 'main=\bbl@tempc' as\%
504     option. Reported}%
505   \fi
506 \else
507   \DeclareOption{\bbl@opt@main}{\bbl@loadmain}
508   \ExecuteOptions{\bbl@opt@main}
509   \DeclareOption*{}
510   \ProcessOptions*
511 \fi
512 \def\AfterBabelLanguage{%
513   \bbl@error
514   {Too late for \string\AfterBabelLanguage}%
515   {Languages have been loaded, so I can do nothing}}

```

In order to catch the case where the user forgot to specify a language we check whether `\bbl@main@language`, has become defined. If not, no language has been loaded and an error message is displayed.

```

516 \ifx\bbl@main@language\undefined
517   \bbl@info{%
518     You haven't specified a language. I'll use 'nil'\%
519     as the main language. Reported}
520   \bbl@load@language{nil}
521 \fi
522 \</package>
523 \<core>

```

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

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

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

9.1 Tools

```

524 \ifx\ldf@quit\undefined

```

```

525 \else
526   \expandafter\endinput
527 \fi
528 <<Make sure ProvidesFile is defined>>
529 \ProvidesFile{babel.def}[\<date>] \<version>] Babel common definitions]
530 <<Load macros for plain if not LaTeX>>

```

The file `babel.def` expects some definitions made in the $\text{\LaTeX} 2_{\epsilon}$ style file. So, In $\text{\LaTeX} 2.09$ and Plain we must provide at least some predefined values as well some tools to set them (even if not all options are available). There are no package options, and therefore an 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).

```

531 \ifx\bbbl@ifshorthand\@undefined
532   \let\bbbl@opt@shorthands\@nnil
533   \def\bbbl@ifshorthand#1#2#3{#2}%
534   \let\bbbl@language@opts\@empty
535   \ifx\babeloptionstrings\@undefined
536     \let\bbbl@opt@strings\@nnil
537   \else
538     \let\bbbl@opt@strings\babeloptionstrings
539   \fi
540   \def\BabelStringsDefault{generic}
541   \def\bbbl@tempa{normal}
542   \ifx\babeloptionmath\bbbl@tempa
543     \def\bbbl@mathnormal{\noexpand\textormath}
544   \fi
545   \def\AfterBabelLanguage#1#2{}
546   \ifx\BabelModifiers\@undefined\let\BabelModifiers\relax\fi
547   \let\bbbl@afterlang\relax
548   \def\bbbl@opt@safe{BR}
549   \ifx\@uclclist\@undefined\let\@uclclist\@empty\fi
550   \ifx\bbbl@trace\@undefined\def\bbbl@trace#1{}\fi
551   \expandafter\newif\csname ifbbbl@single\endcsname
552 \fi

```

And continue.

```

553 \ifx\bbbl@switchflag\@undefined % Prevent double input
554   \let\bbbl@switchflag\relax
555   \input switch.def\relax
556 \fi
557 \bbbl@trace{Compatibility with language.def}
558 \ifx\bbbl@languages\@undefined
559   \ifx\directlua\@undefined
560     \openin1 = language.def
561     \ifeof1
562       \closein1
563       \message{I couldn't find the file language.def}
564     \else
565       \closein1
566       \begingroup
567         \def\addlanguage#1#2#3#4#5{%
568           \expandafter\ifx\csname lang@#1\endcsname\relax\else
569             \global\expandafter\let\csname l@#1\endcsname\expandafter\endcsname
570             \csname lang@#1\endcsname
571           \fi}%
572         \def\uselanguage#1{}%
573         \input language.def
574       \endgroup

```

```

575 \fi
576 \fi
577 \chardef\l@english\z@
578 \fi
579 <<Load patterns in luatex>>
580 <<Basic macros>>

```

`\addto` For each language four control sequences have to be defined that control the language-specific definitions. To be able to add something to these macro once they have been defined the macro `\addto` is introduced. It takes two arguments, a *<control sequence>* and \TeX -code to be added to the *<control sequence>*.

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

```

581 \def\addto#1#2{%
582   \ifx#1\@undefined
583     \def#1{#2}%
584   \else
585     \ifx#1\relax
586       \def#1{#2}%
587     \else
588       {\toks@\expandafter{#1#2}%
589        \xdef#1{\the\toks@}}%
590   \fi
591 \fi}

```

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

```

592 \def\bbl@withactive#1#2{%
593   \begingroup
594     \lccode`~=#2\relax
595     \lowercase{\endgroup#1~}}

```

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

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

```

596 \def\bbl@redefine#1{%
597   \edef\bbl@tempa{\bbl@stripslash#1}%
598   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
599   \expandafter\def\csname\bbl@tempa\endcsname}

```

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

```

600 \@onlypreamble\bbl@redefine

```

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

```

601 \def\bbl@redefine@long#1{%
602   \edef\bbl@tempa{\bbl@stripslash#1}%
603   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
604   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
605 \@onlypreamble\bbl@redefine@long

```

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

```

606 \def\bbl@redefinero bust#1{%
607   \edef\bbl@tempa{\bbl@stripslash#1}%
608   \bbl@ifunset{\bbl@tempa\space}%
609   {\expandafter\let\csname org@\bbl@tempa\endcsname#1%
610     \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
611   {\bbl@exp{\let\<org@\bbl@tempa\>\<\bbl@tempa\space>}}}%
612   \@namedef{\bbl@tempa\space}}

```

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

```

613 \@onlypreamble\bbl@redefinero bust

```

9.2 Hooks

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

```

614 \bbl@trace{Hooks}
615 \newcommand\AddBabelHook[3][{}{%
616   \bbl@ifunset{bbl@hk@#2}{\EnableBabelHook{#2}}}%
617   \def\bbl@tempa##1,#3=##2,##3\@empty{\def\bbl@tempb{##2}}%
618   \expandafter\bbl@tempa\bbl@evargs,#3=,\@empty
619   \bbl@ifunset{bbl@ev@#2@#3@#1}%
620   {\bbl@csarg\bbl@add{ev@#3@#1}{\bbl@elt{#2}}}%
621   {\bbl@csarg\let{ev@#2@#3@#1}\relax}%
622   \bbl@csarg\newcommand{ev@#2@#3@#1}[\bbl@tempb]}
623 \newcommand\EnableBabelHook[1]{\bbl@csarg\let{hk@#1}\@firstofone}
624 \newcommand\DisableBabelHook[1]{\bbl@csarg\let{hk@#1}\@gobble}
625 \def\bbl@usehooks#1#2{%
626   \def\bbl@elt##1{%
627     \@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1@#2}}%
628     \@nameuse{bbl@ev@#1@}%
629     \ifx\language\@undefined\else % Test required for Plain (?)
630       \def\bbl@elt##1{%
631         \@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1@\language}#2}}%
632         \@nameuse{bbl@ev@#1@\language}%
633       \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).

```

634 \def\bbl@evargs{,% <- don't delete this comma
635   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
636   addialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
637   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
638   hyphenation=2,initiateactive=3,afterreset=0,foreign=0,foreign*=0,%
639   beforestart=0}

```

`\babelensure` The user command just parses the optional argument and creates a new macro named `\bbl@e@<language>`. We register a hook at the `afterextras` event which just executes this macro in a “complete” selection (which, if undefined, is `\relax` and does nothing). This

part is somewhat involved because we have to make sure things are expanded the correct number of times.

The macro `\bbl@e@{language}` contains `\bbl@ensure{<include>}{<exclude>}{<fontenc>}`, which in 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.

```

640 \bbl@trace{Defining babelensure}
641 \newcommand\babelensure[2][{}]{% TODO - revise test files
642   \AddBabelHook{babel-ensure}{afterextras}{%
643     \ifcase\bbl@select@type
644       \@nameuse{\bbl@e@\language}{}%
645     \fi}%
646   \begingroup
647     \let\bbl@ens@include\@empty
648     \let\bbl@ens@exclude\@empty
649     \def\bbl@ens@fontenc{\relax}%
650     \def\bbl@tempb##1{%
651       \ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}%
652     \edef\bbl@tempa{\bbl@tempb#1\@empty}%
653     \def\bbl@tempb##1=##2\@{\@namedef{\bbl@ens@##1}{##2}}%
654     \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
655     \def\bbl@tempc{\bbl@ensure}%
656     \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
657       \expandafter{\bbl@ens@include}}%
658     \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
659       \expandafter{\bbl@ens@exclude}}%
660     \toks@\expandafter{\bbl@tempc}%
661     \bbl@exp{%
662       \endgroup
663       \def\<bbl@e@#2>{\the\toks@{\bbl@ens@fontenc}}}%
664     \def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
665       \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
666         \ifx##1\@undefined % 3.32 - Don't assume the macros exists
667           \edef##1{\noexpand\bbl@nocaption
668             {\bbl@stripslash##1}{\language\bbl@stripslash##1}}%
669         \fi
670         \ifx##1\@empty\else
671           \in@{##1}{#2}%
672         \ifin@ \else
673           \bbl@ifunset{\bbl@ensure@\language}%
674           {\bbl@exp{%
675             \\\DeclareRobustCommand\<bbl@ensure@\language>[1]{%
676               \\\foreignlanguage{\language}%
677               {\ifx\relax#3\else
678                 \\\fontencoding{#3}\selectfont
679               \fi
680               #####1}}}%
681           {}%
682           \toks@\expandafter{##1}%
683           \edef##1{%
684             \bbl@csarg\noexpand{ensure@\language}%
685             {\the\toks@}}%
686         \fi
687         \expandafter\bbl@tempb
688       \fi}%
689     \expandafter\bbl@tempb\bbl@captionslist\today\@empty

```

```

690 \def\bbl@tempa##1{% elt for include list
691   \ifx##1\@empty\else
692     \bbl@csarg\in@{ensure@\language\expandafter}\expandafter{##1}%
693     \ifin\else
694       \bbl@tempb##1\@empty
695       \fi
696     \expandafter\bbl@tempa
697   \fi}%
698 \bbl@tempa#1\@empty}
699 \def\bbl@captionslist{%
700 \prefacename\refname\abstractname\bibname\chaptername\appendixname
701 \contentsname\listfigurename\listtablename\indexname\figurename
702 \tablename\partname\enclname\ccname\headtoname\pagename\seename
703 \alsoname\proofname\glossaryname}

```

9.3 Setting up language files

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

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

Another character that needs to have the correct category code during processing of language definition files is the equals sign, ‘=’, because it is sometimes used in constructions with the `\let` primitive. Therefore we store its current catcode and restore it later on. Now we check whether we should perhaps stop the processing of this file. To do this we first need to check whether the second argument that is passed to `\LdfInit` is a control sequence. We do that by looking at the first token after passing #2 through string. When it is equal to `\@backslashchar` we are dealing with a control sequence which we can compare with `\@undefined`.

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

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

Finally we check `\originalTeX`.

```

704 \bbl@trace{Macros for setting language files up}
705 \def\bbl@ldfinit{%
706   \let\bbl@screset\@empty
707   \let\BabelStrings\bbl@opt@string
708   \let\BabelOptions\@empty
709   \let\BabelLanguages\relax
710   \ifx\originalTeX\@undefined
711     \let\originalTeX\@empty
712   \else
713     \originalTeX
714   \fi}
715 \def\LdfInit#1#2{%
716   \chardef\atcatcode=\catcode`\@
717   \catcode`\@=11\relax
718   \chardef\eqcatcode=\catcode`\=
719   \catcode`\==12\relax
720   \expandafter\if\expandafter\@backslashchar
721     \expandafter\@car\string#2\@nil
722   \ifx#2\@undefined\else

```

```

723 \ldf@quit{#1}%
724 \fi
725 \else
726 \expandafter\ifx\csname#2\endcsname\relax\else
727 \ldf@quit{#1}%
728 \fi
729 \fi
730 \bbl@ldfinit}

```

`\ldf@quit` This macro interrupts the processing of a language definition file.

```

731 \def\ldf@quit#1{%
732 \expandafter\main@language\expandafter{#1}%
733 \catcode\@=\atcatcode \let\atcatcode\relax
734 \catcode\==\eqcatcode \let\eqcatcode\relax
735 \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.

```

736 \def\bbl@afterldf#1{%
737 \bbl@afterlang
738 \let\bbl@afterlang\relax
739 \let\BabelModifiers\relax
740 \let\bbl@screset\relax}%
741 \def\ldf@finish#1{%
742 \loadlocalcfg{#1}%
743 \bbl@afterldf{#1}%
744 \expandafter\main@language\expandafter{#1}%
745 \catcode\@=\atcatcode \let\atcatcode\relax
746 \catcode\==\eqcatcode \let\eqcatcode\relax}

```

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

```

747 \@onlypreamble\LdfInit
748 \@onlypreamble\ldf@quit
749 \@onlypreamble\ldf@finish

```

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

```

750 \def\main@language#1{%
751 \def\bbl@main@language{#1}%
752 \let\language\bbl@main@language
753 \bbl@id@assign
754 \bbl@patterns{\language}}

```

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

```

755 \def\bbl@beforestart{%
756 \bbl@usehooks{beforestart}}}%
757 \global\let\bbl@beforestart\relax}
758 \AtBeginDocument{%
759 \@nameuse{bbl@beforestart}%
760 \if@filesw

```

```

761 \immediate\write\@mainaux{\string\@nameuse{bbl@beforestart}}%
762 \fi
763 \expandafter\selectlanguage\expandafter{\bbl@main@language}%
764 \ifbbl@single % must go after the line above
765 \renewcommand\selectlanguage[1]{}%
766 \renewcommand\foreignlanguage[2]{#2}%
767 \global\let\babel@aux\@gobbletwo % Also as flag
768 \fi
769 \ifcase\bbl@engine\or\pagedir\bodydir\fi} % TODO - a better place

```

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

```

770 \def\select@language@x#1{%
771 \ifcase\bbl@select@type
772 \bbl@ifsamestring\languagename{#1}{\select@language{#1}}%
773 \else
774 \select@language{#1}%
775 \fi}

```

9.4 Shorthands

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

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

```

776 \bbl@trace{Shorhands}
777 \def\bbl@add@special#1{% 1:a macro like "\", \?, etc.
778 \bbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
779 \bbl@ifunset{@sanitize}{\bbl@add\@sanitize{\@makeother#1}}%
780 \ifx\nfss@catcodes\@undefined\else % TODO - same for above
781 \begingroup
782 \catcode`#1\active
783 \nfss@catcodes
784 \ifnum\catcode`#1=\active
785 \endgroup
786 \bbl@add\nfss@catcodes{\@makeother#1}%
787 \else
788 \endgroup
789 \fi
790 \fi}

```

`\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.

```

791 \def\bbl@remove@special#1{%
792 \begingroup
793 \def\x##1##2{\ifnum`#1=##2\noexpand\@empty
794 \else\noexpand##1\noexpand##2\fi}%
795 \def\do{\x\do}%
796 \def\@makeother{\x\@makeother}%
797 \edef\x{\endgroup
798 \def\noexpand\dospecials{\dospecials}%
799 \expandafter\ifx\csname @sanitize\endcsname\relax\else
800 \def\noexpand\@sanitize{\@sanitize}%
801 \fi}%
802 \x}

```

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

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

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

```
803 \def\bbl@active@def#1#2#3#4{%
804   \@namedef{#3#1}{%
805     \expandafter\ifx\csname#2@sh@#1@endcsname\relax
806       \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
807     \else
808       \bbl@afterfi\csname#2@sh@#1@endcsname
809     \fi}%

```

When there is also no current-level shorthand with an argument we will check whether there is a next-level defined shorthand for this active character.

```
810   \long\@namedef{#3@arg#1}##1{%
811     \expandafter\ifx\csname#2@sh@#1\string##1@endcsname\relax
812       \bbl@afterelse\csname#4#1@endcsname##1%
813     \else
814       \bbl@afterfi\csname#2@sh@#1\string##1@endcsname
815     \fi}}%

```

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

```
816 \def\@initiate@active@char#1#2#3{%
817   \bbl@ifunset{active@char\string#1}%
818   {\bbl@withactive
819     {\expandafter\@initiate@active@char\expandafter}#1\string#1#1}%
820   {}}

```

The very first thing to do is saving the original catcode and the original definition, even if not active, which is possible (undefined characters require a special treatment to avoid making them `\relax`).

```
821 \def\@initiate@active@char#1#2#3{%
822   \bbl@csarg\edef\oricat@#2{\catcode`#2=\the\catcode`#2\relax}%
823   \ifx#1\@undefined
824     \bbl@csarg\edef\oridef@#2{\let\noexpand#1\noexpand\@undefined}%
825   \else
826     \bbl@csarg\let\oridef@#2#1%
827     \bbl@csarg\edef\oridef@#2{%

```

```

828 \let\noexpand#1%
829 \expandafter\noexpand\csname bbl@oridef@@#2\endcsname}%
830 \fi

```

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

```

831 \ifx#1#3\relax
832 \expandafter\let\csname normal@char#2\endcsname#3%
833 \else
834 \bbl@info{Making #2 an active character}%
835 \ifnum\mathcode`#2=\ifodd\bbl@engine"1000000 \else"8000 \fi
836 \@namedef{normal@char#2}{%
837 \textormath{#3}{\csname bbl@oridef@@#2\endcsname}}%
838 \else
839 \@namedef{normal@char#2}{#3}%
840 \fi

```

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

```

841 \bbl@restoreactive{#2}%
842 \AtBeginDocument{%
843 \catcode`#2\active
844 \if@filesw
845 \immediate\write\@mainaux{\catcode`\string#2\active}%
846 \fi}%
847 \expandafter\bbl@add@special\csname#2\endcsname
848 \catcode`#2\active
849 \fi

```

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

```

850 \let\bbl@tempa\@firstoftwo
851 \if\string^#2%
852 \def\bbl@tempa{\noexpand\textormath}%
853 \else
854 \ifx\bbl@mathnormal\@undefined\else
855 \let\bbl@tempa\bbl@mathnormal
856 \fi
857 \fi
858 \expandafter\edef\csname active@char#2\endcsname{%
859 \bbl@tempa
860 {\noexpand\if@safe@actives
861 \noexpand\expandafter
862 \expandafter\noexpand\csname normal@char#2\endcsname
863 \noexpand\else
864 \noexpand\expandafter

```

```

865         \expandafter\noexpand\csname bbl@doactive#2\endcsname
866         \noexpand\fi}%
867     {\expandafter\noexpand\csname normal@char#2\endcsname}}%
868     \bbl@csarg\edef{doactive#2}{%
869         \expandafter\noexpand\csname user@active#2\endcsname}%

```

We now define the default values which the shorthand is set to when activated or deactivated. It is set to the deactivated form (globally), so that the character expands to

`\active@prefix <char> \normal@char <char>`

(where `\active@char <char>` is *one* control sequence!).

```

870     \bbl@csarg\edef{active@#2}{%
871         \noexpand\active@prefix\noexpand#1%
872         \expandafter\noexpand\csname active@char#2\endcsname}%
873     \bbl@csarg\edef{normal@#2}{%
874         \noexpand\active@prefix\noexpand#1%
875         \expandafter\noexpand\csname normal@char#2\endcsname}%
876     \expandafter\let\expandafter#1\csname bbl@normal@#2\endcsname

```

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

```

877     \bbl@active@def#2\user@group{user@active}{language@active}%
878     \bbl@active@def#2\language@group{language@active}{system@active}%
879     \bbl@active@def#2\system@group{system@active}{normal@char}%

```

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

```

880     \expandafter\edef\csname\user@group @sh@#2@@\endcsname
881     {\expandafter\noexpand\csname normal@char#2\endcsname}%
882     \expandafter\edef\csname\user@group @sh@#2@\string\protect\endcsname
883     {\expandafter\noexpand\csname user@active#2\endcsname}%

```

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

```

884     \if\string'#2%
885         \let\prim@s\bbl@prim@s
886         \let\active@math@prime#1%
887     \fi
888     \bbl@usehooks{initiateactive}{\#1}{\#2}{\#3}}

```

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

```

889 <<{*More package options}>> ≡
890 \DeclareOption{math=active}{}
891 \DeclareOption{math=normal}{\def\bbl@mathnormal{\noexpand\textormath}}
892 <</More package options>>

```

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

```

893 \ifpackagewith{babel}{KeepShorthandsActive}%

```

```

894 {\let\bbl@restoreactive\@gobble}%
895 {\def\bbl@restoreactive#1{%
896   \bbl@exp{%
897     \\\AfterBabelLanguage\\CurrentOption
898     {\catcode`#1=\the\catcode`#1\relax}%
899     \\\AtEndOfPackage
900     {\catcode`#1=\the\catcode`#1\relax}}}%
901   \AtEndOfPackage{\let\bbl@restoreactive\@gobble}}

```

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

```

902 \def\bbl@sh@select#1#2{%
903   \expandafter\ifx\csname#1@sh@#2@sel\endcsname\relax
904     \bbl@afterelse\bbl@scndcs
905   \else
906     \bbl@afterfi\csname#1@sh@#2@sel\endcsname
907   \fi}

```

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

```

908 \begingroup
909 \bbl@ifunset{ifincsname}%
910 {\gdef\active@prefix#1{%
911   \ifx\protect\@typeset@protect
912   \else
913     \ifx\protect\@unexpandable@protect
914       \noexpand#1%
915     \else
916       \protect#1%
917     \fi
918     \expandafter\@gobble
919   \fi}}
920 {\gdef\active@prefix#1{%
921   \ifincsname
922     \string#1%
923     \expandafter\@gobble
924   \else
925     \ifx\protect\@typeset@protect
926     \else
927       \ifx\protect\@unexpandable@protect
928         \noexpand#1%
929       \else
930         \protect#1%
931       \fi
932       \expandafter\expandafter\expandafter\@gobble
933     \fi
934   \fi}}
935 \endgroup

```


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

```

936 \newif\if@safe@actives
937 \@safe@activesfalse

```

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

```

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

```

`\bbl@activate` 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`.

```

939 \def\bbl@activate#1{%
940   \bbl@withactive{\expandafter\let\expandafter}\#1%
941   \csname bbl@active@\string#1\endcsname}
942 \def\bbl@deactivate#1{%
943   \bbl@withactive{\expandafter\let\expandafter}\#1%
944   \csname bbl@normal@\string#1\endcsname}

```

`\bbl@firstcs` These macros have two arguments. They use one of their arguments to build a control sequence from.

```

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

```

`\declare@shorthand` 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.

```

947 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil}
948 \def\@decl@short#1#2#3\@nil#4{%
949   \def\bbl@tempa{#3}%
950   \ifx\bbl@tempa\@empty
951     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs
952     \bbl@ifunset{#1@sh@\string#2@}{}%
953     {\def\bbl@tempa{#4}%
954       \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa
955       \else
956         \bbl@info
957           {Redefining #1 shorthand \string#2\\%
958            in language \CurrentOption}%
959       \fi}%
960     \@namedef{#1@sh@\string#2@}{#4}%
961   \else
962     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs
963     \bbl@ifunset{#1@sh@\string#2@\string#3@}{}%
964     {\def\bbl@tempa{#4}%
965       \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
966       \else
967         \bbl@info
968           {Redefining #1 shorthand \string#2\string#3\\%
969            in language \CurrentOption}%

```

```

970      \fi}%
971      \@namedef{#1@sh@string#2@string#3@}{#4}%
972      \fi}

\textormath Some of the shorthands that will be declared by the language definition files have to be
usable in both text and mathmode. To achieve this the helper macro \textormath is
provided.

973 \def\textormath{%
974   \ifmmode
975     \expandafter\@secondoftwo
976   \else
977     \expandafter\@firstoftwo
978   \fi}

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

979 \def\user@group{user}
980 \def\language@group{english}
981 \def\system@group{system}

\useshorthands This is the user level command to tell LATEX that user level shorthands will be used in the
document. It takes one argument, the character that starts a shorthand. First note that this
is user level, and then initialize and activate the character for use as a shorthand character
(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.

982 \def\useshorthands{%
983   \@ifstar\bbl@usesh@s{\bbl@usesh@x{}}
984   \def\bbl@usesh@s#1{%
985     \bbl@usesh@x
986     {\AddBabelHook{babel-sh-}\string#1}{afterextras}{\bbl@activate{#1}}}%
987     {#1}}
988 \def\bbl@usesh@x#1#2{%
989   \bbl@ifshorthand{#2}%
990     {\def\user@group{user}%
991       \initiate@active@char{#2}%
992       #1%
993       \bbl@activate{#2}}%
994     {\bbl@error
995       {Cannot declare a shorthand turned off (\string#2)}
996       {Sorry, but you cannot use shorthands which have been\\%
997         turned off in the package options}}}

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

998 \def\user@language@group{user@\language@group}
999 \def\bbl@set@user@generic#1#2{%
1000   \bbl@ifunset{user@generic@active#1}%
1001     {\bbl@active@def#1\user@language@group{user@active}{user@generic@active}%
1002       \bbl@active@def#1\user@group{user@generic@active}{\language@active}%
1003       \expandafter\edef\csname#2@sh@#1@@\endcsname{%
1004         \expandafter\noexpand\csname normal@char#1\endcsname}%
1005       \expandafter\edef\csname#2@sh@#1@\string\protect\endcsname{%
1006         \expandafter\noexpand\csname user@active#1\endcsname}}%

```

```

1007 \@empty}
1008 \newcommand\defineshorthand[3][user]{%
1009 \edef\bbl@tempa{\zap@space#1 \@empty}%
1010 \bbl@for\bbl@tempb\bbl@tempa{%
1011 \if*\expandafter\@car\bbl@tempb\@nil
1012 \edef\bbl@tempb{user\expandafter\@gobble\bbl@tempb}%
1013 \@expandtwoargs
1014 \bbl@set@user@generic{\expandafter\string\@car#2\@nil}\bbl@tempb
1015 \fi
1016 \declare@shorthand{\bbl@tempb}{#2}{#3}}

```

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

```

1017 \def\languageshorthands#1{\def\language@group{#1}}

```

`\aliasshorthand` First the new shorthand needs to be initialized,

```

1018 \def\aliasshorthand#1#2{%
1019 \bbl@ifshorthand{#2}%
1020 {\expandafter\ifx\csname active@char\string#2\endcsname\relax
1021 \ifx\document\@notprerr
1022 \@notshorthand{#2}%
1023 \else
1024 \initiate@active@char{#2}%

```

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

```

1025 \expandafter\let\csname active@char\string#2\expandafter\endcsname
1026 \csname active@char\string#1\endcsname
1027 \expandafter\let\csname normal@char\string#2\expandafter\endcsname
1028 \csname normal@char\string#1\endcsname
1029 \bbl@activate{#2}%
1030 \fi
1031 \fi}%
1032 {\bbl@error
1033 {Cannot declare a shorthand turned off (\string#2)}
1034 {Sorry, but you cannot use shorthands which have been\\%
1035 turned off in the package options}}}

```

`\@notshorthand`

```

1036 \def\@notshorthand#1{%
1037 \bbl@error{%
1038 The character '\string #1' should be made a shorthand character;\\%
1039 add the command \string\usesshorthands\string{#1\string} to
1040 the preamble.\\%
1041 I will ignore your instruction}%
1042 {You may proceed, but expect unexpected results}}

```

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

```

1043 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
1044 \DeclareRobustCommand*\shorthandoff{%
1045 \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
1046 \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.

```

1047 \def\bbl@switch@sh#1#2{%
1048   \ifx#2\@nnil\else
1049     \bbl@ifunset{\bbl@active@\string#2}%
1050     {\bbl@error
1051       {I cannot switch '\string#2' on or off--not a shorthand}%
1052       {This character is not a shorthand. Maybe you made\\%
1053         a typing mistake? I will ignore your instruction}}}%
1054     {\ifcase#1%
1055       \catcode'#212\relax
1056       \or
1057       \catcode'#2\active
1058       \or
1059       \csname bbl@oricat@\string#2\endcsname
1060       \csname bbl@oridef@\string#2\endcsname
1061       \fi}%
1062     \bbl@afterfi\bbl@switch@sh#1%
1063   \fi}

```

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

```

1064 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
1065 \def\bbl@putsh#1{%
1066   \bbl@ifunset{\bbl@active@\string#1}%
1067   {\bbl@putsh@i#1\@empty\@nnil}%
1068   {\csname bbl@active@\string#1\endcsname}}
1069 \def\bbl@putsh@i#1#2\@nnil{%
1070   \csname\language @sh@\string#1@%
1071     \ifx\@empty#2\else\string#2@\fi\endcsname}
1072 \ifx\bbl@opt@shorthands\@nnil\else
1073   \let\bbl@s@initiate@active@char\initiate@active@char
1074   \def\initiate@active@char#1{%
1075     \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
1076   \let\bbl@s@switch@sh\bbl@switch@sh
1077   \def\bbl@switch@sh#1#2{%
1078     \ifx#2\@nnil\else
1079       \bbl@afterfi
1080       \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
1081       \fi}
1082   \let\bbl@s@activate\bbl@activate
1083   \def\bbl@activate#1{%
1084     \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
1085   \let\bbl@s@deactivate\bbl@deactivate
1086   \def\bbl@deactivate#1{%
1087     \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
1088   \fi

```

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

```

1089 \newcommand\ifbabelshorthand[3]{\bbl@ifunset{\bbl@active@\string#1}{#3}{#2}}

```

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

```

1090 \def\bbl@prim@s{%
1091   \prime\futurelet\@let@token\bbl@pr@m@s}
1092 \def\bbl@if@primes#1#2{%
1093   \ifx#1\@let@token
1094     \expandafter\@firstoftwo
1095   \else\ifx#2\@let@token
1096     \bbl@afterelse\expandafter\@firstoftwo
1097   \else
1098     \bbl@afterfi\expandafter\@secondoftwo
1099   \fi\fi}
1100 \begingroup
1101   \catcode`\^=7 \catcode`\*=\active \lccode`\*='^
1102   \catcode`\'=12 \catcode`\"=\active \lccode`\"=' '
1103   \lowercase{%
1104     \gdef\bbl@pr@m@s{%
1105       \bbl@if@primes""%
1106       \pr@@@s
1107       {\bbl@if@primes*^ \pr@@@t\egroup}}}
1108 \endgroup

```

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

```

1109 \initiate@active@char{~}
1110 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
1111 \bbl@activate{~}

```

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

```

1112 \expandafter\def\csname OT1dqpos\endcsname{127}
1113 \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

```

1114 \ifx\f@encoding\@undefined
1115   \def\f@encoding{OT1}
1116 \fi

```

9.5 Language attributes

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

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

```

1117 \bbl@trace{Language attributes}
1118 \newcommand\languageattribute[2]{%

```

```

1119 \def\bbl@tempc{#1}%
1120 \bbl@fixname\bbl@tempc
1121 \bbl@iflanguage\bbl@tempc{%
1122   \bbl@vforeach{#2}{%

```

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

```

1123   \ifx\bbl@known@attribs\undefined
1124     \in@false
1125   \else

```

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

```

1126     \bbl@xin@{,\bbl@tempc-##1,}{,\bbl@known@attribs,}%
1127   \fi

```

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

```

1128   \ifin@
1129     \bbl@warning{%
1130       You have more than once selected the attribute '##1'\%
1131       for language #1. Reported}%
1132   \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.

```

1133     \bbl@exp{%
1134       \\bbl@add@list\\bbl@known@attribs{\bbl@tempc-##1}}%
1135     \edef\bbl@tempa{\bbl@tempc-##1}%
1136     \expandafter\bbl@ifknown@trib\expandafter{\bbl@tempa}\bbl@attributes%
1137     {\csname\bbl@tempc @attr@##1\endcsname}%
1138     {\@attrerrr{\bbl@tempc}{##1}}%
1139   \fi}}

```

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

```

1140 \@onlypreamble\languageattribute

```

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

```

1141 \newcommand*{\@attrerrr}[2]{%
1142   \bbl@error
1143   {The attribute #2 is unknown for language #1.}%
1144   {Your command will be ignored, type <return> to proceed}}

```

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

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

```

1145 \def\bbl@declare@ttribute#1#2#3{%
1146   \bbl@xin@{,#2,}{,\BabelModifiers,}%
1147   \ifin@
1148     \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
1149   \fi
1150   \bbl@add@list\bbl@attributes{#1-#2}%
1151   \expandafter\def\csname#1@attr@#2\endcsname{#3}}

```

`\bbl@ifattributeset` This internal macro has 4 arguments. It can be used to interpret T_EX 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.

```
1152 \def\bbl@ifattributeset#1#2#3#4{%
```

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

```
1153 \ifx\bbl@known@attribs\@undefined
```

```
1154 \in@false
```

```
1155 \else
```

The we need to check the list of known attributes.

```
1156 \bbl@xin@{,#1-#2,}{,\bbl@known@attribs,}%
```

```
1157 \fi
```

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

```
1158 \ifin@
```

```
1159 \bbl@afterelse#3%
```

```
1160 \else
```

```
1161 \bbl@afterfi#4%
```

```
1162 \fi
```

```
1163 }
```

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

```
1164 \def\bbl@ifknown@ttrib#1#2{%
```

We first assume the attribute is unknown.

```
1165 \let\bbl@tempa\@secondoftwo
```

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

```
1166 \bbl@loopx\bbl@tempb{#2}{%
```

```
1167 \expandafter\in@\expandafter{\expandafter,\bbl@tempb,}{,#1,}%
```

```
1168 \ifin@
```

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

```
1169 \let\bbl@tempa\@firstoftwo
```

```
1170 \else
```

```
1171 \fi}%
```

Finally we execute `\bbl@tempa`.

```
1172 \bbl@tempa
```

```
1173 }
```

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

```
1174 \def\bbl@clear@ttribs{%
```

```
1175 \ifx\bbl@attributes\@undefined\else
```

```
1176 \bbl@loopx\bbl@tempa{\bbl@attributes}{%
```

```
1177 \expandafter\bbl@clear@ttrib\bbl@tempa.
```

```
1178 }%
```

```
1179 \let\bbl@attributes\@undefined
```

```
1180 \fi}
```

```
1181 \def\bbl@clear@ttrib#1-#2.{%
```

```
1182 \expandafter\let\csname#1@attr@#2\endcsname\@undefined}
```

```
1183 \AtBeginDocument{\bbl@clear@ttribs}
```

9.6 Support for saving macro definitions

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

`\babel@savecnt` The initialization of a new save cycle: reset the counter to zero.

```
\babel@beginsave 1184 \bbl@trace{Macros for saving definitions}
                  1185 \def\babel@beginsave{\babel@savecnt\z@}
```

Before it's forgotten, allocate the counter and initialize all.

```
1186 \newcount\babel@savecnt
1187 \babel@beginsave
```

`\babel@save` The macro `\babel@save⟨csname⟩` saves the current meaning of the control sequence `⟨csname⟩` to `\originalTeX`³⁰. To do this, we let the current meaning to a temporary control sequence, the restore commands are appended to `\originalTeX` and the counter is incremented.

```
1188 \def\babel@save#1{%
1189   \expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax
1190   \toks@\expandafter{\originalTeX\let#1=}%
1191   \bbl@exp{%
1192     \def\originalTeX{\the\toks@<\babel@number\babel@savecnt>\relax}}%
1193   \advance\babel@savecnt\@ne}
```

`\babel@savevariable` The macro `\babel@savevariable⟨variable⟩` saves the value of the variable. `⟨variable⟩` can be anything allowed after the `\the` primitive.

```
1194 \def\babel@savevariable#1{%
1195   \toks@\expandafter{\originalTeX #1=}%
1196   \bbl@exp{\def\originalTeX{\the\toks@the#1\relax}}}
```

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

```
1197 \def\bbl@frenchspacing{%
1198   \ifnum\the\sffcode\`.\=@m
1199     \let\bbl@nonfrenchspacing\relax
1200   \else
1201     \frenchspacing
1202     \let\bbl@nonfrenchspacing\nonfrenchspacing
1203   \fi}
1204 \let\bbl@nonfrenchspacing\nonfrenchspacing
```

9.7 Short tags

`\babeltags` This macro is straightforward. After zapping spaces, we loop over the list and define the macros `\text⟨tag⟩` and `\⟨tag⟩`. Definitions are first expanded so that they don't contain `\csname` but the actual macro.

```
1205 \bbl@trace{Short tags}
1206 \def\babeltags#1{%
```

³⁰`\originalTeX` has to be expandable, i.e. you shouldn't let it to `\relax`.


```

1207 \edef\bbl@tempa{\zap@space#1 \@empty}%
1208 \def\bbl@tempb##1=##2\@@{%
1209   \edef\bbl@tempc{%
1210     \noexpand\newcommand
1211     \expandafter\noexpand\csname ##1\endcsname{%
1212       \noexpand\protect
1213       \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
1214     \noexpand\newcommand
1215     \expandafter\noexpand\csname text##1\endcsname{%
1216       \noexpand\foreignlanguage{##2}}
1217   \bbl@tempc}%
1218 \bbl@for\bbl@tempa\bbl@tempa{%
1219   \expandafter\bbl@tempb\bbl@tempa\@@}}

```

9.8 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.

```

1220 \bbl@trace{Hyphens}
1221 \@onlypreamble\babelhyphenation
1222 \AtEndOfPackage{%
1223   \newcommand\babelhyphenation[2][\@empty]{%
1224     \ifx\bbl@hyphenation@relax
1225       \let\bbl@hyphenation@\@empty
1226     \fi
1227     \ifx\bbl@hyphlist\@empty\else
1228       \bbl@warning{%
1229         You must not intermingle \string\selectlanguage\space and\%
1230         \string\babelhyphenation\space or some exceptions will not\%
1231         be taken into account. Reported}%
1232     \fi
1233     \ifx\@empty#1%
1234       \protected@edef\bbl@hyphenation@{\bbl@hyphenation@\space#2}%
1235     \else
1236       \bbl@vforeach{#1}{%
1237         \def\bbl@tempa{##1}%
1238         \bbl@fixname\bbl@tempa
1239         \bbl@iflanguage\bbl@tempa{%
1240           \bbl@csarg\protected@edef{hyphenation@\bbl@tempa}{%
1241             \bbl@ifunset{bbl@hyphenation@\bbl@tempa}%
1242             \@empty
1243             {\csname bbl@hyphenation@\bbl@tempa\endcsname\space}%
1244             #2}}}%
1245     \fi}}

```

`\bbl@allowhyphens` This macro makes hyphenation possible. Basically its definition is nothing more than `\nobreak \hskip Opt plus Opt`³¹.

```

1246 \def\bbl@allowhyphens{\ifvmode\else\nobreak\hskip\zskip\fi}
1247 \def\bbl@t@one{T1}
1248 \def\allowhyphens{\ifx\cf@encoding\bbl@t@one\else\bbl@allowhyphens\fi}

```

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

³¹ \TeX begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.

```

1249 \newcommand\babelnullhyphen{\char\hyphenchar\font}
1250 \def\babelhyphen{\active@prefix\babelhyphen\bb1@hyphen}
1251 \def\bb1@hyphen{%
1252   \@ifstar{\bb1@hyphen@i @}{\bb1@hyphen@i\@empty}}
1253 \def\bb1@hyphen@i#1#2{%
1254   \bb1@ifunset{\bb1@hy@#1#2\@empty}%
1255   {\csname bb1@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%
1256   {\csname bb1@hy@#1#2\@empty\endcsname}}

```

The following two commands are used to wrap the “hyphen” and set the behavior of the rest of the word – the version with a single @ is used when further hyphenation is allowed, while that with @@ if no more 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.

```

1257 \def\bb1@usehyphen#1{%
1258   \leavevmode
1259   \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
1260   \nobreak\hskip\z@skip}
1261 \def\bb1@@usehyphen#1{%
1262   \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

The following macro inserts the hyphen char.

```

1263 \def\bb1@hyphenchar{%
1264   \ifnum\hyphenchar\font=\m@ne
1265     \babelnullhyphen
1266   \else
1267     \char\hyphenchar\font
1268   \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 \bb1@hy@nobreak is redundant.

```

1269 \def\bb1@hy@soft{\bb1@usehyphen{\discretionary{\bb1@hyphenchar}{}}{}}
1270 \def\bb1@hy@soft{\bb1@usehyphen{\discretionary{\bb1@hyphenchar}{}}{}}
1271 \def\bb1@hy@hard{\bb1@usehyphen\bb1@hyphenchar}
1272 \def\bb1@hy@@hard{\bb1@usehyphen\bb1@hyphenchar}
1273 \def\bb1@hy@nobreak{\bb1@usehyphen{\mbox{\bb1@hyphenchar}}}
1274 \def\bb1@hy@nobreak{\mbox{\bb1@hyphenchar}}
1275 \def\bb1@hy@repeat{%
1276   \bb1@usehyphen{%
1277     \discretionary{\bb1@hyphenchar}{\bb1@hyphenchar}{\bb1@hyphenchar}}}
1278 \def\bb1@hy@@repeat{%
1279   \bb1@usehyphen{%
1280     \discretionary{\bb1@hyphenchar}{\bb1@hyphenchar}{\bb1@hyphenchar}}}
1281 \def\bb1@hy@empty{\hskip\z@skip}
1282 \def\bb1@hy@@empty{\discretionary{}{}{}}

```

\bb1@disc For some languages the macro \bb1@disc is used to ease the insertion of discretionaries for letters that behave ‘abnormally’ at a breakpoint.

```

1283 \def\bb1@disc#1#2{\nobreak\discretionary{#2-}{}{#1}\bb1@allowhyphens}

```

9.9 Multiencoding strings

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

Tools But first, a couple of tools. The first one makes global a local variable. This is not the best solution, but it works.

```

1284 \bbl@trace{Multiencoding strings}
1285 \def\bbl@tglobal#1{\global\let#1#1}
1286 \def\bbl@recatcode#1{%
1287   \@tempcnta="7F
1288   \def\bbl@tempa{%
1289     \ifnum\@tempcnta>"FF\else
1290       \catcode\@tempcnta=#1\relax
1291       \advance\@tempcnta\@ne
1292       \expandafter\bbl@tempa
1293     \fi}%
1294   \bbl@tempa}

```

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

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

and starts over (and similarly when lowercasing).

```

1295 \@ifpackagewith{babel}{nocase}%
1296   {\let\bbl@patchuclc\relax}%
1297   {\def\bbl@patchuclc{%
1298     \global\let\bbl@patchuclc\relax
1299     \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}%
1300     \gdef\bbl@uclc##1{%
1301       \let\bbl@encoded\bbl@encoded@uclc
1302       \bbl@ifunset{\language @bbl@uclc}% and resumes it
1303       {##1}%
1304       {\let\bbl@tempa##1\relax % Used by LANG@bbl@uclc
1305        \csname\language @bbl@uclc\endcsname}%
1306       {\bbl@tolower\@empty}{\bbl@toupper\@empty}}%
1307     \gdef\bbl@tolower{\csname\language @bbl@lc\endcsname}%
1308     \gdef\bbl@toupper{\csname\language @bbl@uc\endcsname}}%
1309 <<(*More package options)>> ≡
1310 \DeclareOption{nocase}{}
1311 <</More package options>>

```

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

```

1312 <<(*More package options)>> ≡
1313 \let\bbl@opt@strings\@nnil % accept strings=value
1314 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1315 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1316 \def\BabelStringsDefault{generic}
1317 <</More package options>>

```

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.

```
1318 \@onlypreamble\StartBabelCommands
```

```

1319 \def\StartBabelCommands{%
1320   \begingroup
1321   \bbl@recatcode{11}%
1322   <<Macros local to BabelCommands>>
1323   \def\bbl@provstring##1##2{%
1324     \providecommand##1{##2}%
1325     \bbl@toglobal##1}%
1326   \global\let\bbl@scafter\@empty
1327   \let\StartBabelCommands\bbl@startcmds
1328   \ifx\BabelLanguages\relax
1329     \let\BabelLanguages\CurrentOption
1330   \fi
1331   \begingroup
1332   \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
1333   \StartBabelCommands}
1334 \def\bbl@startcmds{%
1335   \ifx\bbl@screset\@nnil\else
1336     \bbl@usehooks{stopcommands}{}%
1337   \fi
1338   \endgroup
1339   \begingroup
1340   \@ifstar
1341     {\ifx\bbl@opt@strings\@nnil
1342       \let\bbl@opt@strings\BabelStringsDefault
1343     \fi
1344     \bbl@startcmds@i}%
1345   \bbl@startcmds@i}
1346 \def\bbl@startcmds@i#1#2{%
1347   \edef\bbl@L{\zap@space#1 \@empty}%
1348   \edef\bbl@G{\zap@space#2 \@empty}%
1349   \bbl@startcmds@ii}

```

Parse the encoding info to get the label, input, and font parts.

Select the behavior of \SetString. There are two main cases, depending of if there is an optional argument: without it and strings=encoded, strings are defined always; otherwise, they are set only if they are still undefined (ie, fallback values). With labelled blocks and strings=encoded, define the strings, but with another value, define strings only if the current label or font encoding is the value of strings; otherwise (ie, no strings or a block whose label is not in strings=) do nothing.

We presume the current block is not loaded, and therefore set (above) a couple of default values to gobble the arguments. Then, these macros are redefined if necessary according to several parameters.

```

1350 \newcommand\bbl@startcmds@ii[1][\@empty]{%
1351   \let\SetString\@gobbletwo
1352   \let\bbl@stringdef\@gobbletwo
1353   \let\AfterBabelCommands\@gobble
1354   \ifx\@empty#1%
1355     \def\bbl@sc@label{generic}%
1356     \def\bbl@encstring##1##2{%
1357       \ProvideTextCommandDefault##1{##2}%
1358       \bbl@toglobal##1%
1359       \expandafter\bbl@toglobal\csname\string?\string##1\endcsname}%
1360     \let\bbl@sctest\in@true
1361   \else
1362     \let\bbl@sc@charset\space % <- zapped below
1363     \let\bbl@sc@fontenc\space % <- " "
1364     \def\bbl@tempa##1=##2\@nil{%
1365       \bbl@csarg\edef{sc\zap@space##1 \@empty}{##2 }}%

```

```

1366 \bbl@vforeach{label=#1}{\bbl@tempa##1\@nil}%
1367 \def\bbl@tempa##1 ##2{% space -> comma
1368   ##1%
1369   \ifx\@empty##2\else\ifx,##1,\else,\fi\bbl@afterfi\bbl@tempa##2\fi}%
1370 \edef\bbl@sc@fontenc{\expandafter\bbl@tempa\bbl@sc@fontenc\@empty}%
1371 \edef\bbl@sc@label{\expandafter\zap@space\bbl@sc@label\@empty}%
1372 \edef\bbl@sc@charset{\expandafter\zap@space\bbl@sc@charset\@empty}%
1373 \def\bbl@encstring##1##2{%
1374   \bbl@foreach\bbl@sc@fontenc{%
1375     \bbl@ifunset{T####1}%
1376     {}%
1377     {\ProvideTextCommand##1{####1}{##2}%
1378     \bbl@tglobal##1%
1379     \expandafter
1380     \bbl@tglobal\csname####1\string##1\endcsname}}}%
1381 \def\bbl@sctest{%
1382   \bbl@xin@{\,\bbl@opt@strings,}{,\,\bbl@sc@label,\bbl@sc@fontenc,}%
1383 \fi
1384 \ifx\bbl@opt@strings\@nnil % ie, no strings key -> defaults
1385 \else\ifx\bbl@opt@strings\relax % ie, strings=encoded
1386   \let\AfterBabelCommands\bbl@aftercmds
1387   \let\SetString\bbl@setstring
1388   \let\bbl@stringdef\bbl@encstring
1389 \else % ie, strings=value
1390 \bbl@sctest
1391 \ifin@
1392   \let\AfterBabelCommands\bbl@aftercmds
1393   \let\SetString\bbl@setstring
1394   \let\bbl@stringdef\bbl@provstring
1395 \fi\fi\fi
1396 \bbl@scswitch
1397 \ifx\bbl@G\@empty
1398   \def\SetString##1##2{%
1399     \bbl@error{Missing group for string \string##1}%
1400     {You must assign strings to some category, typically\\%
1401     captions or extras, but you set none}}%
1402 \fi
1403 \ifx\@empty#1%
1404   \bbl@usehooks{defaultcommands}{}%
1405 \else
1406   \@expandtwoargs
1407   \bbl@usehooks{encodedcommands}{\bbl@sc@charset}{\bbl@sc@fontenc}}%
1408 \fi}

```

There are two versions of `\bbl@scswitch`. The first version is used when `ldfs` are read, and it makes sure `\langle group \rangle \langle language \rangle` is reset, but only once (`\bbl@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 `\date \langle language \rangle` is defined (after `babel` has been loaded). There are also two version of `\bbl@forlang`. The first one skips the current iteration if the language is not in `\BabelLanguages` (used in `ldfs`), and the second one skips undefined languages (after `babel` has been loaded).

```

1409 \def\bbl@forlang#1##2{%
1410   \bbl@for#1\bbl@L{%
1411     \bbl@xin@{,##1,}{,\,\BabelLanguages,}%
1412     \ifin@##2\relax\fi}}
1413 \def\bbl@scswitch{%
1414   \bbl@forlang\bbl@tempa{%

```

```

1415 \ifx\bb1@G\@empty\else
1416 \ifx\SetString\@gobbletwo\else
1417 \edef\bb1@GL{\bb1@G\bb1@tempa}%
1418 \bb1@xin@{\bb1@GL,}{,\bb1@screset,}%
1419 \ifin@else
1420 \global\expandafter\let\csname\bb1@GL\endcsname\@undefined
1421 \xdef\bb1@screset{\bb1@screset,\bb1@GL}%
1422 \fi
1423 \fi
1424 \fi}}
1425 \AtEndOfPackage{%
1426 \def\bb1@forlang#1#2{\bb1@for#1\bb1@L{\bb1@ifunset{date#1}{\#2}}}%
1427 \let\bb1@scswitch\relax}
1428 \@onlypreamble\EndBabelCommands
1429 \def\EndBabelCommands{%
1430 \bb1@usehooks{stopcommands}{}%
1431 \endgroup
1432 \endgroup
1433 \bb1@scafter}

```

Now we define commands to be used inside \StartBabelCommands.

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

```

1434 \def\bb1@setstring#1#2{%
1435 \bb1@forlang\bb1@tempa{%
1436 \edef\bb1@LC{\bb1@tempa\bb1@stripslash#1}%
1437 \bb1@ifunset{\bb1@LC}% eg, \germanchaptername
1438 {\global\expandafter % TODO - con \bb1@exp ?
1439 \bb1@add\csname\bb1@G\bb1@tempa\expandafter\endcsname\expandafter
1440 {\expandafter\bb1@scset\expandafter#1\csname\bb1@LC\endcsname}}}%
1441 {}}%
1442 \def\BabelString{\#2}%
1443 \bb1@usehooks{stringprocess}{}%
1444 \expandafter\bb1@stringdef
1445 \csname\bb1@LC\expandafter\endcsname\expandafter{\BabelString}}

```

Now, some additional stuff to be used when encoded strings are used. Captions then include \bb1@encoded for string to be expanded in case transformations. It is \relax by default, but in \MakeUppercase and \MakeLowercase its value is a modified expandable \@changed@cmd.

```

1446 \ifx\bb1@opt@strings\relax
1447 \def\bb1@scset#1#2{\def#1{\bb1@encoded#2}}
1448 \bb1@patchuclc
1449 \let\bb1@encoded\relax
1450 \def\bb1@encoded@uclc#1{%
1451 \@inmathwarn#1%
1452 \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
1453 \expandafter\ifx\csname ?\string#1\endcsname\relax
1454 \TextSymbolUnavailable#1%
1455 \else
1456 \csname ?\string#1\endcsname
1457 \fi
1458 \else
1459 \csname\cf@encoding\string#1\endcsname

```

```

1460   \fi}
1461 \else
1462   \def\bbl@scset#1#2{\def#1{#2}}
1463 \fi

```

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

```

1464 <<*Macros local to BabelCommands>> ≡
1465 \def\SetStringLoop##1##2{%
1466   \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
1467   \count@\z@
1468   \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
1469     \advance\count@\@ne
1470     \toks@\expandafter{\bbl@tempa}%
1471     \bbl@exp{%
1472       \\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
1473       \count@=\the\count@\relax}}}%
1474 <</Macros local to BabelCommands>>

```

Delaying code Now the definition of `\AfterBabelCommands` when it is activated.

```

1475 \def\bbl@aftercmds#1{%
1476   \toks@\expandafter{\bbl@scafter#1}%
1477   \xdef\bbl@scafter{\the\toks@}}

```

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

```

1478 <<*Macros local to BabelCommands>> ≡
1479 \newcommand\SetCase[3][]{%
1480   \bbl@patchuclc
1481   \bbl@forlang\bbl@tempa{%
1482     \expandafter\bbl@encstring
1483     \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
1484     \expandafter\bbl@encstring
1485     \csname\bbl@tempa @bbl@uc\endcsname{##2}%
1486     \expandafter\bbl@encstring
1487     \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
1488 <</Macros local to BabelCommands>>

```

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.

```

1489 <<*Macros local to BabelCommands>> ≡
1490 \newcommand\SetHyphenMap[1]{%
1491   \bbl@forlang\bbl@tempa{%
1492     \expandafter\bbl@stringdef
1493     \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
1494 <</Macros local to BabelCommands>>

```

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

```

1495 \newcommand\BabelLower[2]{% one to one.
1496   \ifnum\lccode#1=#2\else
1497     \babel@savevariable{\lccode#1}%
1498     \lccode#1=#2\relax
1499   \fi}

```

```

1500 \newcommand\BabelLowerMM[4]{% many-to-many
1501   \@tempcnta=#1\relax
1502   \@tempcntb=#4\relax
1503   \def\bbl@tempa{%
1504     \ifnum\@tempcnta>#2\else
1505       \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
1506       \advance\@tempcnta#3\relax
1507       \advance\@tempcntb#3\relax
1508       \expandafter\bbl@tempa
1509     \fi}%
1510   \bbl@tempa}
1511 \newcommand\BabelLowerMO[4]{% many-to-one
1512   \@tempcnta=#1\relax
1513   \def\bbl@tempa{%
1514     \ifnum\@tempcnta>#2\else
1515       \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
1516       \advance\@tempcnta#3
1517       \expandafter\bbl@tempa
1518     \fi}%
1519   \bbl@tempa}

```

The following package options control the behavior of hyphenation mapping.

```

1520 <<{*More package options}> ≡
1521 \DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
1522 \DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\@ne}
1523 \DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
1524 \DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@@}
1525 \DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}
1526 <</More package options>>

```

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

```

1527 \AtEndOfPackage{%
1528   \ifx\bbl@opt@hyphenmap\undefined
1529     \bbl@xin@{,}{\bbl@language@opts}%
1530     \chardef\bbl@opt@hyphenmap\ifin4\else\@ne\fi
1531   \fi}

```

9.10 Macros common to a number of languages

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

```

1532 \bbl@trace{Macros related to glyphs}
1533 \def\set@low@box#1{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}%
1534   \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@%
1535   \setbox\z@\hbox{\lower\dimen\z@ \box\z@}\ht\z@\ht\tw@ \dp\z@\dp\tw@}

```

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

```

1536 \def\save@sf@q#1{\leavevmode
1537   \begingroup
1538     \edef\@SF{\spacefactor\the\spacefactor}#1\@SF
1539   \endgroup}

```

9.11 Making glyphs available

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

9.11.1 Quotation marks

`\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.

```
1540 \ProvideTextCommand{\quotedblbase}{OT1}{%  
1541   \save@sf@q{\set@low@box{\textquotedblright\}%  
1542     \box\z@\kern-.04em\bb1@allowhyphens}}
```

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

```
1543 \ProvideTextCommandDefault{\quotedblbase}{%  
1544   \UseTextSymbol{OT1}{\quotedblbase}}
```

`\quotesinglbase` We also need the single quote character at the baseline.

```
1545 \ProvideTextCommand{\quotesinglbase}{OT1}{%  
1546   \save@sf@q{\set@low@box{\textquoteright\}%  
1547     \box\z@\kern-.04em\bb1@allowhyphens}}
```

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

```
1548 \ProvideTextCommandDefault{\quotesinglbase}{%  
1549   \UseTextSymbol{OT1}{\quotesinglbase}}
```

`\guillemotleft` The guillemet characters are not available in OT1 encoding. They are faked.

```
\guillemotright 1550 \ProvideTextCommand{\guillemotleft}{OT1}{%  
1551   \ifmmode  
1552     \ll  
1553   \else  
1554     \save@sf@q{\nobreak  
1555       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bb1@allowhyphens}%  
1556   \fi}  
1557 \ProvideTextCommand{\guillemotright}{OT1}{%  
1558   \ifmmode  
1559     \gg  
1560   \else  
1561     \save@sf@q{\nobreak  
1562       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bb1@allowhyphens}%  
1563   \fi}
```

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

```
1564 \ProvideTextCommandDefault{\guillemotleft}{%  
1565   \UseTextSymbol{OT1}{\guillemotleft}}  
1566 \ProvideTextCommandDefault{\guillemotright}{%  
1567   \UseTextSymbol{OT1}{\guillemotright}}
```

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

```
\guilsinglright 1568 \ProvideTextCommand{\guilsinglleft}{OT1}{%  
1569   \ifmmode  
1570     <%  
1571   \else  
1572     \save@sf@q{\nobreak  
1573       \raise.2ex\hbox{$\scriptscriptstyle<$}\bb1@allowhyphens}%  
1574   \fi}  
1575 \ProvideTextCommand{\guilsinglright}{OT1}{%  
1576   \ifmmode
```

```

1577 >%
1578 \else
1579 \save@sf@q{\nobreak
1580 \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
1581 \fi}

```

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

```

1582 \ProvideTextCommandDefault{\guilsinglleft}{%
1583 \UseTextSymbol{OT1}{\guilsinglleft}}
1584 \ProvideTextCommandDefault{\guilsinglright}{%
1585 \UseTextSymbol{OT1}{\guilsinglright}}

```

9.11.2 Letters

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

```

1586 \DeclareTextCommand{\ij}{OT1}{%
1587 i\kern-0.02em\bbl@allowhyphens j}
1588 \DeclareTextCommand{\IJ}{OT1}{%
1589 I\kern-0.02em\bbl@allowhyphens J}
1590 \DeclareTextCommand{\ij}{T1}{\char188}
1591 \DeclareTextCommand{\IJ}{T1}{\char156}

```

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

```

1592 \ProvideTextCommandDefault{\ij}{%
1593 \UseTextSymbol{OT1}{\ij}}
1594 \ProvideTextCommandDefault{\IJ}{%
1595 \UseTextSymbol{OT1}{\IJ}}

```

`\dj` The croatian language needs the letters `\dj` and `\DJ`; they are available in the T1 encoding, `\DJ` 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).

```

1596 \def\crrtic@{\hrule height0.1ex width0.3em}
1597 \def\crttic@{\hrule height0.1ex width0.33em}
1598 \def\ddj@{%
1599 \setbox0\hbox{d}\dimen@=\ht0
1600 \advance\dimen@1ex
1601 \dimen@.45\dimen@
1602 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1603 \advance\dimen@ii.5ex
1604 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
1605 \def\DDJ@{%
1606 \setbox0\hbox{D}\dimen@=.55\ht0
1607 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1608 \advance\dimen@ii.15ex % correction for the dash position
1609 \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
1610 \dimen\thr@\expandafter\rem@pt\the\fontdimen7\font\dimen@
1611 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
1612 %
1613 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
1614 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}

```

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

```

1615 \ProvideTextCommandDefault{\dj}{%
1616   \UseTextSymbol{OT1}{\dj}}
1617 \ProvideTextCommandDefault{\DJ}{%
1618   \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.

```

1619 \DeclareTextCommand{\SS}{OT1}{SS}
1620 \ProvideTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}

```

9.11.3 Shorthands for quotation marks

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

\glq The ‘german’ single quotes.

```

\grq 1621 \ProvideTextCommandDefault{\glq}{%
1622   \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}

```

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

```

1623 \ProvideTextCommand{\grq}{T1}{%
1624   \textormath{\kern\z@\textquoteleft}{\mbox{\textquoteleft}}}
1625 \ProvideTextCommand{\grq}{TU}{%
1626   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
1627 \ProvideTextCommand{\grq}{OT1}{%
1628   \save@sf@q{\kern-.0125em
1629     \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
1630     \kern.07em\relax}}
1631 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}

```

\glqq The ‘german’ double quotes.

```

\grqq 1632 \ProvideTextCommandDefault{\glqq}{%
1633   \textormath{\quotedblbase}{\mbox{\quotedblbase}}}

```

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

```

1634 \ProvideTextCommand{\grqq}{T1}{%
1635   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1636 \ProvideTextCommand{\grqq}{TU}{%
1637   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1638 \ProvideTextCommand{\grqq}{OT1}{%
1639   \save@sf@q{\kern-.07em
1640     \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
1641     \kern.07em\relax}}
1642 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}

```

\flq The ‘french’ single guillemets.

```

\frq 1643 \ProvideTextCommandDefault{\flq}{%
1644   \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
1645 \ProvideTextCommandDefault{\frq}{%
1646   \textormath{\guilsinglright}{\mbox{\guilsinglright}}}

```

`\flqq` The ‘french’ double guillemets.

```
\frqq 1647 \ProvideTextCommandDefault{\flqq}{%
1648   \textormath{\guillemotleft}{\mbox{\guillemotleft}}}
1649 \ProvideTextCommandDefault{\frqq}{%
1650   \textormath{\guillemotright}{\mbox{\guillemotright}}}
```

9.11.4 Umlauts and tremas

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

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

```
1651 \def\umlauthigh{%
1652   \def\bbl@umlauta##1{\leavevmode\bgroup%
1653     \expandafter\accent\csname\fontencoding dqpos\endcsname
1654     ##1\bbl@allowhyphens\egroup}%
1655   \let\bbl@umlaute\bbl@umlauta}
1656 \def\umlautlow{%
1657   \def\bbl@umlauta{\protect\lower@umlaut}}
1658 \def\umlautelower{%
1659   \def\bbl@umlaute{\protect\lower@umlaut}}
1660 \umlauthigh
```

`\lower@umlaut` The command `\lower@umlaut` is used to position the `\` closer to the letter.
We want the umlaut character lowered, nearer to the letter. To do this we need an extra
(*dimen*) register.

```
1661 \expandafter\ifx\csname U@D\endcsname\relax
1662   \csname newdimen\endcsname\U@D
1663 \fi
```

The following code fools T_EX’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.

```
1664 \def\lower@umlaut#1{%
1665   \leavevmode\bgroup
1666   \U@D 1ex%
1667   {\setbox\z@\hbox{%
1668     \expandafter\char\csname\fontencoding dqpos\endcsname}%
1669     \dimen@ -.45ex\advance\dimen@\ht\z@
1670     \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
1671   \expandafter\accent\csname\fontencoding dqpos\endcsname
1672   \fontdimen5\font\U@D #1%
1673   \egroup}
```

For all vowels we declare `\` to be a composite command which uses `\bbl@umlauta` or `\bbl@umlaute` to position the umlaut character. We need to be sure that these definitions override the ones that are provided when the package `fontenc` with option `OT1` is used.

Therefore these declarations are postponed until the beginning of the document. Note these definitions only apply to some languages, but babel sets them for *all* languages – you may want to redefine `\bbl@umlauta` and/or `\bbl@umlaute` for a language in the corresponding ldf (using the babel switching mechanism, of course).

```

1674 \AtBeginDocument{%
1675   \DeclareTextCompositeCommand{"}{OT1}{a}{\bbl@umlauta{a}}%
1676   \DeclareTextCompositeCommand{"}{OT1}{e}{\bbl@umlaute{e}}%
1677   \DeclareTextCompositeCommand{"}{OT1}{i}{\bbl@umlaute{i}}%
1678   \DeclareTextCompositeCommand{"}{OT1}{\i}{\bbl@umlaute{i}}%
1679   \DeclareTextCompositeCommand{"}{OT1}{o}{\bbl@umlauta{o}}%
1680   \DeclareTextCompositeCommand{"}{OT1}{u}{\bbl@umlauta{u}}%
1681   \DeclareTextCompositeCommand{"}{OT1}{A}{\bbl@umlauta{A}}%
1682   \DeclareTextCompositeCommand{"}{OT1}{E}{\bbl@umlaute{E}}%
1683   \DeclareTextCompositeCommand{"}{OT1}{I}{\bbl@umlaute{I}}%
1684   \DeclareTextCompositeCommand{"}{OT1}{O}{\bbl@umlauta{O}}%
1685   \DeclareTextCompositeCommand{"}{OT1}{U}{\bbl@umlauta{U}}%
1686 }

```

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

```

1687 \ifx\l@english\@undefined
1688   \chardef\l@english\z@
1689 \fi
1690 \main@language{english}

```

9.12 Layout

Work in progress.

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

```

1691 \bbl@trace{Bidi layout}
1692 \providecommand\IfBabelLayout[3]{#3}%
1693 \newcommand\BabelPatchSection[1]{%
1694   \@ifundefined{#1}{%
1695     \bbl@exp{\let\<bbl@ss@#1>\<#1>}%
1696     \@namedef{#1}{%
1697       \@ifstar{\bbl@presec@s{#1}}%
1698       {\@dblarg{\bbl@presec@x{#1}}}%
1699     \def\bbl@presec@x#1[#2]#3{%
1700       \bbl@exp{%
1701         \\select@language@x{\bbl@main@language}%
1702         \\@nameuse{bbl@sspre@#1}%
1703         \\@nameuse{bbl@ss@#1}%
1704         [\\foreignlanguage{\languagename}{\unexpanded{#2}}}%
1705         {\\foreignlanguage{\languagename}{\unexpanded{#3}}}%
1706         \\select@language@x{\languagename}}%
1707       \def\bbl@presec@s#1#2{%
1708         \bbl@exp{%
1709           \\select@language@x{\bbl@main@language}%
1710           \\@nameuse{bbl@sspre@#1}%
1711           \\@nameuse{bbl@ss@#1}%
1712           {\\foreignlanguage{\languagename}{\unexpanded{#2}}}%
1713           \\select@language@x{\languagename}}%
1714       \IfBabelLayout{sectioning}%
1715       {\BabelPatchSection{part}%
1716        \BabelPatchSection{chapter}%
1717        \BabelPatchSection{section}%
1718        \BabelPatchSection{subsection}%

```

```

1719 \BabelPatchSection{subsubsection}%
1720 \BabelPatchSection{paragraph}%
1721 \BabelPatchSection{subparagraph}%
1722 \def\babel@toc#1{%
1723   \select@language@x{\bbl@main@language}}{}
1724 \IfBabelLayout{captions}%
1725   {\BabelPatchSection{caption}}{}

```

9.13 Load engine specific macros

```

1726 \bbl@trace{Input engine specific macros}
1727 \ifcase\bbl@engine
1728   \input txtbabel.def
1729 \or
1730   \input luababel.def
1731 \or
1732   \input xebabel.def
1733 \fi

```

9.14 Creating languages

`\babelprovide` is a general purpose tool for creating and modifying languages. It creates the language infrastructure, and loads, if requested, an ini file. It may be used in conjunction to previously loaded ldf files.

```

1734 \bbl@trace{Creating languages and reading ini files}
1735 \newcommand\babelprovide[2][]{%
1736   \let\bbl@savelangname\language
1737   \edef\bbl@savelocaleid{\the\localeid}%
1738   % Set name and locale id
1739   \def\language#2{%
1740     % \global\@namedef\bbl@lcname@#2}{#2}%
1741     \bbl@id@assign
1742     \let\bbl@KVP@captions\@nil
1743     \let\bbl@KVP@import\@nil
1744     \let\bbl@KVP@main\@nil
1745     \let\bbl@KVP@script\@nil
1746     \let\bbl@KVP@language\@nil
1747     \let\bbl@KVP@hyphenrules\@nil % only for provide@new
1748     \let\bbl@KVP@mapfont\@nil
1749     \let\bbl@KVP@maparabic\@nil
1750     \let\bbl@KVP@mapdigits\@nil
1751     \let\bbl@KVP@intraspace\@nil
1752     \let\bbl@KVP@intrapenalty\@nil
1753     \let\bbl@KVP@onchar\@nil
1754     \let\bbl@KVP@chargroups\@nil
1755     \bbl@forkv{#1}{% TODO - error handling
1756       \in@{.}{##1}%
1757       \ifin@
1758         \bbl@renewinikey##1\@{##2}%
1759       \else
1760         \bbl@csarg\def{KVP@##1}{##2}%
1761       \fi}%
1762   % == import, captions ==
1763   \ifx\bbl@KVP@import\@nil\else
1764     \bbl@exp{\bbl@ifblank{\bbl@KVP@import}}%
1765     {\begin@group
1766       \def\BabelBeforeIni##1##2{\gdef\bbl@KVP@import{##1}\endinput}%
1767       \InputIfFileExists{babel-#2.tex}{}}%

```

```

1768     \endgroup}%
1769   }%
1770 \fi
1771 \ifx\bbbl@KVP@captions\@nil
1772   \let\bbbl@KVP@captions\bbbl@KVP@import
1773 \fi
1774 % Load ini
1775 \bbbl@ifunset{date#2}%
1776   {\bbbl@provide@new{#2}}%
1777   {\bbbl@ifblank{#1}%
1778     {\bbbl@error
1779       {If you want to modify `#2' you must tell how in\\%
1780       the optional argument. See the manual for the\\%
1781       available options.}%
1782       {Use this macro as documented}}}%
1783   {\bbbl@provide@renew{#2}}}%
1784 % Post tasks
1785 \bbbl@exp{\bbabelensure[exclude=\\today]{#2}}%
1786 \bbbl@ifunset{bbbl@ensure@\language}%
1787   {\bbbl@exp{%
1788     \\DeclareRobustCommand\<bbbl@ensure@\language>[1]{%
1789       \\foreignlanguage{\language}%
1790       {###1}}}%
1791   }%
1792 % At this point all parameters are defined if 'import'. Now we
1793 % execute some code depending on them. But what about if nothing was
1794 % imported? We just load the very basic parameters: ids and a few
1795 % more.
1796 \bbbl@ifunset{bbbl@lname@#2}%
1797   {\def\BabelBeforeIni##1##2{%
1798     \begingroup
1799       \catcode`\[=12 \catcode`\]=12 \catcode`\==12 %
1800       \let\bbbl@ini@captions@aux\@gobbletwo
1801       \def\bbbl@inidate #####1.####2.####3.####4\relax #####5####6}%
1802       \bbbl@read@ini{##1}{basic data}%
1803       \bbbl@exportkey{chrng}{characters.ranges}{}%
1804       \bbbl@exportkey{dgnat}{numbers.digits.native}{}%
1805       \bbbl@exportkey{hyphr}{typography.hyphenrules}{}%
1806       \bbbl@exportkey{intsp}{typography.intraspace}{}%
1807       \endgroup}%
1808       {\setbox\z@\hbox{\InputIfFileExists{babel-#2.tex}{}}}%
1809   }%
1810 % -
1811 % == script, language ==
1812 % Override the values from ini or defines them
1813 \ifx\bbbl@KVP@script\@nil\else
1814   \bbbl@csarg\edef{sname@#2}{\bbbl@KVP@script}%
1815 \fi
1816 \ifx\bbbl@KVP@language\@nil\else
1817   \bbbl@csarg\edef{lname@#2}{\bbbl@KVP@language}%
1818 \fi
1819 % == onchar ==
1820 \ifx\bbbl@KVP@onchar\@nil\else
1821   \bbbl@luahyphenate
1822   \directlua{
1823     if Babel.locale_mapped == nil then
1824       Babel.locale_mapped = true
1825       Babel.linebreaking.add_before(Babel.locale_map)
1826       Babel.loc_to_scr = {}

```

```

1827     Babel.chr_to_loc = {}
1828   end}%
1829   \bbl@xin@{ ids }{ \bbl@KVP@onchar\space}%
1830   \ifin@
1831     % TODO - error/warning if no script
1832     \directlua{
1833       if Babel.script_blocks['\bbl@cs{sbc@}\language'] then
1834         Babel.loc_to_scr[\the\localeid] =
1835           Babel.script_blocks['\bbl@cs{sbc@}\language']
1836         Babel.locale_props[\the\localeid].lc = \the\localeid\space
1837         Babel.locale_props[\the\localeid].lg = \the\@nameuse{1@}\language\space
1838       end
1839     }%
1840   \fi
1841   \bbl@xin@{ fonts }{ \bbl@KVP@onchar\space}%
1842   \ifin@
1843     \bbl@ifunset{bbl@lsys@\language}{\bbl@provide@lsys@\language}{}%
1844     \bbl@ifunset{bbl@wdir@\language}{\bbl@provide@dirs@\language}{}%
1845     \directlua{
1846       if Babel.script_blocks['\bbl@cs{sbc@}\language'] then
1847         Babel.loc_to_scr[\the\localeid] =
1848           Babel.script_blocks['\bbl@cs{sbc@}\language']
1849       end}
1850   \ifx\bbl@mapselect\undefined
1851     \AtBeginDocument{%
1852       \expandafter\bbl@add\csname selectfont \endcsname{\bbl@mapselect}%
1853       {\selectfont}}%
1854     \def\bbl@mapselect{%
1855       \let\bbl@mapselect\relax
1856       \edef\bbl@prefontid{\fontid\font}}%
1857     \def\bbl@mapdir##1{%
1858       {\def\language{##1}%
1859       \let\bbl@ifrestoring\@firstoftwo % To avoid font warning
1860       \bbl@switchfont
1861       \directlua{
1862         Babel.locale_props[\the\csname bbl@id@##1\endcsname]
1863           [\bbl@prefontid] = \fontid\font\space}}}%
1864     \fi
1865     \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\language}}}%
1866   \fi
1867   % TODO - catch non-valid values
1868 \fi
1869 % \ifx\bbl@KVP@chargroups\@nil\else
1870 %   \bbl@chargroups
1871 % \fi
1872 % == mapfont ==
1873 % For bidi texts, to switch the font based on direction
1874 \ifx\bbl@KVP@mapfont\@nil\else
1875   \bbl@ifsamestring{\bbl@KVP@mapfont}{direction}{}%
1876   {\bbl@error{Option '\bbl@KVP@mapfont' unknown for\%
1877     mapfont. Use 'direction'.%
1878     {See the manual for details.}}}%
1879   \bbl@ifunset{bbl@lsys@\language}{\bbl@provide@lsys@\language}{}%
1880   \bbl@ifunset{bbl@wdir@\language}{\bbl@provide@dirs@\language}{}%
1881   \ifx\bbl@mapselect\undefined
1882     \AtBeginDocument{%
1883       \expandafter\bbl@add\csname selectfont \endcsname{\bbl@mapselect}%
1884       {\selectfont}}%
1885     \def\bbl@mapselect{%

```



```

1886 \let\bbl@mapselect\relax
1887 \edef\bbl@prefontid{\fontid\font}}%
1888 \def\bbl@mapdir##1{%
1889   {\def\language{##1}%
1890     \let\bbl@ifrestoring\@firstoftwo % avoid font warning
1891     \bbl@switchfont
1892     \directlua{Babel.fontmap
1893       [\the\csname bbl@wdir@##1\endcsname]%
1894       [\bbl@prefontid]=\fontid\font}}}%
1895 \fi
1896 \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\language}}}%
1897 \fi
1898 % == intraspace, intrapenalty ==
1899 % For CJK, East Asian, Southeast Asian, if interspace in ini
1900 \ifx\bbl@KVP@intraspace\@nil\else % We can override the ini or set
1901   \bbl@csarg\edef{intsp@#2}{\bbl@KVP@intraspace}%
1902 \fi
1903 \bbl@provide@intraspace
1904 % == maparabic ==
1905 % Native digits, if provided in ini (TeX level, xe and lua)
1906 \ifcase\bbl@engine\else
1907   \bbl@ifunset{\bbl@dgnat@language}{}%
1908   {\expandafter\ifx\csname bbl@dgnat@language\endcsname\@empty\else
1909     \expandafter\expandafter\expandafter
1910     \bbl@setdigits\csname bbl@dgnat@language\endcsname
1911     \ifx\bbl@KVP@maparabic\@nil\else
1912       \ifx\bbl@latinarabic\undefined
1913         \expandafter\let\expandafter\@arabic
1914         \csname bbl@counter@language\endcsname
1915         % ie, if layout=counters, which redefines \@arabic
1916         \expandafter\let\expandafter\bbl@latinarabic
1917         \csname bbl@counter@language\endcsname
1918       \fi
1919     \fi
1920   \fi}%
1921 \fi
1922 % == mapdigits ==
1923 % Native digits (lua level).
1924 \ifodd\bbl@engine
1925   \ifx\bbl@KVP@mapdigits\@nil\else
1926     \bbl@ifunset{\bbl@dgnat@language}{}%
1927     {\RequirePackage{luatexbase}%
1928       \bbl@activate@preotf
1929       \directlua{
1930         Babel = Babel or {} %%% -> presets in luababel
1931         Babel.digits_mapped = true
1932         Babel.digits = Babel.digits or {}
1933         Babel.digits[\the\localeid] =
1934           table.pack(string.utfvalue('\bbl@cs{dgnat@language}'))
1935         if not Babel.numbers then
1936           function Babel.numbers(head)
1937             local LOCALE = luatexbase.registernumber'bbl@attr@locale'
1938             local GLYPH = node.id'glyph'
1939             local inmath = false
1940             for item in node.traverse(head) do
1941               if not inmath and item.id == GLYPH then
1942                 local temp = node.get_attribute(item, LOCALE)
1943                 if Babel.digits[temp] then
1944                   local chr = item.char

```

```

1945         if chr > 47 and chr < 58 then
1946             item.char = Babel.digits[temp][chr-47]
1947         end
1948     end
1949     elseif item.id == node.id'math' then
1950         inmath = (item.subtype == 0)
1951     end
1952 end
1953 return head
1954 end
1955 end
1956 }}
1957 \fi
1958 \fi
1959 % == require.babel in ini ==
1960 % To load or reload the babel-*.tex, if require.babel in ini
1961 \bbl@ifunset{bbl@rqtex@\language\language}\fi%
1962 {\expandafter\ifx\csname bbl@rqtex@\language\endcsname\@empty\else
1963     \let\BabelBeforeIni\@gobbletwo
1964     \chardef\atcatcode=\catcode`\@
1965     \catcode`\@=11\relax
1966     \InputIfFileExists{babel-\bbl@cs{rqtex@\language}.tex}\fi}%
1967     \catcode`\@=\atcatcode
1968     \let\atcatcode\relax
1969 \fi}%
1970 % == main ==
1971 \ifx\bbl@KVP@main\@nil % Restore only if not 'main'
1972     \let\language\bbl@savelangname
1973     \chardef\localeid\bbl@savelocaleid\relax
1974 \fi}

```

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

```

1975 \def\bbl@setdigits#1#2#3#4#5{%
1976     \bbl@exp{%
1977         \def\<\language digits>####1{% ie, \langdigits
1978             \<bbl@digits@\language>####1\\\@nil}%
1979         \def\<\language counter>####1{% ie, \langcounter
1980             \expandafter\<bbl@counter@\language>%
1981             \csname c@####1\endcsname}%
1982         \def\<bbl@counter@\language>####1{% ie, \bbl@counter@lang
1983             \expandafter\<bbl@digits@\language>%
1984             \number####1\\\@nil}}%
1985     \def\bbl@tempa##1##2##3##4##5{%
1986         \bbl@exp{% Wow, quite a lot of hashes! :-(
1987             \def\<bbl@digits@\language>#####1{%
1988                 \ifx#####1\\\@nil % ie, \bbl@digits@lang
1989                 \else
1990                     \ifx0#####1#1%
1991                     \else\ifx1#####1#2%
1992                     \else\ifx2#####1#3%
1993                     \else\ifx3#####1#4%
1994                     \else\ifx4#####1#5%
1995                     \else\ifx5#####1##1%
1996                     \else\ifx6#####1##2%
1997                     \else\ifx7#####1##3%
1998                     \else\ifx8#####1##4%
1999                     \else\ifx9#####1##5%
2000                     \else#####1%

```


The hyphenrules option is handled with an auxiliary macro.

```

2057 \def\bbl@provide@hyphens#1{%
2058   \let\bbl@tempa\relax
2059   \ifx\bbl@KVP@hyphenrules\@nil\else
2060     \bbl@replace\bbl@KVP@hyphenrules{ }{,}%
2061     \bbl@foreach\bbl@KVP@hyphenrules{%
2062       \ifx\bbl@tempa\relax % if not yet found
2063         \bbl@ifsamestring{##1}{+}%
2064         {\bbl@exp{\addlanguage\<l@##1>}}}%
2065       {}%
2066       \bbl@ifunset{l@##1}%
2067       {}%
2068       {\bbl@exp{\let\bbl@tempa\<l@##1>}}}%
2069     \fi}%
2070 \fi
2071 \ifx\bbl@tempa\relax % if no opt or no language in opt found
2072   \ifx\bbl@KVP@import\@nil\else % if importing
2073     \bbl@exp{%
2074       \bbl@ifblank{\@nameuse{bbl@hyphr@#1}}}%
2075     {}%
2076     {\let\bbl@tempa\<l@\@nameuse{bbl@hyphr@language}>}}}%
2077 \fi
2078 \fi
2079 \bbl@ifunset{bbl@tempa}% ie, relax or undefined
2080 {\bbl@ifunset{l@#1}% no hyphenrules found - fallback
2081  {\bbl@exp{\adddialect\<l@#1>\language}}}%
2082  {}% so, l@<lang> is ok - nothing to do
2083  {\bbl@exp{\adddialect\<l@#1>\bbl@tempa}}% found in opt list or ini
2084 \bbl@ifunset{bbl@prehc@language}%
2085 {}% TODO - XeTeX, based on \babelfont and HyphenChar?
2086 {\ifodd\bbl@engine\bbl@exp{%
2087   \bbl@ifblank{\@nameuse{bbl@prehc@#1}}}%
2088   {}%
2089   {\AddBabelHook[language]{babel-prehc-\language}{patterns}%
2090    {\prehyphenchar=\@nameuse{bbl@prehc@language}\relax}}}%
2091 \fi}}

```

The reader of ini files. There are 3 possible cases: a section name (in the form [. . .]), a comment (starting with ;) and a key/value pair.

```

2092 \ifx\bbl@readstream\@undefined
2093   \csname newread\endcsname\bbl@readstream
2094 \fi
2095 \def\bbl@read@ini#1#2{%
2096   \global\@namedef{bbl@lini@language}{#1}%
2097   \openin\bbl@readstream=babel-#1.ini
2098   \ifeof\bbl@readstream
2099     \bbl@error
2100     {There is no ini file for the requested language\%
2101      (#1). Perhaps you misspelled it or your installation\%
2102      is not complete.}%
2103     {Fix the name or reinstall babel.}%
2104   \else
2105     \let\bbl@section\@empty
2106     \let\bbl@savestrings\@empty
2107     \let\bbl@savetoday\@empty
2108     \let\bbl@savestate\@empty
2109     \def\bbl@inipreread##1=##2\@{%
2110       \bbl@trim\def\bbl@tempa{##1}% Redundant below !!

```

```

2111 % Move trims here ??
2112 \bbl@ifunset{bbl@KVP@\bbl@section.\bbl@tempa}%
2113 {\expandafter\bbl@inireader\bbl@tempa=##2\@@}%
2114 {}}%
2115 \let\bbl@inireader\bbl@iniskip
2116 \bbl@info{Importing #2 for \language\%
2117 from babel-#1.ini. Reported}%
2118 \loop
2119 \if T\ifeof\bbl@readstream F\fi T\relax % Trick, because inside \loop
2120 \endlinechar\m@ne
2121 \read\bbl@readstream to \bbl@line
2122 \endlinechar\^^M
2123 \ifx\bbl@line\empty\else
2124 \expandafter\bbl@inline\bbl@line\bbl@inline
2125 \fi
2126 \repeat
2127 \bbl@foreach\bbl@renewlist{%
2128 \bbl@ifunset{bbl@renew@##1}{\bbl@inisec[##1]\@@}%
2129 \global\let\bbl@renewlist\empty
2130 % Ends last section. See \bbl@inisec
2131 \def\bbl@elt##1##2{\bbl@inireader##1=##2\@@}%
2132 \@nameuse{bbl@renew@\bbl@section}%
2133 \global\bbl@csarg\let{renew@\bbl@section}\relax
2134 \@nameuse{bbl@secpost@\bbl@section}%
2135 \fi}
2136 \def\bbl@inline#1\bbl@inline{%
2137 \@ifnextchar[\bbl@inisec{\@ifnextchar;\bbl@iniskip\bbl@inipreread}#1\@@}% ]

```

The special cases for comment lines and sections are handled by the two following commands. In sections, we provide the possibility to take extra actions at the end or at the start (TODO - but note the last section is not ended). By default, key=val pairs are ignored. The secpost “hook” is used only by ‘identification’, while secpre only by date.gregorian.licr.

```

2138 \def\bbl@iniskip#1\@@{% if starts with ;
2139 \def\bbl@inisec[#1]#2\@@{% if starts with opening bracket
2140 \def\bbl@elt##1##2{\bbl@inireader##1=##2\@@}%
2141 \@nameuse{bbl@renew@\bbl@section}%
2142 \global\bbl@csarg\let{renew@\bbl@section}\relax
2143 \@nameuse{bbl@secpost@\bbl@section}% ends previous section
2144 \def\bbl@section{#1}% starts current section
2145 \def\bbl@elt##1##2{%
2146 \@namedef{bbl@KVP@#1..##1}{}}%
2147 \@nameuse{bbl@renew@#1}%
2148 \@nameuse{bbl@secpre@#1}% pre-section 'hook'
2149 \bbl@ifunset{bbl@inikv@#1}%
2150 {\let\bbl@inireader\bbl@iniskip}%
2151 {\bbl@exp{\let\bbl@inireader<\bbl@inikv@#1>}}}
2152 \let\bbl@renewlist\empty
2153 \def\bbl@renewinikey#1..#2\@@#3{%
2154 \bbl@ifunset{bbl@renew@#1}%
2155 {\bbl@add@list\bbl@renewlist{#1}}%
2156 {}%
2157 \bbl@csarg\bbl@add{renew@#1}{\bbl@elt{#2}{#3}}}

```

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

```

2158 \def\bbl@inikv#1=##2\@@{% key=value
2159 \bbl@trim@def\bbl@tempa{#1}%
2160 \bbl@trim\toks@{#2}%
2161 \bbl@csarg\edef{kv@\bbl@section.\bbl@tempa}{\the\toks@}}

```

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

```

2162 \def\bbl@exportkey#1#2#3{%
2163   \bbl@ifunset{\bbl@kv@#2}%
2164     {\bbl@csarg\gdef{#1@\language}\{#3}}%
2165     {\expandafter\ifx\csname\bbl@kv@#2\endcsname\@empty
2166       \bbl@csarg\gdef{#1@\language}\{#3}}%
2167     \else
2168       \bbl@exp{\global\let\<bbl@#1@\language>\<bbl@kv@#2>}%
2169       \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@secpost@identification` is called always (via `\bbl@inisec`), while `\bbl@after@ini` must be called explicitly after `\bbl@read@ini` if necessary.

```

2170 \let\bbl@inikv@identification\bbl@inikv
2171 \def\bbl@secpost@identification{%
2172   \bbl@exportkey{elname}{identification.name.english}{}%
2173   \bbl@exp{\bbl@exportkey{lname}{identification.name.opentype}%
2174     {\csname\bbl@elname@\language\endcsname}}%
2175   \bbl@exportkey{lbcpr}{identification.tag.bcp47}{}%
2176   \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
2177   \bbl@exportkey{esname}{identification.script.name}{}%
2178   \bbl@exp{\bbl@exportkey{sname}{identification.script.name.opentype}%
2179     {\csname\bbl@esname@\language\endcsname}}%
2180   \bbl@exportkey{sbcpr}{identification.script.tag.bcp47}{}%
2181   \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}}
2182 \let\bbl@inikv@typography\bbl@inikv
2183 \let\bbl@inikv@characters\bbl@inikv
2184 \let\bbl@inikv@numbers\bbl@inikv
2185 \def\bbl@after@ini{%
2186   \bbl@exportkey{lftm}{typography.lefthyphenmin}{2}%
2187   \bbl@exportkey{rgthm}{typography.righthyphenmin}{3}%
2188   \bbl@exportkey{prehc}{typography.prehyphenchar}{}%
2189   \bbl@exportkey{hyphr}{typography.hyphenrules}{}%
2190   \bbl@exportkey{intsp}{typography.intraspace}{}%
2191   \bbl@exportkey{jstfy}{typography.justify}{w}%
2192   \bbl@exportkey{chrng}{characters.ranges}{}%
2193   \bbl@exportkey{dgnat}{numbers.digits.native}{}%
2194   \bbl@exportkey{rqtex}{identification.require.babel}{}%
2195   \bbl@xin@{0.5}{\@nameuse{\bbl@kv@identification.version}}%
2196   \ifin@
2197     \bbl@warning{%
2198       There are neither captions nor date in '\language'.\%
2199       It may not be suitable for proper typesetting, and it\%
2200       could change. Reported}%
2201   \fi
2202   \bbl@xin@{0.9}{\@nameuse{\bbl@kv@identification.version}}%
2203   \ifin@
2204     \bbl@warning{%
2205       The '\language' date format may not be suitable\%
2206       for proper typesetting, and therefore it very likely will\%
2207       change in a future release. Reported}%
2208   \fi
2209   \bbl@toglobal\bbl@savetoday
2210   \bbl@toglobal\bbl@savestate}

```

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.

```
2211 \ifcase\bbl@engine
2212   \bbl@csarg\def{inikv@captions.licr}#1=#2\@@{%
2213     \bbl@ini@captions@aux{#1}{#2}}
2214 \else
2215   \def\bbl@inikv@captions#1=#2\@@{%
2216     \bbl@ini@captions@aux{#1}{#2}}
2217 \fi
```

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

```
2218 \def\bbl@ini@captions@aux#1#2{%
2219   \bbl@trim@def\bbl@tempa{#1}%
2220   \bbl@ifblank{#2}%
2221   {\bbl@exp{%
2222     \toks@{\bbl@nocaption{\bbl@tempa}{\language\bbl@tempa name}}}%
2223   {\bbl@trim\toks@{#2}}}%
2224   \bbl@exp{%
2225     \bbl@add\bbl@savestrings{%
2226       \SetString\<\bbl@tempa name>{\the\toks@}}}
```

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

TODO. Remove copypaste pattern.

```
2227 \bbl@csarg\def{inikv@date.gregorian}#1=#2\@@{%           for defaults
2228   \bbl@inidate#1...\relax{#2}{}}
2229 \bbl@csarg\def{inikv@date.islamic}#1=#2\@@{%
2230   \bbl@inidate#1...\relax{#2}{islamic}}
2231 \bbl@csarg\def{inikv@date.hebrew}#1=#2\@@{%
2232   \bbl@inidate#1...\relax{#2}{hebrew}}
2233 \bbl@csarg\def{inikv@date.persian}#1=#2\@@{%
2234   \bbl@inidate#1...\relax{#2}{persian}}
2235 \bbl@csarg\def{inikv@date.indian}#1=#2\@@{%
2236   \bbl@inidate#1...\relax{#2}{indian}}
2237 \ifcase\bbl@engine
2238   \bbl@csarg\def{inikv@date.gregorian.licr}#1=#2\@@{%      override
2239     \bbl@inidate#1...\relax{#2}{}}
2240   \bbl@csarg\def{secpre@date.gregorian.licr}{%             discard uni
2241     \ifcase\bbl@engine\let\bbl@savestate\empty\fi}
2242 \fi
2243 % eg: 1=months, 2=wide, 3=1, 4=dummy
2244 \def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
2245   \bbl@trim@def\bbl@tempa{#1.#2}%
2246   \bbl@ifsamestring{\bbl@tempa}{months.wide}%              to savedate
2247   {\bbl@trim@def\bbl@tempa{#3}%
2248     \bbl@trim\toks@{#5}%
2249     \bbl@exp{%
2250       \bbl@add\bbl@savestate{%
2251         \SetString\<month\romannumeral\bbl@tempa#6name>{\the\toks@}}}%
2252     {\bbl@ifsamestring{\bbl@tempa}{date.long}%              defined now
2253       {\bbl@trim@def\bbl@toreplace{#5}%
2254         \bbl@TG@date
2255         \global\bbl@csarg\let{date@\language}\bbl@toreplace
2256         \bbl@exp{%
2257           \gdef\<\language date>{\protect\<\language date >}}%
2258           \gdef\<\language date >###1###2###3{%
2259             \bbl@usedategrouptue
2260             \<\bbl@ensure@\language>{%
```

```

2261 \<bbl@date@language>{####1}{####2}{####3}}}%
2262 \\bbl@add\\bbl@savetoday{%
2263 \\SetString\\today{%
2264 \<language date>{\\the\year}{\\the\month}{\\the\day}}}%
2265 {}

```

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.

```

2266 \let\bbl@calendar\@empty
2267 \newcommand\BabelDateSpace{\nobreakspace}
2268 \newcommand\BabelDateDot{.\@}
2269 \newcommand\BabelDated[1]{\number#1}
2270 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
2271 \newcommand\BabelDateM[1]{\number#1}
2272 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}
2273 \newcommand\BabelDateMMMM[1]{%
2274 \csname month\romannumeral#1\bbl@calendar name\endcsname}%
2275 \newcommand\BabelDatey[1]{\number#1}%
2276 \newcommand\BabelDateyy[1]{%
2277 \ifnum#1<10 0\number#1 %
2278 \else\ifnum#1<100 \number#1 %
2279 \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
2280 \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
2281 \else
2282 \bbl@error
2283 {Currently two-digit years are restricted to the\
2284 range 0-9999.}%
2285 {There is little you can do. Sorry.}%
2286 \fi\fi\fi}}
2287 \newcommand\BabelDateyyyy[1]{\number#1} % FIXME - add leading 0
2288 \def\bbl@replace@finish@iii#1{%
2289 \bbl@exp{\def\#1####1####2####3{\the\toks@}}
2290 \def\bbl@TG@date{%
2291 \bbl@replace\bbl@toreplace{[ ]}{\BabelDateSpace}}%
2292 \bbl@replace\bbl@toreplace{[.]}{\BabelDateDot}}%
2293 \bbl@replace\bbl@toreplace{[d]}{\BabelDated{####3}}%
2294 \bbl@replace\bbl@toreplace{[dd]}{\BabelDatedd{####3}}%
2295 \bbl@replace\bbl@toreplace{[M]}{\BabelDateM{####2}}%
2296 \bbl@replace\bbl@toreplace{[MM]}{\BabelDateMM{####2}}%
2297 \bbl@replace\bbl@toreplace{[MMM]}{\BabelDateMMMM{####2}}%
2298 \bbl@replace\bbl@toreplace{[y]}{\BabelDatey{####1}}%
2299 \bbl@replace\bbl@toreplace{[yy]}{\BabelDateyy{####1}}%
2300 \bbl@replace\bbl@toreplace{[yyyy]}{\BabelDateyyyy{####1}}%
2301 % Note after \bbl@replace \toks@ contains the resulting string.
2302 % TODO - Using this implicit behavior doesn't seem a good idea.
2303 \bbl@replace@finish@iii\bbl@toreplace}

```

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

```

2304 \def\bbl@provide@lsys#1{%
2305 \bbl@ifunset{bbl@lname@#1}%
2306 {\bbl@ini@basic{#1}}%
2307 {}%
2308 \bbl@csarg\let{lsys@#1}\@empty
2309 \bbl@ifunset{bbl@sname@#1}{\bbl@csarg\gdef{sname@#1}{Default}}}%
2310 \bbl@ifunset{bbl@sotf@#1}{\bbl@csarg\gdef{sotf@#1}{DFLT}}}%
2311 \bbl@csarg\bbl@add@list{lsys@#1}{Script=\bbl@cs{sname@#1}}%
2312 \bbl@ifunset{bbl@lname@#1}}%

```



```

2313    {\bbl@csarg\bbl@add@list{lsys@#1}{Language=\bbl@cs{lname@#1}}}%
2314    \bbl@csarg\bbl@tglobal{lsys@#1}}

```

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

```

2315 \def\bbl@ini@basic#1{%
2316   \def\BabelBeforeIni##1##2{%
2317     \begingroup
2318       \bbl@add\bbl@secpost@identification{\closein\bbl@readstream}%
2319       \catcode`\[=12 \catcode`\]=12 \catcode`\==12 %
2320       \bbl@read@ini{##1}{font and identification data}%
2321       \endinput          % babel- .tex may contain onlypreamble's
2322       \endgroup}%        boxed, to avoid extra spaces:
2323   {\setbox\z@\hbox{\InputIfFileExists{babel-#1.tex}{}}}}

```

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

```

2324 \newcommand\localeinfo[1]{%
2325   \bbl@ifunset{bbl@csname bbl@info@#1\endcsname @\languagename}%
2326   {\bbl@error{I've found no info for the current locale.\\%
2327     The corresponding ini file has not been loaded\\%
2328     Perhaps it doesn't exist}%
2329     {See the manual for details.}}%
2330   {\@nameuse{bbl@csname bbl@info@#1\endcsname @\languagename}}}
2331 % \@namedef{bbl@info@name.locale}{lname}
2332 \@namedef{bbl@info@tag.ini}{lini}
2333 \@namedef{bbl@info@name.english}{elname}
2334 \@namedef{bbl@info@name.opentype}{lname}
2335 \@namedef{bbl@info@tag.bcp47}{lbc47}
2336 \@namedef{bbl@info@tag.opentype}{lotf}
2337 \@namedef{bbl@info@script.name}{esname}
2338 \@namedef{bbl@info@script.name.opentype}{sname}
2339 \@namedef{bbl@info@script.tag.bcp47}{sbcp}
2340 \@namedef{bbl@info@script.tag.opentype}{sotf}
2341 \let\bbl@ensureinfo\@gobble
2342 \newcommand\BabelEnsureInfo{%
2343   \def\bbl@ensureinfo##1{%
2344     \ifx\InputIfFileExists\@undefined\else % not in plain
2345       \bbl@ifunset{bbl@lname@##1}{\bbl@ini@basic{##1}}}%
2346     \fi}}

```

10 Adjusting the Babel behavior

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

```

2347 \newcommand\babeladjust[1]{% TODO. Error handling.
2348   \bbl@forkv{#1}{\@nameuse{bbl@ADJ@##1@##2}}
2349 %
2350 \def\bbl@adjust@lua#1#2{%
2351   \ifvmode
2352     \ifnum\currentgrouplevel=\z@
2353       \directlua{ Babel.#2 }%
2354       \expandafter\expandafter\expandafter\@gobble
2355     \fi
2356   \fi

```

```

2357 {\bbl@error % The error is gobbled if everything went ok.
2358 {Currently, #1 related features can be adjusted only\\%
2359 in the main vertical list.}%
2360 {Maybe things change in the future, but this is what it is.}}}
2361 \@namedef{bbl@ADJ@bidi.mirroring@on}{%
2362 \bbl@adjust@lua{bidi}{mirroring_enabled=true}}
2363 \@namedef{bbl@ADJ@bidi.mirroring@off}{%
2364 \bbl@adjust@lua{bidi}{mirroring_enabled=false}}
2365 \@namedef{bbl@ADJ@bidi.text@on}{%
2366 \bbl@adjust@lua{bidi}{bidi_enabled=true}}
2367 \@namedef{bbl@ADJ@bidi.text@off}{%
2368 \bbl@adjust@lua{bidi}{bidi_enabled=false}}
2369 \@namedef{bbl@ADJ@bidi.mapdigits@on}{%
2370 \bbl@adjust@lua{bidi}{digits_mapped=true}}
2371 \@namedef{bbl@ADJ@bidi.mapdigits@off}{%
2372 \bbl@adjust@lua{bidi}{digits_mapped=false}}
2373 %
2374 \@namedef{bbl@ADJ@linebreak.sea@on}{%
2375 \bbl@adjust@lua{linebreak}{sea_enabled=true}}
2376 \@namedef{bbl@ADJ@linebreak.sea@off}{%
2377 \bbl@adjust@lua{linebreak}{sea_enabled=false}}
2378 \@namedef{bbl@ADJ@linebreak.cjk@on}{%
2379 \bbl@adjust@lua{linebreak}{cjk_enabled=true}}
2380 \@namedef{bbl@ADJ@linebreak.cjk@off}{%
2381 \bbl@adjust@lua{linebreak}{cjk_enabled=false}}
2382 %
2383 \def\bbl@adjust@layout#1{%
2384 \ifvmode
2385 #1%
2386 \expandafter\@gobble
2387 \fi
2388 {\bbl@error % The error is gobbled if everything went ok.
2389 {Currently, layout related features can be adjusted only\\%
2390 in vertical mode.}%
2391 {Maybe things change in the future, but this is what it is.}}}
2392 \@namedef{bbl@ADJ@layout.tabular@on}{%
2393 \bbl@adjust@layout{\let\@tabular\bbl@NL@@tabular}}
2394 \@namedef{bbl@ADJ@layout.tabular@off}{%
2395 \bbl@adjust@layout{\let\@tabular\bbl@OL@@tabular}}
2396 \@namedef{bbl@ADJ@layout.lists@on}{%
2397 \bbl@adjust@layout{\let\list\bbl@NL@list}}
2398 \@namedef{bbl@ADJ@layout.lists@off}{%
2399 \bbl@adjust@layout{\let\list\bbl@OL@list}}
2400 \@namedef{bbl@ADJ@hyphenation.extra@on}{%
2401 \directlua{
2402 Babel.linebreaking.add_after(Babel.post_hyphenate_replace)
2403 }}

```

11 The kernel of Babel (babel.def for L^AT_EX only)

11.1 The redefinition of the style commands

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

the group is closed. This is accomplished by the command `\aftergroup`.

```
2404 {\def\format{plain}
2405 \ifx\fmtname\format
2406 \else
2407   \def\format{LaTeX2e}
2408   \ifx\fmtname\format
2409     \else
2410       \aftergroup\endinput
2411     \fi
2412 \fi}
```

11.2 Cross referencing macros

The \LaTeX book states:

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

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

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

The only way to accomplish this in most cases is to use the trick described in the \TeX book [4] (Appendix D, page 382). The primitive `\meaning` applied to a token expands to the current meaning of this token. For example, ‘`\meaning\A`’ with `\A` defined as ‘`\def\A#1{\B}`’ expands to the characters ‘`macro:#1->\B`’ with all category codes set to ‘other’ or ‘space’.

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

```
2413 %\bbl@redefine\newlabel#1#2{%
2414 % \safe@activetrue\org@newlabel{#1}{#2}\@safe@activesfalse}
```

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

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

```
2415 <<{*More package options}>> ≡
2416 \DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
2417 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
2418 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
2419 <</More package options>>
```

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

```
2420 \bbl@trace{Cross referencing macros}
2421 \ifx\bbl@opt@safe\@empty\else
2422   \def\@newl@bel#1#2#3{%
2423     {\@safe@activetrue
2424       \bbl@ifunset{#1@#2}%
2425       \relax
2426       {\gdef\@multiplelabels{%
2427         \latex@warning@no@line{There were multiply-defined labels}}%
2428         \latex@warning@no@line{Label `#2' multiply defined}}%
2429       \global\@namedef{#1@#2}{#3}}}
```

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

```
2430 \CheckCommand*\@testdef[3]{%
2431   \def\reserved@a{#3}%
2432   \expandafter\ifx\csname#1@#2\endcsname\reserved@a
2433   \else
2434     \@tempswatrue
2435   \fi}
```

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

```
2436 \def\@testdef#1#2#3{%
2437   \@safe@activetrue
```

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

```
2438   \expandafter\let\expandafter\bbl@tempa\csname #1@#2\endcsname
```

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

```
2439   \def\bbl@tempb{#3}%
2440   \@safe@activesfalse
```

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

```
2441   \ifx\bbl@tempa\relax
2442   \else
2443     \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
2444   \fi
```

We do the same for `\bbl@tempb`.

```
2445   \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
```

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

```
2446   \ifx\bbl@tempa\bbl@tempb
2447   \else
2448     \@tempswatrue
2449   \fi}
2450 \fi
```

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

```
2451 \bbl@xin@{R}\bbl@opt@safe
2452 \ifin@
2453   \bbl@redefineroobust\ref#1{%
2454     \@safe@activetrue\org@ref{#1}\@safe@activesfalse}
2455   \bbl@redefineroobust\pageref#1{%
2456     \@safe@activetrue\org@pageref{#1}\@safe@activesfalse}
2457 \else
2458   \let\org@ref\ref
2459   \let\org@pageref\pageref
2460 \fi
```

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

```
2461 \bbl@xin@{B}\bbl@opt@safe
2462 \ifin@
2463 \bbl@redefine\@citex[#1]#2{%
2464   \@safe@activetrue\edef\@tempa{#2}\@safe@activesfalse
2465   \org@@citex[#1]{\@tempa}}
```

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.

```
2466 \AtBeginDocument{%
2467   \@ifpackageloaded{natbib}{%
```

Notice that we use `\def` here instead of `\bbl@redefine` because `\org@@citex` is already defined and we don't want to overwrite that definition (it would result in parameter stack overflow because of a circular definition).

(Recent versions of `natbib` change dynamically `\@citex`, so PR4087 doesn't seem fixable in a simple way. Just load `natbib` before.)

```
2468   \def\@citex[#1][#2]#3{%
2469     \@safe@activetrue\edef\@tempa{#3}\@safe@activesfalse
2470     \org@@citex[#1][#2]{\@tempa}}%
2471   }{}}
```

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

```
2472 \AtBeginDocument{%
2473   \@ifpackageloaded{cite}{%
2474     \def\@citex[#1]#2{%
2475       \@safe@activetrue\org@@citex[#1][#2]\@safe@activesfalse}%
2476     }{}}
```

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

```
2477 \bbl@redefine\nocite#1{%
2478   \@safe@activetrue\org@nocite{#1}\@safe@activesfalse}
```

`\bibcite` The macro that is used in the `.aux` file to define citation labels. When packages such as `natbib` or `cite` are not loaded its second argument is used to typeset the citation label. In that case, this second argument can contain active characters but is used in an environment where `\@safe@activetrue` 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.

```
2479 \bbl@redefine\bibcite{%
2480   \bbl@cite@choice
2481   \bibcite}
```

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

```
2482 \def\bbl@bibcite#1#2{%
2483   \org@bibcite{#1}{\@safe@activesfalse#2}}
```

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

```
2484 \def\bbl@cite@choice{%
2485   \global\let\bibcite\bbl@bibcite
```

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

```
2486   \@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{}%
2487   \@ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{}%
```

Make sure this only happens once.

```
2488   \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.

```
2489   \AtBeginDocument{\bbl@cite@choice}
```

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

```
2490 \bbl@redefine\@bibitem#1{%
2491   \@safe@activetrue\org@@bibitem{#1}\@safe@activetruefalse}
2492 \else
2493   \let\org@nocite\nocite
2494   \let\org@@citex\@citex
2495   \let\org@bibcite\bibcite
2496   \let\org@@bibitem\@bibitem
2497 \fi
```

11.3 Marks

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

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

```
2498 \bbl@trace{Marks}
2499 \IfBabelLayout{sectioning}
2500   {\ifx\bbl@opt@headfoot\@nnil
2501     \g@addto@macro\@resetactivechars{%
2502       \set@typeset@protect
2503       \expandafter\select@language@x\expandafter{\bbl@main@language}%
2504       \let\protect\noexpand
2505       \edef\thepage{%
2506         \noexpand\babelsublr{\unexpanded\expandafter{\thepage}}}%
2507     \fi}
2508   {\ifbbl@single\else
2509     \bbl@ifunset{markright} \bbl@redefine\bbl@redefineroobust
2510     \markright#1{%
2511       \bbl@ifblank{#1}%
2512       {\org@markright}}}%
2513     {\toks@{#1}}%
2514     \bbl@exp{%
```

```

2515          \org@markright{\protect\foreignlanguage{\language}%
2516          {\protect\bb1@restore@actives\the\toks@}}}%

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

2517 \ifx\@mkboth\markboth
2518 \def\bb1@tempc{\let\@mkboth\markboth}
2519 \else
2520 \def\bb1@tempc{}
2521 \fi
2522 \bb1@ifunset{markboth }\bb1@redefine\bb1@redefineroobust
2523 \markboth#1#2{%
2524 \protected@edef\bb1@tempb##1{%
2525 \protect\foreignlanguage
2526 {\language}{\protect\bb1@restore@actives##1}}%
2527 \bb1@ifblank{#1}%
2528 {\toks@{}}%
2529 {\toks@\expandafter{\bb1@tempb{#1}}}%
2530 \bb1@ifblank{#2}%
2531 {\@temptokena{}}%
2532 {\@temptokena\expandafter{\bb1@tempb{#2}}}%
2533 \bb1@exp{\org@markboth{\the\toks@}{\the\@temptokena}}
2534 \bb1@tempc
2535 \fi} % end ifbb1@single, end \IfBabelLayout

```

11.4 Preventing clashes with other packages

11.4.1 ifthen

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

```

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

```

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

The first thing we need to do is check if the package `ifthen` is loaded. This should be done at `\begin{document}` time.

```

2536 \bb1@trace{Preventing clashes with other packages}
2537 \bb1@xin@{R}\bb1@opt@safe
2538 \ifin@
2539 \AtBeginDocument{%
2540 \ifpackageloaded{ifthen}{%

```

Then we can redefine `\ifthenelse`:

```

2541 \bb1@redefine@long\ifthenelse#1#2#3{%

```

We want to revert the definition of `\pageref` and `\ref` to their original definition for the first argument of `\ifthenelse`, so we first need to store their current meanings.

```

2542      \let\bbl@temp@pref\pageref
2543      \let\pageref\org@pageref
2544      \let\bbl@temp@ref\ref
2545      \let\ref\org@ref

```

Then we can set the `\@safe@actives` switch and call the original `\ifthenelse`. In order to be able to use shorthands in the second and third arguments of `\ifthenelse` the resetting of the switch *and* the definition of `\pageref` happens inside those arguments. When the package wasn't loaded we do nothing.

```

2546      \@safe@activetrue
2547      \org@ifthenelse{#1}%
2548      {\let\pageref\bbl@temp@pref
2549      \let\ref\bbl@temp@ref
2550      \@safe@activesfalse
2551      #2}%
2552      {\let\pageref\bbl@temp@pref
2553      \let\ref\bbl@temp@ref
2554      \@safe@activesfalse
2555      #3}%
2556      }%
2557      }{}%
2558      }

```

11.4.2 varioref

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

```

2559 \AtBeginDocument{%
2560   \ifpackageloaded{varioref}{%
2561     \bbl@redefine\@@vpageref#1[#2]#3{%
2562       \@safe@activetrue
2563       \org@@@vpageref{#1}[#2]{#3}%
2564       \@safe@activesfalse}%
2565     \bbl@redefine\vrefpagenum#1#2{%
2566       \@safe@activetrue
2567       \org@vrefpagenum{#1}{#2}%
2568       \@safe@activesfalse}%

```

The package `varioref` defines `\Ref` to be a robust command which uppercases the first character of the reference text. In order to be able to do that it needs to access the 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.

```

2569   \expandafter\def\csname Ref \endcsname#1{%
2570     \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
2571   }{}%
2572   }
2573 \fi

```

11.4.3 hhline

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

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

```
2574 \AtEndOfPackage{%
2575   \AtBeginDocument{%
2576     \ifpackageloaded{hhline}%
```

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

```
2577     {\expandafter\ifx\csname normal@char\string\endcsname\relax
2578       \else
```

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

```
2579         \makeatletter
2580         \def\@currname{hhline}\input{hhline.sty}\makeatother
2581         \fi}%
2582     {}}}
```

11.4.4 hyperref

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

```
2583 \AtBeginDocument{%
2584   \ifx\pdfstringdefDisableCommands\undefined\else
2585     \pdfstringdefDisableCommands{\languageshortands{system}}%
2586   \fi}
```

11.4.5 fancyhdr

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

```
2587 \DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
2588   \lowercase{\foreignlanguage{#1}}}
```

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

```
2589 \def\substitutefontfamily#1#2#3{%
2590   \lowercase{\immediate\openout15=#1#2.fd\relax}%
2591   \immediate\write15{%
2592     \string\ProvidesFile{#1#2.fd}%
2593     [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
2594     \space generated font description file]^^J
2595     \string\DeclareFontFamily{#1}{#2}{^^J
2596     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{^^J
2597     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{^^J
2598     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{^^J
2599     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{^^J
2600     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{^^J
2601     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{^^J
2602     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{^^J
2603     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{^^J
2604     }%
2605   \closeout15
2606 }
```

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

```
2607 \@onlypreamble\substitutefontfamily
```

11.5 Encoding and fonts

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

`\ensureascii`

```
2608 \bbl@trace{Encoding and fonts}
2609 \newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU,PU,PD1}
2610 \newcommand\BabelNonText{TS1,T3,TS3}
2611 \let\org@TeX\TeX
2612 \let\org@LaTeX\LaTeX
2613 \let\ensureascii\@firstofone
2614 \AtBeginDocument{%
2615   \in@false
2616   \bbl@foreach\BabelNonASCII{% is there a text non-ascii enc?
2617     \ifin@
2618       \lowercase{\bbl@xin@{,#1enc.def,}\@filelist,}%
2619     \fi}%
2620   \ifin@ % if a text non-ascii has been loaded
2621     \def\ensureascii#1{{\fontencoding{OT1}\selectfont#1}}%
2622     \DeclareTextCommandDefault{\TeX}{\org@TeX}%
2623     \DeclareTextCommandDefault{\LaTeX}{\org@LaTeX}%
2624     \def\bbl@tempb#1\@@{\uppercase{\bbl@tempc#1}ENC.DEF\empty\@@}%
2625     \def\bbl@tempc#1ENC.DEF#2\@@{%
2626       \ifx\empty#2\else
2627         \bbl@ifunset{T#1}%
2628         {}%
2629         {\bbl@xin@{,#1,}\@filelist,\BabelNonASCII,\BabelNonText,}%
2630       \ifin@
2631         \DeclareTextCommand{\TeX}{#1}{\ensureascii{\org@TeX}}%
2632         \DeclareTextCommand{\LaTeX}{#1}{\ensureascii{\org@LaTeX}}%
2633       \else
2634         \def\ensureascii##1{{\fontencoding{#1}\selectfont##1}}%
2635       \fi}%
2636     \fi}%
2637   \bbl@foreach\@filelist{\bbl@tempb#1\@@}% TODO - \@@ de mas??
2638   \bbl@xin@{\cf@encoding,}\@filelist,\BabelNonASCII,\BabelNonText,%
2639   \ifin@
2640     \edef\ensureascii#1{{%
2641       \noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}}%
2642     \fi
2643   \fi}
```

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

`\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.

```
2644 \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.

```

2645 \AtBeginDocument{%
2646   \ifpackageloaded{fontspec}%
2647     {\xdef\latinencoding{%
2648       \ifx\UTFencname\@undefined
2649         EU\ifcase\bbl@engine\or2\or1\fi
2650       \else
2651         \UTFencname
2652       \fi}}%
2653   {\gdef\latinencoding{OT1}%
2654     \ifx\cf@encoding\bbl@t@one
2655       \xdef\latinencoding{\bbl@t@one}%
2656     \else
2657       \@ifl@aded{def}{t1enc}{\xdef\latinencoding{\bbl@t@one}}{}%
2658     \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.

```

2659 \DeclareRobustCommand{\latintext}{%
2660   \fontencoding{\latinencoding}\selectfont
2661   \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.

```

2662 \ifx\@undefined\DeclareTextFontCommand
2663   \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
2664 \else
2665   \DeclareTextFontCommand{\textlatin}{\latintext}
2666 \fi

```

11.6 Basic bidi support

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

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

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

- `pdftex` provides a minimal support for bidi text, and it must be done by hand. Vertical typesetting is not possible.
- `xetex` is somewhat better, thanks to its font engine (even if not always reliable) and a few additional tools. However, very little is done at the paragraph level. Another challenging problem is text direction does not honour \TeX grouping.
- `luatex` can provide the most complete solution, as we can manipulate almost freely the node list, the generated lines, and so on, but bidi text does not work out of the box and

some development is necessary. It also provides tools to properly set left-to-right and right-to-left page layouts. As Lua_T_EX-j_a shows, vertical typesetting is possible, too. Its main drawback is font handling is often considered to be less mature than xetex, mainly in Indic scripts (but there are steps to make HarfBuzz, the xetex font engine, available in luatex; see <<https://github.com/tatzetwerk/luatex-harfbuzz>>).

```

2667 \bbl@trace{Basic (internal) bidi support}
2668 \def\bbl@alscripts{,Arabic,Syriac,Thaana,}
2669 \def\bbl@rscripts{%
2670   ,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
2671   Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaeen,%
2672   Manichaeen,Meroitic Cursive,Meroitic,Old North Arabian,%
2673   Nabataean,N'Ko,Orkhon,Palmyrene,Inscriptional Pahlavi,%
2674   Psalter Pahlavi,Phoenician,Inscriptional Parthian,Samaritan,%
2675   Old South Arabian,}%
2676 \def\bbl@provide@dirs#1{%
2677   \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts\bbl@rscripts}%
2678   \ifin@
2679     \global\bbl@csarg\chardef{wdir@#1}\@ne
2680     \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts}%
2681     \ifin@
2682       \global\bbl@csarg\chardef{wdir@#1}\tw@ % useless in xetex
2683       \fi
2684     \else
2685       \global\bbl@csarg\chardef{wdir@#1}\z@
2686       \fi
2687     \ifodd\bbl@engine
2688       \bbl@csarg\ifcase{wdir@#1}%
2689         \directlua{ Babel.locale_props[\the\localeid].texmdir = 'l' }%
2690         \or
2691         \directlua{ Babel.locale_props[\the\localeid].texmdir = 'r' }%
2692         \or
2693         \directlua{ Babel.locale_props[\the\localeid].texmdir = 'al' }%
2694         \fi
2695       \fi}
2696 \def\bbl@switchdir{%
2697   \bbl@ifunset{bbl@lsys@\languagename}{\bbl@provide@lsys{\languagename}}{}%
2698   \bbl@ifunset{bbl@wdir@\languagename}{\bbl@provide@dirs{\languagename}}{}%
2699   \bbl@exp{\bbl@setdirs\bbl@cs{wdir@\languagename}}%
2700 \def\bbl@setdirs#1{% TODO - math
2701   \ifcase\bbl@select@type % TODO - strictly, not the right test
2702     \bbl@bodydir{#1}%
2703     \bbl@pardir{#1}%
2704   \fi
2705   \bbl@texmdir{#1}}
2706 \ifodd\bbl@engine % luatex=1
2707   \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
2708   \DisableBabelHook{babel-bidi}
2709   \chardef\bbl@thetexmdir\z@
2710   \chardef\bbl@thepardir\z@
2711   \def\bbl@getluadir#1{%
2712     \directlua{
2713       if tex.#1dir == 'TLT' then
2714         tex.sprint('0')
2715       elseif tex.#1dir == 'TRT' then
2716         tex.sprint('1')
2717       end}}
2718   \def\bbl@setluadir#1#2#3{% 1=text/par.. 2=\texmdir.. 3=0 lr/1 rl
2719     \ifcase#3\relax

```

```

2720     \ifcase\bb1@getluadir{#1}\relax\else
2721     #2 TLT\relax
2722     \fi
2723 \else
2724     \ifcase\bb1@getluadir{#1}\relax
2725     #2 TRT\relax
2726     \fi
2727 \fi}
2728 \def\bb1@textdir#1{%
2729     \bb1@setluadir{text}\textdir{#1}%
2730     \chardef\bb1@thetextdir#1\relax
2731     \setattribute\bb1@attr@dir{\numexpr\bb1@thepardir*3+#1}}
2732 \def\bb1@pardir#1{%
2733     \bb1@setluadir{par}\pardir{#1}%
2734     \chardef\bb1@thepardir#1\relax}
2735 \def\bb1@bodydir{\bb1@setluadir{body}\bodydir}
2736 \def\bb1@pagedir{\bb1@setluadir{page}\pagedir}
2737 \def\bb1@dirparastext{\pardir\the\textdir\relax}%   %%%
2738 % Sadly, we have to deal with boxes in math with basic.
2739 % Activated every math with the package option bidi=:
2740 \def\bb1@mathboxdir{%
2741     \ifcase\bb1@thetextdir\relax
2742     \everyhbox{\textdir TLT\relax}%
2743     \else
2744     \everyhbox{\textdir TRT\relax}%
2745     \fi}
2746 \else % pdftex=0, xetex=2
2747     \AddBabelHook{babel-bidi}{afterextras}{\bb1@switchdir}
2748     \DisableBabelHook{babel-bidi}
2749     \newcount\bb1@dirlevel
2750     \chardef\bb1@thetextdir\z@
2751     \chardef\bb1@thepardir\z@
2752     \def\bb1@textdir#1{%
2753         \ifcase#1\relax
2754             \chardef\bb1@thetextdir\z@
2755             \bb1@textdir@i\beginL\endL
2756         \else
2757             \chardef\bb1@thetextdir@ne
2758             \bb1@textdir@i\beginR\endR
2759         \fi}
2760     \def\bb1@textdir@i#1#2{%
2761         \ifhmode
2762             \ifnum\currentgrouplevel>\z@
2763                 \ifnum\currentgrouplevel=\bb1@dirlevel
2764                     \bb1@error{Multiple bidi settings inside a group}%
2765                     {I'll insert a new group, but expect wrong results.}%
2766                     \bgroup\aftergroup#2\aftergroup\egroup
2767                 \else
2768                     \ifcase\currentgrouptype\or % 0 bottom
2769                     \aftergroup#2% 1 simple {}
2770                     \or
2771                     \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
2772                     \or
2773                     \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
2774                     \or\or\or % vbox vtop align
2775                     \or
2776                     \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
2777                     \or\or\or\or\or\or % output math disc insert vcent mathchoice
2778                     \or

```

```

2779         \aftergroup#2% 14 \begingroup
2780     \else
2781         \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
2782     \fi
2783 \fi
2784     \bbl@dirlevel\currentgrouplevel
2785 \fi
2786 #1%
2787 \fi}
2788 \def\bbl@pardir#1{\chardef\bbl@thepardir#1\relax}
2789 \let\bbl@bodydir\@gobble
2790 \let\bbl@pagedir\@gobble
2791 \def\bbl@dirparastext{\chardef\bbl@thepardir\bbl@thetextdir}

```

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

```

2792 \def\bbl@xebidipar{%
2793     \let\bbl@xebidipar\relax
2794     \TeXeTstate\@ne
2795     \def\bbl@xeverypar{%
2796         \ifcase\bbl@thepardir
2797             \ifcase\bbl@thetextdir\else\beginR\fi
2798         \else
2799             {\setbox\z@\lastbox\beginR\box\z@}%
2800         \fi}%
2801     \let\bbl@severypar\everypar
2802     \newtoks\everypar
2803     \everypar=\bbl@severypar
2804     \bbl@severypar{\bbl@xeverypar\the\everypar}}
2805 \@ifpackagewith{babel}{bidi=bidi}%
2806 {\let\bbl@textdir\i\@gobbletwo
2807     \let\bbl@xebidipar\@empty
2808     \AddBabelHook{bidi}{foreign}{%
2809         \def\bbl@tempa{\def\BabelText####1}%
2810         \ifcase\bbl@thetextdir
2811             \expandafter\bbl@tempa\expandafter{\BabelText{\LR{##1}}}%
2812         \else
2813             \expandafter\bbl@tempa\expandafter{\BabelText{\RL{##1}}}%
2814         \fi}
2815     \def\bbl@pardir#1{\ifcase#1\relax\setLR\else\setRL\fi}}
2816 {}%
2817 \fi

```

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

```

2818 \DeclareRobustCommand\babelsublr[1]{\leavevmode{\bbl@textdir\z@#1}}
2819 \AtBeginDocument{%
2820     \ifx\pdfstringdefDisableCommands\@undefined\else
2821         \ifx\pdfstringdefDisableCommands\relax\else
2822             \pdfstringdefDisableCommands{\let\babelsublr\@firstofone}%
2823         \fi
2824     \fi}

```

11.7 Local Language Configuration

`\loadlocalcfg` At some sites it may be necessary to add site-specific actions to a language definition file. This can be done by creating a file with the same name as the language definition file, but

with the extension .cfg. For instance the file norsk.cfg will be loaded when the language definition file norsk.ldf is loaded.

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

```
2825 \bbl@trace{Local Language Configuration}
2826 \ifx\loadlocalcfg\@undefined
2827   \@ifpackagewith{babel}{noconfigs}%
2828   {\let\loadlocalcfg@gobble}%
2829   {\def\loadlocalcfg#1{%
2830     \InputIfFileExists{#1.cfg}%
2831     {\typeout{*****^J%
2832               * Local config file #1.cfg used^^J%
2833               *}}%
2834     \@empty}}
2835 \fi
```

Just to be compatible with L^AT_EX 2.09 we add a few more lines of code:

```
2836 \ifx\@unexpandable@protect\@undefined
2837   \def\@unexpandable@protect{\noexpand\protect\noexpand}
2838   \long\def\protected@write#1#2#3{%
2839     \begingroup
2840       \let\thepage\relax
2841       #2%
2842       \let\protect\@unexpandable@protect
2843       \edef\reserved@a{\write#1{#3}}%
2844       \reserved@a
2845     \endgroup
2846     \if@nobreak\ifvmode\nobreak\fi\fi}
2847 \fi
2848 </core>
2849 <*kernel>
```

12 Multiple languages (switch.def)

Plain T_EX 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.

```
2850 <<Make sure ProvidesFile is defined>>
2851 \ProvidesFile{switch.def}[\<date>] [\<version>] Babel switching mechanism]
2852 <<Load macros for plain if not LaTeX>>
2853 <<Define core switching macros>>
```

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

```
2854 \def\bbl@version{\<version>}
2855 \def\bbl@date{\<date>}
2856 \def\adddialect#1#2{%
2857   \global\chardef#1#2\relax
2858   \bbl@usehooks{adddialect}{\#1}{\#2}%
2859   \begingroup
2860     \count@#1\relax
2861     \def\bbl@elt##1##2##3##4{%
2862       \ifnum\count@=##2\relax
2863         \bbl@info{\string#1 = using hyphenrules for ##1\%
2864                   (\string\language\the\count@)}%
2865         \def\bbl@elt####1####2####3####4{%
```

```

2866     \fi}%
2867     \bbl@languages
2868 \endgroup}

```

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

```

2869 \def\bbl@fixname#1{%
2870   \begingroup
2871   \def\bbl@tempe{#1}%
2872   \edef\bbl@tempd{\noexpand\@ifundefined{\noexpand\bbl@tempe#1}}%
2873   \bbl@tempd
2874   {\lowercase\expandafter{\bbl@tempd}}%
2875   {\uppercase\expandafter{\bbl@tempd}}%
2876   \@empty
2877   {\edef\bbl@tempd{\def\noexpand#1{#1}}}%
2878   {\uppercase\expandafter{\bbl@tempd}}}%
2879   {\edef\bbl@tempd{\def\noexpand#1{#1}}}%
2880   {\lowercase\expandafter{\bbl@tempd}}}%
2881   \@empty
2882   \edef\bbl@tempd{\endgroup\def\noexpand#1{#1}}%
2883   \bbl@tempd}
2884 \def\bbl@iflanguage#1{%
2885   \@ifundefined{l@#1}{\@nolanerr{#1}\@gobble}\@firstofone}

```

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

```

2886 \def\iflanguage#1{%
2887   \bbl@iflanguage{#1}{%
2888     \ifnum\csname l@#1\endcsname=\language
2889       \expandafter\@firstoftwo
2890     \else
2891       \expandafter\@secondoftwo
2892     \fi}}

```

12.1 Selecting the language

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

To allow the call of `\selectlanguage` either with a control sequence name or with a simple string as argument, we have to use a trick to delete the optional escape character. To convert a control sequence to a string, we use the `\string` primitive. Next we have to look at the first character of this string and compare it with the escape character. Because this escape character can be changed by setting the internal integer `\escapechar` to a character number, we have to compare this number with the character of the string. To do this we have to use T_EX’s backquote notation to specify the character as a number. If the first character of the `\string`’ed argument is the current escape character, the comparison has stripped this character and the rest in the ‘then’ part consists of the rest of

the control sequence name. Otherwise we know that either the argument is not a control sequence or `\escapechar` is set to a value outside of the character range 0–255. If the user gives an empty argument, we provide a default argument for `\string`. This argument should expand to nothing.

```
2893 \let\bbl@select@type\z@
2894 \edef\selectlanguage{%
2895   \noexpand\protect
2896   \expandafter\noexpand\csname selectlanguage \endcsname}
```

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

```
2897 \ifx\@undefined\protect\let\protect\relax\fi
```

As \LaTeX 2.09 writes to files *expanded* whereas \LaTeX 2_ε takes care *not* to expand the arguments of `\write` statements we need to be a bit clever about the way we add information to .aux files. Therefore we introduce the macro `\xstring` which should expand to the right amount of `\string`'s.

```
2898 \ifx\documentclass\@undefined
2899   \def\xstring{\string\string\string}
2900 \else
2901   \let\xstring\string
2902 \fi
```

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

`\bbl@pop@language` But when the language change happens *inside* a group the end of the group doesn't write anything to the auxiliary files. Therefore we need \TeX 's `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.

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

```
2903 \def\bbl@language@stack{}
```

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

`\bbl@push@language` The stack is simply a list of languagenames, separated with a '+' sign; the push function can be simple:

`\bbl@pop@language`

```
2904 \def\bbl@push@language{%
2905   \xdef\bbl@language@stack{\language+\bbl@language@stack}}
```

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

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

```
2906 \def\bbl@pop@lang#1+#2-#3{%
2907   \edef\language{#1}\xdef#3{#2}}
```

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

```
2908 \let\bbl@ifrestoring\@secondoftwo
2909 \def\bbl@pop@language{%
2910   \expandafter\bbl@pop@lang\bbl@language@stack-\bbl@language@stack
2911   \let\bbl@ifrestoring\@firstoftwo
2912   \expandafter\bbl@set@language\expandafter{\language}%
2913   \let\bbl@ifrestoring\@secondoftwo}
```

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

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

```
2914 \chardef\localeid\z@
2915 \def\bbl@id@last{0} % No real need for a new counter
2916 \def\bbl@id@assign{%
2917   \bbl@ifunset{bbl@id@@\language}%
2918   {\count@bbl@id@last\relax
2919    \advance\count@\@ne
2920    \bbl@csarg\chardef{id@@\language}\count@
2921    \edef\bbl@id@last{\the\count@}%
2922    \ifcase\bbl@engine\or
2923      \directlua{
2924        Babel = Babel or {}
2925        Babel.locale_props = Babel.locale_props or {}
2926        Babel.locale_props[\bbl@id@last] = {}
2927        Babel.locale_props[\bbl@id@last].name = '\language'
2928      }%
2929    \fi}%
2930  }%
2931  \chardef\localeid\@nameuse{bbl@id@@\language}}
```

The unprotected part of `\selectlanguage`.

```
2932 \expandafter\def\csname selectlanguage \endcsname#1{%
2933   \ifnum\bbl@hymapsel=\@ccclv\let\bbl@hymapsel\tw@\fi
2934   \bbl@push@language
2935   \aftergroup\bbl@pop@language
2936   \bbl@set@language{#1}}
```

`\bbl@set@language` The macro `\bbl@set@language` takes care of switching the language environment *and* of writing entries on the auxiliary files. For historical reasons, language names can be either language of `\language`. To catch either form a trick is used, but unfortunately as a side effect the catcodes of letters in `\language` 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.

```
2937 \def\BabelContentsFiles{toc,lof,lot}
2938 \def\bbl@set@language#1{% from selectlanguage, pop@
```

```

2939 \edef\language{\ifnum\escapechar=\expandafter\string#1\@empty
2940 \else\string#1\@empty\fi}%
2941 \else\string#1\@empty\fi}%
2942 % \@namedef{bbl@lcnam@#1}{#1}%
2943 \select@language{\language}%
2944 % write to aux
2945 \expandafter\ifx\csname date\language\endcsname\relax\else
2946 \if@filesw
2947 \ifx\babel@aux@gobbletwo\else % Set if single in the first, redundant
2948 \protected@write\@auxout{}\string\babel@aux{\language}{}}%
2949 \fi
2950 \bbl@usehooks{write}}}%
2951 \fi
2952 \fi}
2953 \def\select@language#1{% from set@, babel@aux
2954 % set hmap
2955 \ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi
2956 % set name
2957 \edef\language{#1}%
2958 \bbl@fixname\language
2959 \bbl@iflanguage\language{%
2960 \expandafter\ifx\csname date\language\endcsname\relax
2961 \bbl@error
2962 {Unknown language `#1'. Either you have\\%
2963 misspelled its name, it has not been installed,\\%
2964 or you requested it in a previous run. Fix its name,\\%
2965 install it or just rerun the file, respectively. In\\%
2966 some cases, you may need to remove the aux file}%
2967 {You may proceed, but expect wrong results}%
2968 \else
2969 % set type
2970 \let\bbl@select@type\z@
2971 \expandafter\bbl@switch\expandafter{\language}%
2972 \fi}}
2973 \def\babel@aux#1#2{%
2974 \expandafter\ifx\csname date#1\endcsname\relax
2975 \expandafter\ifx\csname bbl@auxwarn@#1\endcsname\relax
2976 \@namedef{bbl@auxwarn@#1}{}%
2977 \bbl@warning
2978 {Unknown language `#1'. Very likely you\\%
2979 requested it in a previous run. Expect some\\%
2980 wrong results in this run, which should vanish\\%
2981 in the next one. Reported}%
2982 \fi
2983 \else
2984 \select@language{#1}%
2985 \bbl@foreach\BabelContentsFiles{%
2986 \@writefile{##1}{\babel@toc{#1}{#2}}}% % TODO - ok in plain?
2987 \fi}
2988 \def\babel@toc#1#2{%
2989 \select@language{#1}}

```

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

```
2990 \let\select@language\x\select@language
```

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 `\language`.

Then we have to *redefine* `\originalTeX` to compensate for the things that have been activated. To save memory space for the macro definition of `\originalTeX`, we construct the control sequence name for the `\noextras<lang>` command at definition time by expanding the `\csname` primitive.

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

The switching of the values of `\lefthyphenmin` and `\righthyphenmin` is somewhat different. First we save their current values, then we check if `\<lang>hyphenmins` is defined. If it is not, we set default values (2 and 3), otherwise the values in `\<lang>hyphenmins` will be used.

```

2991 \newif\ifbbl@usedategroup
2992 \def\bbl@switch#1{% from select@, foreign@
2993 % make sure there is info for the language if so requested
2994 \bbl@ensureinfo{#1}%
2995 % restore
2996 \originalTeX
2997 \expandafter\def\expandafter\originalTeX\expandafter{%
2998   \csname noextras#1\endcsname
2999   \let\originalTeX\@empty
3000   \babel@beginsave}%
3001 \bbl@usehooks{afterreset}{}%
3002 \languageshortands{none}%
3003 % set the locale id
3004 \bbl@id@assign
3005 % switch captions, date
3006 \ifcase\bbl@select@type
3007   \ifhmode
3008     \hskip\z@skip % trick to ignore spaces
3009     \csname captions#1\endcsname\relax
3010     \csname date#1\endcsname\relax
3011     \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
3012   \else
3013     \csname captions#1\endcsname\relax
3014     \csname date#1\endcsname\relax
3015   \fi
3016 \else
3017   \ifbbl@usedategroup % if \foreign... within \<lang>date
3018     \bbl@usedategroupfalse
3019     \ifhmode
3020       \hskip\z@skip % trick to ignore spaces
3021       \csname date#1\endcsname\relax
3022       \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
3023     \else
3024       \csname date#1\endcsname\relax
3025     \fi
3026   \fi
3027 \fi
3028 % switch extras
3029 \bbl@usehooks{beforeextras}{}%
3030 \csname extras#1\endcsname\relax
3031 \bbl@usehooks{afterextras}{}%
3032 % > babel-ensure
3033 % > babel-sh-<short>
3034 % > babel-bidi
3035 % > babel-fontspec
3036 % hyphenation - case mapping

```

```

3037 \ifcase\bb1@opt@hyphenmap\or
3038 \def\BabelLower##1##2{\lccode##1=##2\relax}%
3039 \ifnum\bb1@hymapsel>4\else
3040 \csname\language @bb1@hyphenmap\endcsname
3041 \fi
3042 \chardef\bb1@opt@hyphenmap\z@
3043 \else
3044 \ifnum\bb1@hymapsel>\bb1@opt@hyphenmap\else
3045 \csname\language @bb1@hyphenmap\endcsname
3046 \fi
3047 \fi
3048 \global\let\bb1@hymapsel\@cclv
3049 % hyphenation - patterns
3050 \bb1@patterns{#1}%
3051 % hyphenation - mins
3052 \babel@savevariable\lefthyphenmin
3053 \babel@savevariable\righthyphenmin
3054 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
3055 \set@hyphenmins\tw@\thr@@\relax
3056 \else
3057 \expandafter\expandafter\expandafter\set@hyphenmins
3058 \csname #1hyphenmins\endcsname\relax
3059 \fi}

```

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

```

3060 \long\def\otherlanguage#1{%
3061 \ifnum\bb1@hymapsel=\@cclv\let\bb1@hymapsel\thr@@\fi
3062 \csname selectlanguage \endcsname{#1}%
3063 \ignorespaces}

```

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

```

3064 \long\def\endotherlanguage{%
3065 \global\@ignoretrue\ignorespaces}

```

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

```

3066 \expandafter\def\csname otherlanguage*\endcsname#1{%
3067 \ifnum\bb1@hymapsel=\@cclv\chardef\bb1@hymapsel4\relax\fi
3068 \foreign@language{#1}}

```

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

```

3069 \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.

```

3070 \providecommand\bbl@beforeforeign{}
3071 \edef\foreignlanguage{%
3072   \noexpand\protect
3073   \expandafter\noexpand\csname foreignlanguage \endcsname}
3074 \expandafter\def\csname foreignlanguage \endcsname{%
3075   \@ifstar\bbl@foreign@s\bbl@foreign@x}
3076 \def\bbl@foreign@x#1#2{%
3077   \begingroup
3078     \let\BabelText\@firstofone
3079     \bbl@beforeforeign
3080     \foreign@language{#1}%
3081     \bbl@usehooks{foreign}{}%
3082     \BabelText{#2}% Now in horizontal mode!
3083   \endgroup}
3084 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \setpar, ?\@@par
3085   \begingroup
3086     {\par}%
3087     \let\BabelText\@firstofone
3088     \foreign@language{#1}%
3089     \bbl@usehooks{foreign*}{}%
3090     \bbl@dirparastext
3091     \BabelText{#2}% Still in vertical mode!
3092   {\par}%
3093   \endgroup}

```

`\foreign@language` This macro does the work for `\foreignlanguage` and the `otherlanguage*` 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`.

```

3094 \def\foreign@language#1{%
3095   % set name
3096   \edef\language#1%
3097   % \@namedef{\bbl@lcn#1}{#1}%
3098   \bbl@fixname\language
3099   \bbl@iflanguage\language%
3100   \expandafter\ifx\csname \language\endcsname\relax
3101     \bbl@warning % TODO - why a warning, not an error?
3102       {Unknown language `#1'. Either you have\\%
3103         misspelled its name, it has not been installed,\\%
3104         or you requested it in a previous run. Fix its name,\\%
3105         install it or just rerun the file, respectively. In\\%

```

```

3106         some cases, you may need to remove the aux file.\%
3107         I'll proceed, but expect wrong results.\%
3108         Reported}%
3109     \fi
3110     % set type
3111     \let\bbbl@select@type\@ne
3112     \expandafter\bbbl@switch\expandafter{\language}%

```

\bbbl@patterns This macro selects the hyphenation patterns by changing the `\language` register. If special hyphenation patterns are available specifically for the current font encoding, use them instead of the default.

It also sets hyphenation exceptions, but only once, because they are global (here language `\lccode's` has been set, too). `\bbbl@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.

```

3113 \let\bbbl@hyphlist\@empty
3114 \let\bbbl@hyphenation@ \relax
3115 \let\bbbl@pttnlist\@empty
3116 \let\bbbl@patterns@ \relax
3117 \let\bbbl@hymapsel=\@cclv
3118 \def\bbbl@patterns#1{%
3119     \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
3120         \csname l@#1\endcsname
3121         \edef\bbbl@tempa{#1}%
3122     \else
3123         \csname l@#1:\f@encoding\endcsname
3124         \edef\bbbl@tempa{#1:\f@encoding}%
3125     \fi
3126     \@expandtwoargs\bbbl@usehooks{patterns}{#1}{\bbbl@tempa}%
3127     % > luatex
3128     \@ifundefined{bbbl@hyphenation@}{% Can be \relax!
3129         \begingroup
3130             \bbbl@xin@{\number\language,}{,\bbbl@hyphlist}%
3131         \ifin@ \else
3132             \@expandtwoargs\bbbl@usehooks{hyphenation}{#1}{\bbbl@tempa}%
3133             \hyphenation{%
3134                 \bbbl@hyphenation@
3135                 \@ifundefined{bbbl@hyphenation@#1}%
3136                 \@empty
3137                 {\space\csname bbbl@hyphenation@#1\endcsname}}%
3138             \xdef\bbbl@hyphlist{\bbbl@hyphlist\number\language,}%
3139         \fi
3140     \endgroup}}

```

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

```

3141 \def\hyphenrules#1{%
3142     \edef\bbbl@tempf{#1}%
3143     \bbbl@fixname\bbbl@tempf
3144     \bbbl@iflanguage\bbbl@tempf{%
3145         \expandafter\bbbl@patterns\expandafter{\bbbl@tempf}%
3146         \languageshorthands{none}%
3147         \expandafter\ifx\csname\bbbl@tempf hyphenmins\endcsname\relax

```

```

3148     \set@hyphenmins\tw@\thr@@\relax
3149     \else
3150         \expandafter\expandafter\expandafter\set@hyphenmins
3151         \csname\bbl@tempf hyphenmins\endcsname\relax
3152     \fi}}
3153 \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 \langhyphenmins is already defined this command has no effect.

3154 \def\providehyphenmins#1#2{%
3155     \expandafter\ifx\csname #1hyphenmins\endcsname\relax
3156         \@namedef{#1hyphenmins}{#2}%
3157     \fi}

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

3158 \def\set@hyphenmins#1#2{%
3159     \lefthyphenmin#1\relax
3160     \righthyphenmin#2\relax}

\ProvidesLanguage The identification code for each file is something that was introduced in LATEX 2ε. 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.

3161 \ifx\ProvidesFile\@undefined
3162     \def\ProvidesLanguage#1[#2 #3 #4]{%
3163         \wlog{Language: #1 #4 #3 <#2>}%
3164     }
3165 \else
3166     \def\ProvidesLanguage#1{%
3167         \begingroup
3168             \catcode`\ 10 %
3169             \@makeother\/%
3170             \@ifnextchar[%]
3171                 {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}
3172     \def\@provideslanguage#1[#2]{%
3173         \wlog{Language: #1 #2}%
3174         \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
3175     \endgroup}
3176 \fi

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

3177 \def\LdfInit{%
3178     \chardef\atcatcode=\catcode`\@
3179     \catcode`\@=11\relax
3180     \input babel.def\relax
3181     \catcode`\@=\atcatcode \let\atcatcode\relax
3182     \LdfInit}

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

3183 \ifx\originalTeX\@undefined\let\originalTeX\@empty\fi

```


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

```
3184 \ifx\babel@beginsave\@undefined\let\babel@beginsave\relax\fi
```

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

```
3185 \providecommand\setlocale{%
3186   \bbl@error
3187   {Not yet available}%
3188   {Find an armchair, sit down and wait}}
3189 \let\uselocale\setlocale
3190 \let\locale\setlocale
3191 \let\selectlocale\setlocale
3192 \let\textlocale\setlocale
3193 \let\textlanguage\setlocale
3194 \let\languagetext\setlocale
```

12.2 Errors

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

`\@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 $\text{\LaTeX 2}_{\epsilon}$, 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.

```
3195 \edef\bbl@nulllanguage{\string\language=0}
3196 \ifx\PackageError\@undefined
3197   \def\bbl@error#1#2{%
3198     \begingroup
3199     \newlinechar=`^^J
3200     \def\{^^J(babel) }%
3201     \errhelp{#2}\errmessage{\#1}%
3202   \endgroup}
3203 \def\bbl@warning#1{%
3204   \begingroup
3205   \newlinechar=`^^J
3206   \def\{^^J(babel) }%
3207   \message{\#1}%
3208   \endgroup}
3209 \let\bbl@infowarn\bbl@warning
3210 \def\bbl@info#1{%
3211   \begingroup
3212   \newlinechar=`^^J
3213   \def\{^^J}%
3214   \wlog{#1}%
3215   \endgroup}
3216 \else
3217   \def\bbl@error#1#2{%
3218     \begingroup
```

```

3219     \def\{\MessageBreak}%
3220     \PackageError{babel}{#1}{#2}%
3221   \endgroup}
3222   \def\bbl@warning#1{%
3223     \begingroup
3224       \def\{\MessageBreak}%
3225       \PackageWarning{babel}{#1}%
3226     \endgroup}
3227   \def\bbl@infowarn#1{%
3228     \begingroup
3229       \def\{\MessageBreak}%
3230       \GenericWarning
3231         {(babel) \@spaces\@spaces\@spaces}%
3232         {Package babel Info: #1}%
3233     \endgroup}
3234   \def\bbl@info#1{%
3235     \begingroup
3236       \def\{\MessageBreak}%
3237       \PackageInfo{babel}{#1}%
3238     \endgroup}
3239 \fi
3240 \@ifpackagewith{babel}{silent}
3241   {\let\bbl@info\@gobble
3242    \let\bbl@infowarn\@gobble
3243    \let\bbl@warning\@gobble}
3244   {}
3245   \def\bbl@nocaption{\protect\bbl@nocaption@i}
3246   \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
3247     \global\@namedef{#2}{\textbf{?#1?}}%
3248     \@nameuse{#2}%
3249     \bbl@warning{%
3250       \@backslashchar#2 not set. Please, define\\%
3251       it in the preamble with something like:\\%
3252       \string\renewcommand\@backslashchar#2{..}\\%
3253       Reported}}
3254   \def\bbl@tentative{\protect\bbl@tentative@i}
3255   \def\bbl@tentative@i#1{%
3256     \bbl@warning{%
3257       Some functions for '#1' are tentative.\\%
3258       They might not work as expected and their behavior\\%
3259       could change in the future.\\%
3260       Reported}}
3261   \def\@nolanerr#1{%
3262     \bbl@error
3263     {You haven't defined the language #1\space yet}%
3264     {Your command will be ignored, type <return> to proceed}}
3265   \def\@nopatterns#1{%
3266     \bbl@warning
3267     {No hyphenation patterns were preloaded for\\%
3268      the language `#1' into the format.\\%
3269      Please, configure your TeX system to add them and\\%
3270      rebuild the format. Now I will use the patterns\\%
3271      preloaded for \bbl@nulllanguage\space instead}}
3272   \let\bbl@usehooks\@gobbletwo
3273 </kernel>
3274 <*patterns>

```

13 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` can be used to include this code in the file `hyphen.cfg`. Code is written with lower level macros.

We want to add a message to the message `LATEX 2.09` puts in the `\everyjob` register. This could be done by the following code:

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

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

- When someone wants to use several hyphenation patterns with `LATEX` the above scheme won't work. The reason is that `LATEX` overwrites the contents of the `\everyjob` register with its own message.
- Plain `TEX` does not use the `\everyjob` register so the message would not be displayed.

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

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

This new definition starts by adding an instruction to write a message on the terminal and in the transcript file to inform the user of the preloaded hyphenation patterns.

Then everything is restored to the old situation and the format is dumped.

```
3275 <<Make sure ProvidesFile is defined>>
3276 \ProvidesFile{hyphen.cfg}[<<date>>] <<version>> Babel hyphens]
3277 \xdef\bb1@format{\jobname}
3278 \ifx\AtBeginDocument\@undefined
3279   \def\@empty{}
3280   \let\orig@dump\dump
3281   \def\dump{%
3282     \ifx\@ztryfc\@undefined
3283       \else
3284         \toks0=\expandafter{\@preamblecmds}%
3285         \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
3286         \def\@begindocumenthook{}%
3287       \fi
3288       \let\dump\orig@dump\let\orig@dump\@undefined\dump}
3289 \fi
3290 <<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.

```

3291 \def\process@line#1#2 #3 #4 {%
3292   \ifx=#1%
3293     \process@synonym{#2}%
3294   \else
3295     \process@language{#1#2}{#3}{#4}%
3296   \fi
3297   \ignorespaces}

```

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

```

3298 \toks@{}
3299 \def\bbl@languages{}

```

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

Otherwise the name will be a synonym for the language loaded last.

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

```

3300 \def\process@synonym#1{%
3301   \ifnum\last@language=\m@ne
3302     \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
3303   \else
3304     \expandafter\chardef\csname l@#1\endcsname\last@language
3305     \wlog{\string\l@#1=\string\language\the\last@language}%
3306     \expandafter\let\csname #1hyphenmins\endcsname\expandafter\endcsname
3307     \csname\language\endcsname hyphenmins\endcsname
3308     \let\bbl@elt\relax
3309     \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}}}%
3310   \fi}

```

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

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

Pattern files may contain assignments to `\lefthyphenmin` and `\righthyphenmin`. \TeX does not keep track of these assignments. Therefore we try to detect such assignments and store them in the `\langhyphenmins` 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.

```

3311 \def\process@language#1#2#3{%
3312   \expandafter\addlanguage\csname l@#1\endcsname
3313   \expandafter\language\csname l@#1\endcsname
3314   \edef\language#1{%
3315     \bbl@hook@everylanguage{#1}%
3316     % > luatex
3317     \bbl@get@enc#1::\@@@
3318   \begingroup
3319     \lefthyphenmin\m@ne
3320     \bbl@hook@loadpatterns{#2}%
3321     % > luatex
3322     \ifnum\lefthyphenmin=\m@ne
3323     \else
3324       \expandafter\xdef\csname #1hyphenmins\endcsname{%
3325         \the\lefthyphenmin\the\righthyphenmin}%
3326     \fi
3327   \endgroup
3328   \def\bbl@tempa{#3}%
3329   \ifx\bbl@tempa\@empty\else
3330     \bbl@hook@loadexceptions{#3}%
3331     % > luatex
3332   \fi
3333   \let\bbl@elt\relax
3334   \edef\bbl@languages{%
3335     \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
3336   \ifnum\the\language=\z@
3337     \expandafter\ifx\csname #1hyphenmins\endcsname\relax
3338       \set@hyphenmins\tw@\thr@@\relax
3339     \else
3340       \expandafter\expandafter\expandafter\set@hyphenmins
3341       \csname #1hyphenmins\endcsname
3342     \fi
3343     \the\toks@
3344     \toks@{}%
3345   \fi}

```

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

```

3346 \def\bbl@get@enc#1:#2:#3\@@{\def\bbl@hyph@enc{#2}}

```

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

```

3347 \def\bbl@hook@everylanguage#1{}
3348 \def\bbl@hook@loadpatterns#1{\input #1\relax}
3349 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
3350 \let\bbl@hook@loadkernel\bbl@hook@loadpatterns
3351 \begingroup
3352   \def\AddBabelHook#1#2{%
3353     \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
3354       \def\next{\toks1}%

```

```

3355 \else
3356 \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname###1}%
3357 \fi
3358 \next}
3359 \ifx\directlua\undefined
3360 \ifx\XeTeXinputencoding\undefined\else
3361 \input xebabel.def
3362 \fi
3363 \else
3364 \input luababel.def
3365 \fi
3366 \openin1 = babel-\bbl@format.cfg
3367 \ifeof1
3368 \else
3369 \input babel-\bbl@format.cfg\relax
3370 \fi
3371 \closein1
3372 \endgroup
3373 \bbl@hook@loadkernel{switch.def}

```

\readconfigfile The configuration file can now be opened for reading.

```

3374 \openin1 = language.dat

```

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

```

3375 \def\language{english}%
3376 \ifeof1
3377 \message{I couldn't find the file language.dat,\space
3378 I will try the file hyphen.tex}
3379 \input hyphen.tex\relax
3380 \chardef\l@english\z@
3381 \else

```

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

```

3382 \last@language\m@ne

```

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

```

3383 \loop

```

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

```

3384 \endlinechar\m@ne
3385 \read1 to \bbl@line
3386 \endlinechar\^^M

```

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

```

3387 \if T\ifeof1F\fi T\relax
3388 \ifx\bbl@line\@empty\else
3389 \edef\bbl@line{\bbl@line\space\space\space}%
3390 \expandafter\process@line\bbl@line\relax
3391 \fi
3392 \repeat

```

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

```

3393 \begingroup
3394   \def\bbl@elt#1#2#3#4{%
3395     \global\language=#2\relax
3396     \gdef\language{#1}%
3397     \def\bbl@elt##1##2##3##4{}}%
3398   \bbl@languages
3399 \endgroup
3400 \fi

```

and close the configuration file.

```

3401 \closein1

```

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

```

3402 \if/\the\toks@/\else
3403   \errhelp{language.dat loads no language, only synonyms}
3404   \errmessage{Orphan language synonym}
3405 \fi

```

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

```

3406 \let\bbl@line\@undefined
3407 \let\process@line\@undefined
3408 \let\process@synonym\@undefined
3409 \let\process@language\@undefined
3410 \let\bbl@get@enc\@undefined
3411 \let\bbl@hyph@enc\@undefined
3412 \let\bbl@tempa\@undefined
3413 \let\bbl@hook@loadkernel\@undefined
3414 \let\bbl@hook@everylanguage\@undefined
3415 \let\bbl@hook@loadpatterns\@undefined
3416 \let\bbl@hook@loadexceptions\@undefined
3417 \</patterns>

```

Here the code for `iniTEX` ends.

14 Font handling with fontspec

Add the bidi handler just before `luaotfload`, 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].

```

3418 <<{*More package options}>> ≡
3419 \ifodd\bbl@engine
3420   \DeclareOption{bidi=basic-r}%
3421   {\ExecuteOptions{bidi=basic}}
3422   \DeclareOption{bidi=basic}%
3423   {\let\bbl@beforeforeign\leavevmode
3424     % TODO - to locale_props, not as separate attribute
3425     \newattribute\bbl@attr@dir
3426     % I don't like it, hackish:
3427     \frozen@everymath\expandafter{%
3428       \expandafter\bbl@mathboxdir\the\frozen@everymath}%
3429     \frozen@everydisplay\expandafter{%
3430       \expandafter\bbl@mathboxdir\the\frozen@everydisplay}%

```

```

3431 \bbl@exp{\output{\bodydir\pagedir\the\output}}%
3432 \AtEndOfPackage{\EnableBabelHook{babel-bidi}}%
3433 \else
3434 \DeclareOption{bidi=basic-r}%
3435 {\ExecuteOptions{bidi=basic}}
3436 \DeclareOption{bidi=basic}%
3437 {\bbl@error
3438 {The bidi method 'basic' is available only in\%
3439 luatex. I'll continue with 'bidi=default', so\%
3440 expect wrong results}%
3441 {See the manual for further details.}%
3442 \let\bbl@beforeforeign\leavevmode
3443 \AtEndOfPackage{%
3444 \EnableBabelHook{babel-bidi}%
3445 \bbl@xebidipar}}
3446 \def\bbl@loadxebidi#1{%
3447 \ifx\RTLfootnotetext\@undefined
3448 \AtEndOfPackage{%
3449 \EnableBabelHook{babel-bidi}%
3450 \ifx\fontspec\@undefined
3451 \usepackage{fontspec}% bidi needs fontspec
3452 \fi
3453 \usepackage#1{bidi}}%
3454 \fi}
3455 \DeclareOption{bidi=bidi}%
3456 {\bbl@tentative{bidi=bidi}%
3457 \bbl@loadxebidi{}}
3458 \DeclareOption{bidi=bidi-r}%
3459 {\bbl@tentative{bidi=bidi-r}%
3460 \bbl@loadxebidi{[rldocument]}}
3461 \DeclareOption{bidi=bidi-l}%
3462 {\bbl@tentative{bidi=bidi-l}%
3463 \bbl@loadxebidi{}}
3464 \fi
3465 \DeclareOption{bidi=default}%
3466 {\let\bbl@beforeforeign\leavevmode
3467 \ifodd\bbl@engine
3468 \newattribute\bbl@attr@dir
3469 \bbl@exp{\output{\bodydir\pagedir\the\output}}%
3470 \fi
3471 \AtEndOfPackage{%
3472 \EnableBabelHook{babel-bidi}%
3473 \ifodd\bbl@engine\else
3474 \bbl@xebidipar
3475 \fi}}
3476 <</More package options>>

```

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

```

3477 <<*Font selection>> ≡
3478 \bbl@trace{Font handling with fontspec}
3479 \@onlypreamble\babelfont
3480 \newcommand\babelfont[2][{}]{% 1=langs/scripts 2=fam
3481 \edef\bbl@tempa{#1}%
3482 \def\bbl@tempb{#2}% Used by \bbl@bblfont
3483 \ifx\fontspec\@undefined
3484 \usepackage{fontspec}%
3485 \fi

```



```

3486 \EnableBabelHook{babel-fontspec}% Just calls \bbl@switchfont
3487 \bbl@bblfont}
3488 \newcommand\bbl@bblfont[2][]{% 1=features 2=fontname, @font=rm|sf|tt
3489 \bbl@ifunset{\bbl@tempb family}%
3490 {\bbl@providedefam{\bbl@tempb}}%
3491 {\bbl@exp{%
3492   \\bbl@sreplace\<\bbl@tempb family >%
3493   {\@nameuse{\bbl@tempb default}}{\<\bbl@tempb default>}}}%
3494 % For the default font, just in case:
3495 \bbl@ifunset{\bbl@lsys@\language name}{\bbl@provide@lsys{\language name}}}%
3496 \expandafter\bbl@ifblank\expandafter{\bbl@tempa}%
3497 {\bbl@csarg\edef{\bbl@tempb dflt@}{<{#1}{#2}}% save bbl@rmdflt@
3498 \bbl@exp{%
3499   \let\<bbl@\bbl@tempb dflt@\language name>\<bbl@\bbl@tempb dflt@>%
3500   \\bbl@font@set\<bbl@\bbl@tempb dflt@\language name>%
3501   \<\bbl@tempb default>\<\bbl@tempb family>}}%
3502 {\bbl@foreach\bbl@tempa{% ie bbl@rmdflt@lang / *scrt
3503   \bbl@csarg\def{\bbl@tempb dflt@##1}{<{#1}{#2}}}}}%

```

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

```

3504 \def\bbl@providedefam#1{%
3505   \bbl@exp{%
3506     \\newcommand\<#1default>{}% Just define it
3507     \\bbl@add@list\\bbl@font@fams{#1}%
3508     \\DeclareRobustCommand\<#1family>%
3509     \\not@math@alphabet\<#1family>\relax
3510     \\fontfamily\<#1default>\\selectfont}%
3511     \\DeclareTextFontCommand{\<text#1>}{\<#1family>}}%

```

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

```

3512 \def\bbl@nostdfont#1{%
3513   \bbl@ifunset{\bbl@WFF@\f@family}%
3514   {\bbl@csarg\gdef{\bbl@WFF@\f@family}}{% Flag, to avoid dupl warns
3515     \bbl@infowarn{The current font is not a babel standard family:\\%
3516       #1%
3517       \fontname\font\\%
3518       There is nothing intrinsically wrong with this warning, and\\%
3519       you can ignore it altogether if you do not need these\\%
3520       families. But if they are used in the document, you should be\\%
3521       aware 'babel' will no set Script and Language for them, so\\%
3522       you may consider defining a new family with \string\babelfont.\\%
3523       See the manual for further details about \string\babelfont.\\%
3524       Reported}}
3525   }%
3526 \gdef\bbl@switchfont{%
3527   \bbl@ifunset{\bbl@lsys@\language name}{\bbl@provide@lsys{\language name}}}%
3528   \bbl@exp{% eg Arabic -> arabic
3529     \lowercase{\edef\\bbl@tempa{\bbl@cs{sname@\language name}}}%
3530   \bbl@foreach\bbl@font@fams{%
3531     \bbl@ifunset{\bbl@##1dflt@\language name}% (1) language?
3532     {\bbl@ifunset{\bbl@##1dflt@*\bbl@tempa}% (2) from script?
3533       {\bbl@ifunset{\bbl@##1dflt@}% 2=F - (3) from generic?
3534         }% 123=F - nothing!
3535         {\bbl@exp{% 3=T - from generic
3536           \global\let\<bbl@##1dflt@\language name>%
3537           \<bbl@##1dflt@>}}}%
3538       {\bbl@exp{% 2=T - from script
3539         \global\let\<bbl@##1dflt@\language name>%

```

```

3540         \<bbl@##1dflt@*bbl@tempa>}}}%
3541     {}}}%
3542     \def\bbl@tempa{\bbl@nostdfont{}}}%
3543     \bbl@foreach\bbl@font@fams{%      don't gather with prev for
3544         \bbl@ifunset{bbl@##1dflt@language}%
3545         {\bbl@cs{famrst@##1}%
3546         \global\bbl@csarg\let{famrst@##1}\relax}%
3547         {\bbl@exp{% order is relevant
3548             \\bbl@add\\originalTeX{%
3549                 \\bbl@font@rst{\bbl@cs{##1dflt@language}}}%
3550                 \<##1default>\<##1family>{##1}}}%
3551             \\bbl@font@set\bbl@##1dflt@language>% the main part!
3552             \<##1default>\<##1family>}}}%
3553     \bbl@ifrestoring{}}{\bbl@tempa}}}%

```

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

```

3554 \ifx\fontfamily\undefined\else      % if latex
3555     \ifcase\bbl@engine                % if pdftex
3556         \let\bbl@ckeckstdfonts\relax
3557     \else
3558         \def\bbl@ckeckstdfonts{%
3559             \begingroup
3560             \global\let\bbl@ckeckstdfonts\relax
3561             \let\bbl@tempa\@empty
3562             \bbl@foreach\bbl@font@fams{%
3563                 \bbl@ifunset{bbl@##1dflt@}%
3564                 {\nameuse{##1family}%
3565                 \bbl@csarg\gdef{WFF@f@family}}}% Flag
3566                 \bbl@exp{\\bbl@add\\bbl@tempa{* \<##1family>= \fontfamily\\
3567                     \space\space\fontname\font\\}}}%
3568                 \bbl@csarg\xdef{##1dflt@}{\fontfamily}%
3569                 \expandafter\xdef\csname ##1default\endcsname{\fontfamily}}}%
3570             {}}}%
3571     \ifx\bbl@tempa\@empty\else
3572         \bbl@infowarn{The following fonts are not babel standard families:\\
3573             \bbl@tempa
3574             There is nothing intrinsically wrong with it, but\\
3575             'babel' will no set Script and Language. Consider\\
3576             defining a new family with \string\babelfont.\\
3577             Reported}%
3578     \fi
3579 \endgroup}
3580 \fi
3581 \fi

```

Now the macros defining the font with `fontspec`.

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

```

3582 \def\bbl@font@set#1#2#3{% eg \bbl@rmdflt@lang \rmdefault \rmfamily
3583     \bbl@xin@{<>}{#1}%
3584     \ifin@
3585         \bbl@exp{\\bbl@fontspec@set\\#1\expandafter\@gobbletwo#1\\#3}%
3586     \fi
3587     \bbl@exp{%
3588         \def\\#2{#1}%          eg, \rmdefault{\bbl@rmdflt@lang}
3589         \\bbl@ifsamestring{#2}{\fontfamily}{\\#3\let\\bbl@tempa\relax}}}%

```

```

3590%      TODO - next should be global?, but even local does its job. I'm
3591%      still not sure -- must investigate:
3592\def\bb1@fontspec@set#1#2#3#4{% eg \bb1@rmdflt@lang fnt-opt fnt-nme \xxfamily
3593  \let\bb1@tempe\bb1@mapselect
3594  \let\bb1@mapselect\relax
3595  \let\bb1@temp@fam#4%      eg, '\rmfamily', to be restored below
3596  \let#4\relax      % So that can be used with \newfontfamily
3597  \bb1@exp{%
3598    \let\\bb1@temp@pfam\<\bb1@stripslash#4\space>% eg, '\rmfamily '
3599    \<keys_if_exist:nnF>{fontspec-opentype}%
3600      {Script/\bb1@cs{sname@\language}}}%
3601      {\newfontscript{\bb1@cs{sname@\language}}}%
3602      {\bb1@cs{sotf@\language}}}%
3603    \<keys_if_exist:nnF>{fontspec-opentype}%
3604      {Language/\bb1@cs{lname@\language}}}%
3605      {\newfontlanguage{\bb1@cs{lname@\language}}}%
3606      {\bb1@cs{lotf@\language}}}%
3607    \newfontfamily\\#4%
3608    [\bb1@cs{lsys@\language},#2]{#3}% ie \bb1@exp{.}{#3}
3609  \begingroup
3610    #4%
3611    \xdef#1{f@family}%      eg, \bb1@rmdflt@lang{FreeSerif(0)}
3612  \endgroup
3613  \let#4\bb1@temp@fam
3614  \bb1@exp{\let\<\bb1@stripslash#4\space>}\bb1@temp@pfam
3615  \let\bb1@mapselect\bb1@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.

```

3616\def\bb1@font@rst#1#2#3#4{%
3617  \bb1@csarg\def{famrst@#4}{\bb1@font@set{#1}#2#3}}

```

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

```

3618\def\bb1@font@fams{rm,sf,tt}

```

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

```

3619\newcommand\babelFSstore[2][{%
3620  \bb1@ifblank{#1}%
3621    {\bb1@csarg\def{sname@#2}{Latin}}%
3622    {\bb1@csarg\def{sname@#2}{#1}}%
3623  \bb1@provide@dirs{#2}%
3624  \bb1@csarg\ifnum{wdir@#2}>\z@
3625    \let\bb1@beforeforeign\leavevmode
3626    \EnableBabelHook{babel-bidi}%
3627  \fi
3628  \bb1@foreach{#2}{%
3629    \bb1@FSstore{##1}{rm}\rmdefault\bb1@save@rmdefault
3630    \bb1@FSstore{##1}{sf}\sfdefault\bb1@save@sfdefault
3631    \bb1@FSstore{##1}{tt}\ttdefault\bb1@save@ttdefault}}
3632\def\bb1@FSstore#1#2#3#4{%
3633  \bb1@csarg\edef{#2default#1}{#3}%
3634  \expandafter\addto\csname extras#1\endcsname{%
3635    \let#4#3%
3636    \ifx#3\f@family
3637      \edef#3{\csname \bb1@#2default#1\endcsname}%
3638    \fontfamily{#3}\selectfont

```

```

3639 \else
3640 \edef#3{\csname bbl@#2default#1\endcsname}%
3641 \fi}%
3642 \expandafter\addto\csname noextras#1\endcsname{%
3643 \ifx#3\fontfamily
3644 \fontfamily{#4}\selectfont
3645 \fi
3646 \let#3#4}}
3647 \let\bbl@langfeatures\@empty
3648 \def\babelFSfeatures{% make sure \fontspec is redefined once
3649 \let\bbl@ori@fontspec\fontspec
3650 \renewcommand\fontspec[1][{}]{%
3651 \bbl@ori@fontspec[\bbl@langfeatures##1]}
3652 \let\babelFSfeatures\bbl@FSfeatures
3653 \babelFSfeatures}
3654 \def\bbl@FSfeatures#1#2{%
3655 \expandafter\addto\csname extras#1\endcsname{%
3656 \babel@save\bbl@langfeatures
3657 \edef\bbl@langfeatures{#2,}}
3658 <</Font selection>>

```

15 Hooks for XeTeX and LuaTeX

15.1 XeTeX

Unfortunately, the current encoding cannot be retrieved and therefore it is reset always to `utf8`, which seems a sensible default.

\LaTeX sets many “codes” just before loading `hyphen.cfg`. That is not a problem in `luatex`, but in `xetex` they must be reset to the proper value. Most of the work is done in `xe(la)tex.ini`, so here we just “undo” some of the changes done by \LaTeX . Anyway, for consistency `LuaTeX` also resets the catcodes.

```

3659 <<*Restore Unicode catcodes before loading patterns>> ≡
3660 \begingroup
3661 % Reset chars "80-"C0 to category "other", no case mapping:
3662 \catcode`\@=11 \count@=128
3663 \loop\ifnum\count@<192
3664 \global\uccode\count@=0 \global\lccode\count@=0
3665 \global\catcode\count@=12 \global\sffcode\count@=1000
3666 \advance\count@ by 1 \repeat
3667 % Other:
3668 \def\O ##1 {%
3669 \global\uccode"##1=0 \global\lccode"##1=0
3670 \global\catcode"##1=12 \global\sffcode"##1=1000 }%
3671 % Letter:
3672 \def\L ##1 ##2 ##3 {\global\catcode"##1=11
3673 \global\uccode"##1="##2
3674 \global\lccode"##1="##3
3675 % Uppercase letters have sffcode=999:
3676 \ifnum"##1="##3 \else \global\sffcode"##1=999 \fi }%
3677 % Letter without case mappings:
3678 \def\l ##1 {\L ##1 ##1 ##1 }%
3679 \l 00AA
3680 \L 00B5 039C 00B5
3681 \l 00BA
3682 \O 00D7
3683 \l 00DF
3684 \O 00F7

```

```

3685 \L 00FF 0178 00FF
3686 \endgroup
3687 \input #1\relax
3688 <</Restore Unicode catcodes before loading patterns>>

```

Some more common code.

```

3689 <<(*Footnote changes)>> ≡
3690 \bbl@trace{Bidi footnotes}
3691 \ifx\bbl@beforeforeign\leavevmode
3692 \def\bbl@footnote#1#2#3{%
3693 \@ifnextchar[%
3694 {\bbl@footnote@o{#1}{#2}{#3}}%
3695 {\bbl@footnote@x{#1}{#2}{#3}}}
3696 \def\bbl@footnote@x#1#2#3#4{%
3697 \bgroup
3698 \select@language@x{\bbl@main@language}%
3699 \bbl@fn@footnote{#2#1{\ignorespaces#4}#3}%
3700 \egroup}
3701 \def\bbl@footnote@o#1#2#3[#4]#5{%
3702 \bgroup
3703 \select@language@x{\bbl@main@language}%
3704 \bbl@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}%
3705 \egroup}
3706 \def\bbl@footnotetext#1#2#3{%
3707 \@ifnextchar[%
3708 {\bbl@footnotetext@o{#1}{#2}{#3}}%
3709 {\bbl@footnotetext@x{#1}{#2}{#3}}}
3710 \def\bbl@footnotetext@x#1#2#3#4{%
3711 \bgroup
3712 \select@language@x{\bbl@main@language}%
3713 \bbl@fn@footnotetext{#2#1{\ignorespaces#4}#3}%
3714 \egroup}
3715 \def\bbl@footnotetext@o#1#2#3[#4]#5{%
3716 \bgroup
3717 \select@language@x{\bbl@main@language}%
3718 \bbl@fn@footnotetext[#4]{#2#1{\ignorespaces#5}#3}%
3719 \egroup}
3720 \def\BabelFootnote#1#2#3#4{%
3721 \ifx\bbl@fn@footnote\@undefined
3722 \let\bbl@fn@footnote\footnote
3723 \fi
3724 \ifx\bbl@fn@footnotetext\@undefined
3725 \let\bbl@fn@footnotetext\footnotetext
3726 \fi
3727 \bbl@ifblank{#2}%
3728 {\def#1{\bbl@footnote{\@firstofone}{#3}{#4}}
3729 \@namedef{\bbl@stripslash#1text}%
3730 {\bbl@footnotetext{\@firstofone}{#3}{#4}}}%
3731 {\def#1{\bbl@exp{\bbl@footnote{\foreignlanguage{#2}}}{#3}{#4}}%
3732 \@namedef{\bbl@stripslash#1text}%
3733 {\bbl@exp{\bbl@footnotetext{\foreignlanguage{#2}}}{#3}{#4}}}%
3734 \fi
3735 <</Footnote changes>>

```

Now, the code.

```

3736 (*xetex)
3737 \def\BabelStringsDefault{unicode}
3738 \let\xebbl@stop\relax
3739 \AddBabelHook{xetex}{encodedcommands}{%

```

```

3740 \def\bb1@tempa{#1}%
3741 \ifx\bb1@tempa\@empty
3742 \XeTeXinputencoding"bytes"%
3743 \else
3744 \XeTeXinputencoding"#1"%
3745 \fi
3746 \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
3747 \AddBabelHook{xetex}{stopcommands}{%
3748 \xebbl@stop
3749 \let\xebbl@stop\relax}
3750 \def\bb1@intraspace#1 #2 #3\@@{%
3751 \bb1@csarg\gdef{\xeisp@\bb1@cs{sbc}\language}}%
3752 {\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}}
3753 \def\bb1@intrapenalty#1\@@{%
3754 \bb1@csarg\gdef{\xeipn@\bb1@cs{sbc}\language}}%
3755 {\XeTeXlinebreakpenalty #1\relax}}
3756 \def\bb1@provide@intraspace{%
3757 \bb1@xin@{\bb1@cs{sbc}\language}}{Thai,Lao,Khmr}%
3758 \ifin@ % sea (currently ckj not handled)
3759 \bb1@ifunset{\bb1@intsp@\language}}{%
3760 \expandafter\ifx\csname \bb1@intsp@\language\endcsname\@empty\else
3761 \ifx\bb1@KVP@intraspace\@nil
3762 \bb1@exp{%
3763 \bb1@intraspace\bb1@cs{intsp@\language}\@@}%
3764 \fi
3765 \ifx\bb1@KVP@intrapenalty\@nil
3766 \bb1@intrapenalty0\@@
3767 \fi
3768 \fi
3769 \ifx\bb1@KVP@intraspace\@nil\else % We may override the ini
3770 \expandafter\bb1@intraspace\bb1@KVP@intraspace\@@
3771 \fi
3772 \ifx\bb1@KVP@intrapenalty\@nil\else
3773 \expandafter\bb1@intrapenalty\bb1@KVP@intrapenalty\@@
3774 \fi
3775 \ifx\bb1@ispace\@undefined
3776 \AtBeginDocument{%
3777 \expandafter\bb1@add
3778 \csname selectfont\endcsname{\bb1@ispace}}%
3779 \def\bb1@ispace{\bb1@cs{\xeisp@\bb1@cs{sbc}\language}}}%
3780 \fi}%
3781 \fi}
3782 \AddBabelHook{xetex}{loadkernel}{%
3783 <<Restore Unicode catcodes before loading patterns>>}
3784 \ifx\DisableBabelHook\@undefined\endinput\fi
3785 \AddBabelHook{babel-fontspec}{afterextras}{\bb1@switchfont}
3786 \AddBabelHook{babel-fontspec}{beforestart}{\bb1@ckeckstdfonts}
3787 \DisableBabelHook{babel-fontspec}
3788 <<Font selection>>
3789 \input txtbabel.def
3790 </xetex>

```

15.2 Layout

In progress.

Note elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titlesp, and geometry.

\bb1@startskip and \bb1@endskip are available to package authors. Thanks to the T_EX

expansion mechanism the following constructs are valid: `\adim\bbbl@startskip`,
`\advance\bbbl@startskip\adim`, `\bbbl@startskip\adim`.
Consider `txtbabel` as a shorthand for *tex-xet babel*, which is the bidi model in both `pdfTeX`
and `xetex`.

```

3791 (*texxet)
3792 \providecommand\bbbl@provide@intraspace{}
3793 \bbbl@trace{Redefinitions for bidi layout}
3794 \def\bbbl@sspre@caption{%
3795   \bbbl@exp{\everyhbox{\bbbl@textdir\bbbl@cs{wdir@\bbbl@main@language}}}}
3796 \ifx\bbbl@opt@layout\@nnil\endinput\fi % No layout
3797 \def\bbbl@startskip{\ifcase\bbbl@thepardir\leftskip\else\rightskip\fi}
3798 \def\bbbl@endskip{\ifcase\bbbl@thepardir\rightskip\else\leftskip\fi}
3799 \ifx\bbbl@beforeforeign\leavevmode % A poor test for bidi=
3800   \def\hangfrom#1{%
3801     \setbox\@tempboxa\hbox{#1}%
3802     \hangindent\ifcase\bbbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
3803     \noindent\box\@tempboxa}
3804 \def\raggedright{%
3805   \let\@centercr
3806   \bbbl@startskip\z@skip
3807   \@rightskip\@flushglue
3808   \bbbl@endskip\@rightskip
3809   \parindent\z@
3810   \parfillskip\bbbl@startskip}
3811 \def\raggedleft{%
3812   \let\@centercr
3813   \bbbl@startskip\@flushglue
3814   \bbbl@endskip\z@skip
3815   \parindent\z@
3816   \parfillskip\bbbl@endskip}
3817 \fi
3818 \IfBabelLayout{lists}
3819   {\bbbl@sreplace\list
3820     {\@totalleftmargin\leftmargin}{\@totalleftmargin\bbbl@listleftmargin}%
3821     \def\bbbl@listleftmargin{%
3822       \ifcase\bbbl@thepardir\leftmargin\else\rightmargin\fi}%
3823     \ifcase\bbbl@engine
3824       \def\labelenumii{}\theenumii{}\pdfTeX doesn't reverse ()
3825       \def\p@enumiii{\p@enumii}\theenumii{}
3826     \fi
3827     \bbbl@sreplace\@verbatim
3828     {\leftskip\@totalleftmargin}%
3829     {\bbbl@startskip\textwidth
3830       \advance\bbbl@startskip-\linewidth}%
3831     \bbbl@sreplace\@verbatim
3832     {\rightskip\z@skip}%
3833     {\bbbl@endskip\z@skip}}%
3834   {}
3835 \IfBabelLayout{contents}
3836   {\bbbl@sreplace\@dottedtocline{\leftskip}{\bbbl@startskip}%
3837     \bbbl@sreplace\@dottedtocline{\rightskip}{\bbbl@endskip}}
3838   {}
3839 \IfBabelLayout{columns}
3840   {\bbbl@sreplace\@outputdblcol{\hb@xt@\textwidth}{\bbbl@outputbox}%
3841     \def\bbbl@outputbox#1{%
3842       \hb@xt@\textwidth{%
3843         \hskip\columnwidth
3844         \hfil

```

```

3845      {\normalcolor\vrule \@width\columnseprule}%
3846      \hfil
3847      \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
3848      \hskip-\textwidth
3849      \hb@xt@\columnwidth{\box\@outputbox \hss}%
3850      \hskip\columnsep
3851      \hskip\columnwidth}}}%
3852  {}
3853  <<Footnote changes>>
3854  \IfBabelLayout{footnotes}%
3855  {\BabelFootnote\footnote\language\language{}{}}%
3856  \BabelFootnote\localfootnote\language\language{}{}}%
3857  \BabelFootnote\mainfootnote{}{}}{}
3858  {}

```

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.

```

3859 \IfBabelLayout{counters}%
3860  {\let\bb@latinarabic=\@arabic
3861   \def\@arabic#1{\babelsublr{\bb@latinarabic#1}}}%
3862   \let\bb@asciroman=\@roman
3863   \def\@roman#1{\babelsublr{\ensureascii{\bb@asciroman#1}}}%
3864   \let\bb@asciiRoman=\@Roman
3865   \def\@Roman#1{\babelsublr{\ensureascii{\bb@asciiRoman#1}}}}{}
3866  </texxet>

```

15.3 LuaTeX

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

The names `\l@<language>` are defined and take some value from the beginning because all `ldf` files assume this for the corresponding language to be considered valid, but patterns are not loaded (except the first one). This is done later, when the language is first selected (which usually means when the `ldf` finishes). If a language has been loaded, `\bb@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, they are added, but note if the language patterns have not been preloaded they won't at run time.

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

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

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

We need catcode tables, but no format (targeted by `babel`) provide a command to allocate them (although there are packages like `ctablestack`). For the moment, a dangerous

approach is used – just allocate a high random number and cross the fingers. To complicate things, etex.sty changes the way languages are allocated.

```

3867 (*luatex)
3868 \ifx\AddBabelHook\@undefined
3869 \bbl@trace{Read language.dat}
3870 \ifx\bbl@readstream\@undefined
3871   \csname newread\endcsname\bbl@readstream
3872 \fi
3873 \begingroup
3874   \toks@{}
3875   \count@\z@ % 0=start, 1=0th, 2=normal
3876   \def\bbl@process@line#1#2 #3 #4 {%
3877     \ifx=#1%
3878       \bbl@process@synonym{#2}%
3879     \else
3880       \bbl@process@language{#1#2}{#3}{#4}%
3881     \fi
3882     \ignorespaces}
3883   \def\bbl@manylang{%
3884     \ifnum\bbl@last>\@ne
3885       \bbl@info{Non-standard hyphenation setup}%
3886     \fi
3887     \let\bbl@manylang\relax}
3888   \def\bbl@process@language#1#2#3{%
3889     \ifcase\count@
3890       \@ifundefined{zth#1}{\count@\tw@}{\count@\@ne}%
3891     \or
3892       \count@\tw@
3893     \fi
3894     \ifnum\count@=\tw@
3895       \expandafter\addlanguage\csname l@#1\endcsname
3896       \language\allocationnumber
3897       \chardef\bbl@last\allocationnumber
3898       \bbl@manylang
3899       \let\bbl@elt\relax
3900       \xdef\bbl@languages{%
3901         \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
3902     \fi
3903     \the\toks@
3904     \toks@{}}
3905   \def\bbl@process@synonym@aux#1#2{%
3906     \global\expandafter\chardef\csname l@#1\endcsname#2\relax
3907     \let\bbl@elt\relax
3908     \xdef\bbl@languages{%
3909       \bbl@languages\bbl@elt{#1}{#2}{}}}%
3910   \def\bbl@process@synonym#1{%
3911     \ifcase\count@
3912       \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
3913     \or
3914       \@ifundefined{zth#1}{\bbl@process@synonym@aux{#1}{}}}%
3915     \else
3916       \bbl@process@synonym@aux{#1}{\the\bbl@last}%
3917     \fi}
3918   \ifx\bbl@languages\@undefined % Just a (sensible?) guess
3919     \chardef\l@english\z@
3920     \chardef\l@USenglish\z@
3921     \chardef\bbl@last\z@
3922     \global\@namedef{bbl@hyphendata@0}{\hyphen.tex}}

```

```

3923 \gdef\bbl@languages{%
3924     \bbl@elt{english}{0}{hyphen.tex}}%
3925     \bbl@elt{USenglish}{0}{}%}}
3926 \else
3927     \global\let\bbl@languages@format\bbl@languages
3928     \def\bbl@elt#1#2#3#4{% Remove all except language 0
3929         \ifnum#2>\z@ \else
3930             \noexpand\bbl@elt{#1}{#2}{#3}{#4}%
3931         \fi}%
3932     \xdef\bbl@languages{\bbl@languages}%
3933 \fi
3934 \def\bbl@elt#1#2#3#4{\@namedef{zth#1}{} } % Define flags
3935 \bbl@languages
3936 \openin\bbl@readstream=language.dat
3937 \ifeof\bbl@readstream
3938     \bbl@warning{I couldn't find language.dat. No additional\\%
3939                 patterns loaded. Reported}%
3940 \else
3941     \loop
3942         \endlinechar\m@ne
3943         \read\bbl@readstream to \bbl@line
3944         \endlinechar\^^M
3945         \if T\ifeof\bbl@readstream F\fi T\relax
3946         \ifx\bbl@line\@empty\else
3947             \edef\bbl@line{\bbl@line\space\space\space}%
3948             \expandafter\bbl@process@line\bbl@line\relax
3949         \fi
3950     \repeat
3951 \fi
3952 \endgroup
3953 \bbl@trace{Macros for reading patterns files}
3954 \def\bbl@get@enc#1:#2:#3\@@{\def\bbl@hyph@enc{#2}}
3955 \ifx\babelcatcodetablenum\@undefined
3956     \def\babelcatcodetablenum{5211}
3957 \fi
3958 \def\bbl@luapatterns#1#2{%
3959     \bbl@get@enc#1::\@@
3960     \setbox\z@\hbox\bgroup
3961         \begingroup
3962             \ifx\catcodetable\@undefined
3963                 \let\savecatcodetable\luatexsavecatcodetable
3964                 \let\initcatcodetable\luatexinitcatcodetable
3965                 \let\catcodetable\luatexcatcodetable
3966             \fi
3967             \savecatcodetable\babelcatcodetablenum\relax
3968             \initcatcodetable\numexpr\babelcatcodetablenum+1\relax
3969             \catcodetable\numexpr\babelcatcodetablenum+1\relax
3970             \catcode`\#=6 \catcode`\$=3 \catcode`\&=4 \catcode`\^=7
3971             \catcode`\_ =8 \catcode`\{=1 \catcode`\}=2 \catcode`\~ =13
3972             \catcode`\@ =11 \catcode`\^^I=10 \catcode`\^^J=12
3973             \catcode`\<=12 \catcode`\>=12 \catcode`\*=12 \catcode`\.=12
3974             \catcode`\-=12 \catcode`\/=12 \catcode`\[=12 \catcode`\]=12
3975             \catcode`\`=12 \catcode`\'=12 \catcode`\`=12
3976             \input #1\relax
3977             \catcodetable\babelcatcodetablenum\relax
3978         \endgroup
3979     \def\bbl@tempa{#2}%
3980     \ifx\bbl@tempa\@empty\else
3981         \input #2\relax

```

```

3982 \fi
3983 \egroup}%
3984 \def\bbl@patterns@lua#1{%
3985 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
3986 \csname l@#1\endcsname
3987 \edef\bbl@tempa{#1}%
3988 \else
3989 \csname l@#1:\f@encoding\endcsname
3990 \edef\bbl@tempa{#1:\f@encoding}%
3991 \fi\relax
3992 \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
3993 \@ifundefined{bbl@hyphendata@the\language}%
3994 {\def\bbl@elt##1##2##3##4{%
3995 \ifnum##2=\csname l@bbl@tempa\endcsname % #2=spanish, dutch:OT1...
3996 \def\bbl@tempb{##3}%
3997 \ifx\bbl@tempb\empty\else % if not a synonymous
3998 \def\bbl@tempc{##3}{##4}%
3999 \fi
4000 \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4001 \fi}%
4002 \bbl@languages
4003 \@ifundefined{bbl@hyphendata@the\language}%
4004 {\bbl@info{No hyphenation patterns were set for\%
4005 language '\bbl@tempa'. Reported}}%
4006 {\expandafter\expandafter\expandafter\bbl@luapatterns
4007 \csname bbl@hyphendata@the\language\endcsname}}}%
4008 \endinput\fi
4009 \begingroup
4010 \catcode`\%=12
4011 \catcode`\'=12
4012 \catcode`\%=12
4013 \catcode`\:=12
4014 \directlua{
4015 Babel = Babel or {}
4016 function Babel.bytes(line)
4017 return line:gsub(".",
4018 function (chr) return unicode.utf8.char(string.byte(chr)) end)
4019 end
4020 function Babel.begin_process_input()
4021 if luatexbase and luatexbase.add_to_callback then
4022 luatexbase.add_to_callback('process_input_buffer',
4023 Babel.bytes,'Babel.bytes')
4024 else
4025 Babel.callback = callback.find('process_input_buffer')
4026 callback.register('process_input_buffer',Babel.bytes)
4027 end
4028 end
4029 function Babel.end_process_input ()
4030 if luatexbase and luatexbase.remove_from_callback then
4031 luatexbase.remove_from_callback('process_input_buffer','Babel.bytes')
4032 else
4033 callback.register('process_input_buffer',Babel.callback)
4034 end
4035 end
4036 function Babel.addpatterns(pp, lg)
4037 local lg = lang.new(lg)
4038 local pats = lang.patterns(lg) or ''
4039 lang.clear_patterns(lg)
4040 for p in pp:gmatch('[^%s]+') do

```

```

4041     ss = ''
4042     for i in string.utfcharacters(p:gsub('%d', '')) do
4043         ss = ss .. '%d?' .. i
4044     end
4045     ss = ss:gsub('^%%d%?%', '%%.') .. '%d?'
4046     ss = ss:gsub('%.%%d%?$', '%%.')
4047     pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
4048     if n == 0 then
4049         tex.sprint(
4050             [[\string\csname\space bbl@info\endcsname{New pattern: }]]
4051             .. p .. [[]])
4052         pats = pats .. ' ' .. p
4053     else
4054         tex.sprint(
4055             [[\string\csname\space bbl@info\endcsname{Renew pattern: }]]
4056             .. p .. [[]])
4057     end
4058 end
4059 lang.patterns(lg, pats)
4060 end
4061 }
4062 \endgroup
4063 \ifx\newattribute\@undefined\else
4064   \newattribute\bbl@attr@locale
4065   \AddBabelHook{luatex}{beforeextras}{%
4066     \setattribute\bbl@attr@locale\localeid}
4067 \fi
4068 \def\BabelStringsDefault{unicode}
4069 \let\luabbl@stop\relax
4070 \AddBabelHook{luatex}{encodedcommands}{%
4071   \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
4072   \ifx\bbl@tempa\bbl@tempb\else
4073     \directlua{Babel.begin_process_input()}%
4074     \def\luabbl@stop{%
4075       \directlua{Babel.end_process_input()}}%
4076   \fi}%
4077 \AddBabelHook{luatex}{stopcommands}{%
4078   \luabbl@stop
4079   \let\luabbl@stop\relax}
4080 \AddBabelHook{luatex}{patterns}{%
4081   \@ifundefined{bbl@hyphendata@the\language}%
4082     {\def\bbl@elt##1##2##3##4{%
4083       \ifnum##2=\csname l@#2\endcsname % #2=spanish, dutch:OT1...
4084       \def\bbl@tempb{##3}%
4085       \ifx\bbl@tempb\@empty\else % if not a synonymous
4086         \def\bbl@tempc{##3}{##4}%
4087       \fi
4088       \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4089     \fi}%
4090   \bbl@languages
4091   \@ifundefined{bbl@hyphendata@the\language}%
4092     {\bbl@info{No hyphenation patterns were set for\%
4093       language '#2'. Reported}}%
4094     {\expandafter\expandafter\expandafter\bbl@luapatterns
4095       \csname bbl@hyphendata@the\language\endcsname}}}%
4096   \@ifundefined{bbl@patterns@}{%
4097     \begingroup
4098     \bbl@xin@{,\number\language,}{,\bbl@pttnlist}%
4099     \ifin@else

```

```

4100     \ifx\bbbl@patterns@\@empty\else
4101         \directlua{ Babel.addpatterns(
4102             [[\bbbl@patterns@]], \number\language) }%
4103     \fi
4104     \@ifundefined{bbbl@patterns@#1}%
4105     \@empty
4106     {\directlua{ Babel.addpatterns(
4107         [[\space\csname bbl@patterns@#1\endcsname]],
4108         \number\language) }}%
4109     \xdef\bbbl@pttnlist{\bbbl@pttnlist\number\language,}%
4110 \fi
4111 \endgroup}}
4112 \AddBabelHook{luatex}{everylanguage}{%
4113 \def\process@language##1##2##3{%
4114 \def\process@line####1####2 ####3 ####4 {}}}
4115 \AddBabelHook{luatex}{loadpatterns}{%
4116 \input #1\relax
4117 \expandafter\gdef\csname bbl@hyphendata@\the\language\endcsname
4118     {\#1}{}}
4119 \AddBabelHook{luatex}{loadexceptions}{%
4120 \input #1\relax
4121 \def\bbbl@tempb##1##2{\#1}{\#1}%
4122 \expandafter\xdef\csname bbl@hyphendata@\the\language\endcsname
4123     {\expandafter\expandafter\expandafter\bbbl@tempb
4124     \csname bbl@hyphendata@\the\language\endcsname}}

```

\babelpatterns This macro adds patterns. Two macros are used to store them: `\bbbl@patterns@` for the global ones and `\bbbl@patterns@<lang>` for language ones. We make sure there is a space between words when multiple commands are used.

```

4125 \@onlypreamble\babelpatterns
4126 \AtEndOfPackage{%
4127 \newcommand\babelpatterns[2][\@empty]{%
4128     \ifx\bbbl@patterns@\relax
4129         \let\bbbl@patterns@\@empty
4130     \fi
4131     \ifx\bbbl@pttnlist@\@empty\else
4132         \bbbl@warning{%
4133             You must not intermingle \string\selectlanguage\space and\\%
4134             \string\babelpatterns\space or some patterns will not\\%
4135             be taken into account. Reported}%
4136     \fi
4137     \ifx\@empty#1%
4138         \protected@edef\bbbl@patterns@\{\bbbl@patterns@\space#2}%
4139     \else
4140         \edef\bbbl@tempb{\zap@space#1 \@empty}%
4141         \bbbl@for\bbbl@tempa\bbbl@tempb{%
4142             \bbbl@fixname\bbbl@tempa
4143             \bbbl@iflanguage\bbbl@tempa{%
4144                 \bbbl@csarg\protected@edef{patterns@\bbbl@tempa}{%
4145                     \@ifundefined{bbbl@patterns@\bbbl@tempa}%
4146                     \@empty
4147                     {\csname bbl@patterns@\bbbl@tempa\endcsname\space}%
4148                     #2}}}%
4149     \fi}}

```

15.4 Southeast Asian scripts

In progress. Replace regular (ie, implicit) discretionaries by spaceskips, based on the previous glyph (which I think makes sense, because the hyphen and the previous char go always together). Other discretionaries are not touched.

For the moment, only 3 SA languages are activated by default (see Unicode UAX 14).

```
4150 \directlua{
4151   Babel = Babel or {}
4152   Babel.linebreaking = Babel.linebreaking or {}
4153   Babel.linebreaking.before = {}
4154   Babel.linebreaking.after = {}
4155   Babel.locale = {} % Free to use, indexed with \localeid
4156   function Babel.linebreaking.add_before(func)
4157     tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
4158     table.insert(Babel.linebreaking.before, func)
4159   end
4160   function Babel.linebreaking.add_after(func)
4161     tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
4162     table.insert(Babel.linebreaking.after, func)
4163   end
4164 }
4165 \def\bbl@intraspace#1 #2 #3\@@{%
4166   \directlua{
4167     Babel = Babel or {}
4168     Babel.intraspace = Babel.intraspace or {}
4169     Babel.intraspace['\csname bbl@sbc@language\endcsname'] = %
4170       {b = #1, p = #2, m = #3}
4171     Babel.locale_props[\the\localeid].intraspace = %
4172       {b = #1, p = #2, m = #3}
4173   }}
4174 \def\bbl@intrapenalty#1\@@{%
4175   \directlua{
4176     Babel = Babel or {}
4177     Babel.intrapenalties = Babel.intrapenalties or {}
4178     Babel.intrapenalties['\csname bbl@sbc@language\endcsname'] = #1
4179     Babel.locale_props[\the\localeid].intrapenalty = #1
4180   }}
4181 \begingroup
4182 \catcode`\%=12
4183 \catcode`\^=14
4184 \catcode`\'=12
4185 \catcode`\~=12
4186 \gdef\bbl@seaintraspace{^
4187   \let\bbl@seaintraspace\relax
4188   \directlua{
4189     Babel = Babel or {}
4190     Babel.sea_enabled = true
4191     Babel.sea_ranges = Babel.sea_ranges or {}
4192     function Babel.set_chranges (script, chrng)
4193       local c = 0
4194       for s, e in string.gmatch(chrng..' ', '(.)%.%.(-)%s') do
4195         Babel.sea_ranges[script..c]={tonumber(s,16), tonumber(e,16)}
4196         c = c + 1
4197       end
4198     end
4199     function Babel.sea_disc_to_space (head)
4200       local sea_ranges = Babel.sea_ranges
4201       local last_char = nil
```

```

4202     local quad = 655360      ^^ 10 pt = 655360 = 10 * 65536
4203   for item in node.traverse(head) do
4204     local i = item.id
4205     if i == node.id'glyph' then
4206       last_char = item
4207     elseif i == 7 and item.subtype == 3 and last_char
4208       and last_char.char > 0x0C99 then
4209       quad = font.getfont(last_char.font).size
4210       for lg, rg in pairs(sea_ranges) do
4211         if last_char.char > rg[1] and last_char.char < rg[2] then
4212           lg = lg:sub(1, 4)  ^^ Remove trailing number of, eg, Cyril1
4213           local intraspace = Babel.intraspaces[lg]
4214           local intrapenalty = Babel.intrapenalties[lg]
4215           local n
4216           if intrapenalty ~= 0 then
4217             n = node.new(14, 0)    ^^ penalty
4218             n.penalty = intrapenalty
4219             node.insert_before(head, item, n)
4220           end
4221           n = node.new(12, 13)    ^^ (glue, spaceskip)
4222           node.setglue(n, intraspace.b * quad,
4223             intraspace.p * quad,
4224             intraspace.m * quad)
4225           node.insert_before(head, item, n)
4226           node.remove(head, item)
4227         end
4228       end
4229     end
4230   end
4231 end
4232 }^^
4233 \bbl@luahyphenate}
4234 \catcode`\%=14
4235 \gdef\bbl@cjkkintraspace{%
4236   \let\bbl@cjkkintraspace\relax
4237   \directlua{
4238     Babel = Babel or {}
4239     require'babel-data-cjk.lua'
4240     Babel.cjk_enabled = true
4241     function Babel.cjk_linebreak(head)
4242       local GLYPH = node.id'glyph'
4243       local last_char = nil
4244       local quad = 655360      % 10 pt = 655360 = 10 * 65536
4245       local last_class = nil
4246       local last_lang = nil
4247
4248       for item in node.traverse(head) do
4249         if item.id == GLYPH then
4250
4251           local lang = item.lang
4252
4253           local LOCALE = node.get_attribute(item,
4254             luatexbase.registernumber'bbl@attr@locale')
4255           local props = Babel.locale_props[LOCALE]
4256
4257           local class = Babel.cjk_class[item.char].c
4258
4259           if class == 'cp' then class = 'cl' end % )] as CL
4260           if class == 'id' then class = 'I' end

```

```

4261
4262     local br = 0
4263     if class and last_class and Babel.cjk_breaks[last_class][class] then
4264         br = Babel.cjk_breaks[last_class][class]
4265     end
4266
4267     if br == 1 and props.linebreak == 'c' and
4268         lang ~= \the\l@nohyphenation\space and
4269         last_lang ~= \the\l@nohyphenation then
4270         local intrapenalty = props.intrapenalty
4271         if intrapenalty ~= 0 then
4272             local n = node.new(14, 0)    % penalty
4273             n.penalty = intrapenalty
4274             node.insert_before(head, item, n)
4275         end
4276         local intraspace = props.intraspace
4277         local n = node.new(12, 13)    % (glue, spaceskip)
4278         node.setglue(n, intraspace.b * quad,
4279             intraspace.p * quad,
4280             intraspace.m * quad)
4281         node.insert_before(head, item, n)
4282     end
4283
4284     quad = font.getfont(item.font).size
4285     last_class = class
4286     last_lang = lang
4287     else % if penalty, glue or anything else
4288         last_class = nil
4289     end
4290 end
4291 lang.hyphenate(head)
4292 end
4293 }%
4294 \bbl@luahyphenate}
4295 \gdef\bbl@luahyphenate{%
4296 \let\bbl@luahyphenate\relax
4297 \directlua{
4298     luatexbase.add_to_callback('hyphenate',
4299     function (head, tail)
4300         if Babel.linebreaking.before then
4301             for k, func in ipairs(Babel.linebreaking.before) do
4302                 func(head)
4303             end
4304         end
4305         if Babel.cjk_enabled then
4306             Babel.cjk_linebreak(head)
4307         end
4308         lang.hyphenate(head)
4309         if Babel.linebreaking.after then
4310             for k, func in ipairs(Babel.linebreaking.after) do
4311                 func(head)
4312             end
4313         end
4314         if Babel.sea_enabled then
4315             Babel.sea_disc_to_space(head)
4316         end
4317     end,
4318     'Babel.hyphenate')
4319 }

```



```

4320 }
4321 \endgroup
4322 \def\bbl@provide@intraspace{%
4323   \bbl@ifunset{\bbl@intsp@language}{}%
4324   {\expandafter\ifx\csname\bbl@intsp@language\endcsname\@empty\else
4325     \bbl@xin@{\bbl@cs{sbc@language}}{Hant,Hans,Jpan,Kore,Kana}%
4326     \ifin@           % cjk
4327     \bbl@cjk@intraspace
4328     \directlua{
4329       Babel = Babel or {}
4330       Babel.locale_props = Babel.locale_props or {}
4331       Babel.locale_props[\the\localeid].linebreak = 'c'
4332     }%
4333     \bbl@exp{\bbl@intraspace\bbl@cs{intsp@language}\@}%
4334     \ifx\bbl@KVP@intrapenalty\@nil
4335       \bbl@intrapenalty0\@
4336     \fi
4337   \else           % sea
4338     \bbl@sea@intraspace
4339     \bbl@exp{\bbl@intraspace\bbl@cs{intsp@language}\@}%
4340     \directlua{
4341       Babel = Babel or {}
4342       Babel.sea_ranges = Babel.sea_ranges or {}
4343       Babel.set_chranges('\bbl@cs{sbc@language}',
4344                           '\bbl@cs{chrng@language}')
4345     }%
4346     \ifx\bbl@KVP@intrapenalty\@nil
4347       \bbl@intrapenalty0\@
4348     \fi
4349   \fi
4350 \fi
4351 \ifx\bbl@KVP@intrapenalty\@nil\else
4352   \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@
4353 \fi}}

```

15.5 CJK line breaking

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

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

Work in progress.

Common stuff.

```

4354 \AddBabelHook{luatex}{loadkernel}{%
4355   <<Restore Unicode catcodes before loading patterns>>}}
4356 \ifx\DisableBabelHook\undefined\endinput\fi
4357 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
4358 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ccheckstdfonts}
4359 \DisableBabelHook{babel-fontspec}
4360 <<Font selection>>

```

15.6 Automatic fonts and ids switching

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

```
4361 \directlua{
4362 Babel.script_blocks = {
4363   ['Arab'] = {{0x0600, 0x06FF}, {0x08A0, 0x08FF}, {0x0750, 0x077F},
4364             {0xFE70, 0xFEFF}, {0xFB50, 0xFDFF}, {0x1EE00, 0x1EEFF}},
4365   ['Armn'] = {{0x0530, 0x058F}},
4366   ['Beng'] = {{0x0980, 0x09FF}},
4367   ['Cher'] = {{0x13A0, 0x13FF}, {0xAB70, 0xABBF}},
4368   ['Cyr1'] = {{0x0400, 0x04FF}, {0x0500, 0x052F}, {0x1C80, 0x1C8F},
4369             {0x2DE0, 0x2DFF}, {0xA640, 0xA69F}},
4370   ['Deva'] = {{0x0900, 0x097F}, {0xA8E0, 0xA8FF}},
4371   ['Ethi'] = {{0x1200, 0x137F}, {0x1380, 0x139F}, {0x2D80, 0x2DDF},
4372             {0xAB00, 0xAB2F}},
4373   ['Geor'] = {{0x10A0, 0x10FF}, {0x2D00, 0x2D2F}},
4374   ['Grek'] = {{0x0370, 0x03FF}, {0x1F00, 0x1FFF}},
4375   ['Hans'] = {{0x2E80, 0x2EFF}, {0x3000, 0x303F}, {0x31C0, 0x31EF},
4376             {0x3300, 0x33FF}, {0x3400, 0x4DBF}, {0x4E00, 0x9FFF},
4377             {0xF900, 0FAFF}, {0xFE30, 0xFE4F}, {0xFF00, 0xFFEF},
4378             {0x20000, 0x2A6DF}, {0x2A700, 0x2B73F},
4379             {0x2B740, 0x2B81F}, {0x2B820, 0x2CEAF},
4380             {0x2CEB0, 0x2EBEF}, {0x2F800, 0x2FA1F}},
4381   ['Hebr'] = {{0x0590, 0x05FF}},
4382   ['Japa'] = {{0x3000, 0x303F}, {0x3040, 0x309F}, {0x30A0, 0x30FF},
4383             {0x4E00, 0x9FAF}, {0xFF00, 0xFFEF}},
4384   ['Khmr'] = {{0x1780, 0x17FF}, {0x19E0, 0x19FF}},
4385   ['Knda'] = {{0x0C80, 0x0CFF}},
4386   ['Kore'] = {{0x1100, 0x11FF}, {0x3000, 0x303F}, {0x3130, 0x318F},
4387             {0x4E00, 0x9FAF}, {0xA960, 0xA97F}, {0xAC00, 0xD7AF},
4388             {0xD7B0, 0xD7FF}, {0xFF00, 0xFFEF}},
4389   ['Laoo'] = {{0x0E80, 0x0EFF}},
4390   ['Latn'] = {{0x0000, 0x007F}, {0x0080, 0x00FF}, {0x0100, 0x017F},
4391             {0x0180, 0x024F}, {0x1E00, 0x1EFF}, {0x2C60, 0x2C7F},
4392             {0xA720, 0xA7FF}, {0xAB30, 0xAB6F}},
4393   ['Mahj'] = {{0x11150, 0x1117F}},
4394   ['Mlym'] = {{0x0D00, 0x0D7F}},
4395   ['Mymr'] = {{0x1000, 0x109F}, {0xAA60, 0xAA7F}, {0xA9E0, 0xA9FF}},
4396   ['Orya'] = {{0x0B00, 0x0B7F}},
4397   ['Sinh'] = {{0x0D80, 0x0DFF}, {0x111E0, 0x111FF}},
4398   ['Taml'] = {{0x0B80, 0x0BFF}},
4399   ['Telu'] = {{0x0C00, 0x0C7F}},
4400   ['Tfng'] = {{0x2D30, 0x2D7F}},
4401   ['Thai'] = {{0x0E00, 0x0E7F}},
4402   ['Tibt'] = {{0x0F00, 0x0FFF}},
4403   ['Vaii'] = {{0xA500, 0xA63F}},
4404   ['Yiii'] = {{0xA490, 0xA4CF}, {0xA000, 0xA48F}}
4405 }
4406
4407 Babel.script_blocks.Hant = Babel.script_blocks.Hans
4408
```

```

4409 function Babel.locale_map(head)
4410   if not Babel.locale_mapped then return head end
4411
4412   local LOCALE = luatexbase.registernumber'bbl@attr@locale'
4413   local GLYPH = node.id('glyph')
4414   local inmath = false
4415   for item in node.traverse(head) do
4416     local toloc
4417     if not inmath and item.id == GLYPH then
4418       % Optimization: build a table with the chars found
4419       if Babel.chr_to_loc[item.char] then
4420         toloc = Babel.chr_to_loc[item.char]
4421       else
4422         for lc, maps in pairs(Babel.loc_to_scr) do
4423           for _, rg in pairs(maps) do
4424             if item.char >= rg[1] and item.char <= rg[2] then
4425               Babel.chr_to_loc[item.char] = lc
4426               toloc = lc
4427               break
4428             end
4429           end
4430         end
4431       end
4432       % Now, take action
4433       if toloc then
4434         if Babel.locale_props[toloc].lg then
4435           item.lang = Babel.locale_props[toloc].lg
4436           node.set_attribute(item, LOCALE, toloc)
4437         end
4438         if Babel.locale_props[toloc]['/'..item.font] then
4439           item.font = Babel.locale_props[toloc]['/'..item.font]
4440         end
4441       end
4442     elseif not inmath and item.id == 7 then
4443       item.replace = item.replace and Babel.locale_map(item.replace)
4444       item.pre = item.pre and Babel.locale_map(item.pre)
4445       item.post = item.post and Babel.locale_map(item.post)
4446     elseif item.id == node.id'math' then
4447       inmath = (item.subtype == 0)
4448     end
4449   end
4450   return head
4451 end
4452 }

```

The code for `\babelcharproperty` is straightforward. Just note the modified lua table can be different.

```

4453 \newcommand\babelcharproperty[1]{%
4454   \count@=#1\relax
4455   \ifvmode
4456     \expandafter\bbl@chprop
4457   \else
4458     \bbl@error{\string\babelcharproperty\space can be used only in\%
4459               vertical mode (preamble or between paragraphs)}%
4460     {See the manual for futher info}%
4461   \fi}
4462 \newcommand\bbl@chprop[3][\the\count@]{%
4463   \@tempcnta=#1\relax
4464   \bbl@ifunset\bbl@chprop@#2}%

```

```

4465     {\bbl@error{No property named '#2'. Allowed values are\\%
4466                 direction (bc), mirror (bmg), and linebreak (lb)}}%
4467     {See the manual for futher info}}%
4468     {}%
4469 \loop
4470   \@nameuse{bbl@chprop@#2}{#3}%
4471 \ifnum\count@<\@tempcnta
4472   \advance\count@\@ne
4473 \repeat}
4474 \def\bbl@chprop@direction#1{%
4475   \directlua{
4476     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
4477     Babel.characters[\the\count@]['d'] = '#1'
4478   }}
4479 \let\bbl@chprop@bc\bbl@chprop@direction
4480 \def\bbl@chprop@mirror#1{%
4481   \directlua{
4482     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
4483     Babel.characters[\the\count@]['m'] = '\number#1'
4484   }}
4485 \let\bbl@chprop@bmg\bbl@chprop@mirror
4486 \def\bbl@chprop@linebreak#1{%
4487   \directlua{
4488     Babel.Babel.cjk_characters[\the\count@] = Babel.Babel.cjk_characters[\the\count@] or {}
4489     Babel.Babel.cjk_characters[\the\count@]['c'] = '#1'
4490   }}
4491 \let\bbl@chprop@lb\bbl@chprop@linebreak

```

Post-handling hyphenation patterns for non-standard rules, like ff to ff-f. There are still some issues with speed (not very slow, but still slow).

After declaring the table containing the patterns with their replacements, we define some auxiliary functions: `str_to_nodes` converts the string returned by a function to a node list, taking the node at base as a model (font, language, etc.); `fetch_word` fetches a series of glyphs and discretionaries, which pattern is matched against (if there is a match, it is called again before trying other patterns, and this is very likely the main bottleneck). `post_hyphenate_replace` is the callback applied after `tex.hyphenate`. This means the automatic hyphenation points are known. As empty captures return a byte position (as explained in the luatex manual), we must convert it to a utf8 position. With `first`, the last byte can be the leading byte in a utf8 sequence, so we just remove it and add 1 to the resulting length. With `last` we must take into account the capture position points to the next character. Here `word_head` points to the starting node of the text to be matched.

```

4492 \begingroup
4493 \catcode`\#=12
4494 \catcode`\%=12
4495 \catcode`\&=14
4496 \directlua{
4497   Babel.linebreaking.replacements = {}
4498
4499   function Babel.str_to_nodes(fn, matches, base)
4500     local n, head, last
4501     if fn == nil then return nil end
4502     for s in string.utfvalues(fn(matches)) do
4503       if base.id == 7 then
4504         base = base.replace
4505       end
4506       n = node.copy(base)
4507       n.char = s
4508       if not head then

```

```

4509         head = n
4510     else
4511         last.next = n
4512     end
4513     last = n
4514 end
4515 return head
4516 end
4517
4518 function Babel.fetch_word(head, funct)
4519     local word_string = ''
4520     local word_nodes = {}
4521     local lang
4522     local item = head
4523
4524     while item do
4525
4526         if item.id == 29
4527             and not(item.char == 124) && ie, not |
4528             and not(item.char == 61) && ie, not =
4529             and (item.lang == lang or lang == nil) then
4530             lang = lang or item.lang
4531             word_string = word_string .. unicode.utf8.char(item.char)
4532             word_nodes[#word_nodes+1] = item
4533
4534             elseif item.id == 7 and item.subtype == 2 then
4535             word_string = word_string .. '='
4536             word_nodes[#word_nodes+1] = item
4537
4538             elseif item.id == 7 and item.subtype == 3 then
4539             word_string = word_string .. '|'
4540             word_nodes[#word_nodes+1] = item
4541
4542             elseif word_string == '' then
4543             && pass
4544
4545             else
4546             return word_string, word_nodes, item, lang
4547             end
4548
4549         item = item.next
4550     end
4551 end
4552
4553 function Babel.post_hyphenate_replace(head)
4554     local u = unicode.utf8
4555     local lbkr = Babel.linebreaking.replacements
4556     local word_head = head
4557
4558     while true do
4559         local w, wn, nw, lang = Babel.fetch_word(word_head)
4560         if not lang then return head end
4561
4562         if not lbkr[lang] then
4563             break
4564         end
4565
4566         for k=1, #lbkr[lang] do
4567             local p = lbkr[lang][k].pattern

```

```

4568     local r = lbkr[lang][k].replace
4569
4570     while true do
4571         local matches = { u.match(w, p) }
4572         if #matches < 2 then break end
4573
4574         local first = table.remove(matches, 1)
4575         local last = table.remove(matches, #matches)
4576
4577         %% Fix offsets, from bytes to unicode.
4578         first = u.len(w:sub(1, first-1)) + 1
4579         last = u.len(w:sub(1, last-1))
4580
4581         local new %% used when inserting and removing nodes
4582         local changed = 0
4583
4584         %% This loop traverses the replace list and takes the
4585         %% corresponding actions
4586         for q = first, last do
4587             local crep = r[q-first+1]
4588             local char_node = wn[q]
4589             local char_base = char_node
4590
4591             if crep and crep.data then
4592                 char_base = wn[crep.data+first-1]
4593             end
4594
4595             if crep == {} then
4596                 break
4597             elseif crep == nil then
4598                 changed = changed + 1
4599                 node.remove(head, char_node)
4600             elseif crep and (crep.pre or crep.no or crep.post) then
4601                 changed = changed + 1
4602                 d = node.new(7, 0) %% (disc, discretionary)
4603                 d.pre = Babel.str_to_nodes(crep.pre, matches, char_base)
4604                 d.post = Babel.str_to_nodes(crep.post, matches, char_base)
4605                 d.replace = Babel.str_to_nodes(crep.no, matches, char_base)
4606                 d.attr = char_base.attr
4607                 if crep.pre == nil then %% TeXbook p96
4608                     d.penalty = crep.penalty or tex.hyphenpenalty
4609                 else
4610                     d.penalty = crep.penalty or tex.exhyphenpenalty
4611                 end
4612                 head, new = node.insert_before(head, char_node, d)
4613                 node.remove(head, char_node)
4614                 if q == 1 then
4615                     word_head = new
4616                 end
4617             elseif crep and crep.string then
4618                 changed = changed + 1
4619                 local str = crep.string(matches)
4620                 if str == '' then
4621                     if q == 1 then
4622                         word_head = char_node.next
4623                     end
4624                     head, new = node.remove(head, char_node)
4625                 elseif char_node.id == 29 and u.len(str) == 1 then
4626                     char_node.char = string.utfvalue(str)

```

```

4627         else
4628             local n
4629             for s in string.utfvalues(str) do
4630                 if char_node.id == 7 then
4631                     log('Automatic hyphens cannot be replaced, just removed.')
4632                 else
4633                     n = node.copy(char_base)
4634                 end
4635                 n.char = s
4636                 if q == 1 then
4637                     head, new = node.insert_before(head, char_node, n)
4638                     word_head = new
4639                 else
4640                     node.insert_before(head, char_node, n)
4641                 end
4642             end
4643
4644             node.remove(head, char_node)
4645             end %% string length
4646         end %% if char and char.string
4647     end %% for char in match
4648     if changed > 20 then
4649         texio.write('Too many changes. Ignoring the rest.')
4650     elseif changed > 0 then
4651         w, wn, nw = Babel.fetch_word(word_head)
4652     end
4653
4654     end %% for match
4655 end %% for patterns
4656 word_head = nw
4657 end %% for words
4658 return head
4659 end
4660
4661 %% Used below
4662 function Babel.capture_func(key, cap)
4663     local ret = "[" .. cap:gsub('{{([0-9])}}', "]]..m[%1]..[" .. "]"
4664     ret = ret:gsub("%[%[%]%]%.%", '')
4665     ret = ret:gsub("%.%[%[%]%]%", '')
4666     return key .. [[=function(m) return ]] .. ret .. [[ end]]
4667 end
4668 }

```

Now the T_EX high level interface, which requires the function defined above for converting strings to functions returning a string. These functions handle the $\{n\}$ syntax. For example, $\text{pre}=\{1\}\{1\}$ - becomes `function(m) return m[1]..m[1]..'-' end`, where m are the matches returned after applying the pattern. The way it is carried out is somewhat tricky, but the effect is not dissimilar to `lua load – save the code as string in a TeX macro, and expand this macro at the appropriate place`. As `\directlua` does not take into account the current catcode of `@`, we just avoid this character in macro names (which explains the internal group, too).

```

4669 \catcode`\#=6
4670 \gdef\babelposthyphenation#1#2#3{&%
4671 \begingroup
4672 \def\babeltempa{\bbl@add@list\babeltempb}&%
4673 \let\babeltempb\@empty
4674 \bbl@foreach{#3}{&%
4675 \bbl@ifsamestring{##1}{remove}&%

```

```

4676      {\bbl@add@list\babeltempb{nil}}&%
4677      {\directlua{
4678          local rep = [[#1]]
4679          rep = rep:gsub( '(no)%s*=%s*([^\s,]*)', Babel.capture_func)
4680          rep = rep:gsub( '(pre)%s*=%s*([^\s,]*)', Babel.capture_func)
4681          rep = rep:gsub( '(post)%s*=%s*([^\s,]*)', Babel.capture_func)
4682          rep = rep:gsub( '(string)%s*=%s*([^\s,]*)', Babel.capture_func)
4683          tex.print([[string\babeltempa{}}] .. rep .. [[]]})
4684      }}&%
4685      \directlua{
4686          local lbkr = Babel.linebreaking.replacements
4687          local u = unicode.utf8
4688          &% Convert pattern:
4689          local patt = string.gsub([[#2]], '%s', '')
4690          if not u.find(patt, '()', nil, true) then
4691              patt = '()' .. patt .. '()'
4692          end
4693          patt = u.gsub(patt, '{(.)}',
4694                      function (n)
4695                          return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
4696                      end)
4697          lbkr[\the\csname l@#1\endcsname] = lbkr[\the\csname l@#1\endcsname] or {}
4698          table.insert(lbkr[\the\csname l@#1\endcsname],
4699                      { pattern = patt, replace = { \babeltempb } })
4700      }&%
4701      \endgroup}
4702 \endgroup

```

15.7 Layout

Work in progress.

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

`\@hangfrom` is useful in many contexts and it is redefined always with the `layout` option. There are, however, a number of issues when the text direction is not the same as the box direction (as set by `\bodydir`), and when `\parbox` and `\hangindent` are involved.

Fortunately, latest releases of luatex simplify a lot the solution with `\shapemode`.

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

```

4703 \bbl@trace{Redefinitions for bidi layout}
4704 \ifx\@eqnnum\@undefined\else
4705   \ifx\bbl@attr@dir\@undefined\else
4706     \edef\@eqnnum{%
4707       \unexpanded{\ifcase\bbl@attr@dir\else\bbl@textdir\@ne\fi}%
4708       \unexpanded\expandafter{\@eqnnum}}
4709   \fi
4710 \fi
4711 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
4712 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
4713   \def\bbl@nextfake#1{% non-local changes, use always inside a group!
4714     \bbl@exp{%
4715       \mathdir\the\bodydir
4716       #1%           Once entered in math, set boxes to restore values

```



```

4717 \<ifmode>%
4718 \everyvbox{%
4719 \the\everyvbox
4720 \bodydir\the\bodydir
4721 \mathdir\the\mathdir
4722 \everyhbox{\the\everyhbox}%
4723 \everyvbox{\the\everyvbox}}%
4724 \everyhbox{%
4725 \the\everyhbox
4726 \bodydir\the\bodydir
4727 \mathdir\the\mathdir
4728 \everyhbox{\the\everyhbox}%
4729 \everyvbox{\the\everyvbox}}%
4730 \<fi>}}%
4731 \def\hangfrom#1{%
4732 \setbox\@tempboxa\hbox{{#1}}%
4733 \hangindent\wd\@tempboxa
4734 \ifnum\bb1@getluadir{page}=\bb1@getluadir{par}\else
4735 \shapemode\@ne
4736 \fi
4737 \noindent\box\@tempboxa}
4738 \fi
4739 \IfBabelLayout{tabular}
4740 {\let\bb1@OL@tabular\@tabular
4741 \bb1@replace\@tabular{$}{\bb1@nextfake$}%
4742 \let\bb1@NL@tabular\@tabular
4743 \AtBeginDocument{%
4744 \ifx\bb1@NL@tabular\@tabular\else
4745 \bb1@replace\@tabular{$}{\bb1@nextfake$}%
4746 \let\bb1@NL@tabular\@tabular
4747 \fi}}
4748 {}
4749 \IfBabelLayout{lists}
4750 {\let\bb1@OL@list\list
4751 \bb1@sreplace\list{\parshape}{\bb1@listparshape}%
4752 \let\bb1@NL@list\list
4753 \def\bb1@listparshape#1#2#3{%
4754 \parshape #1 #2 #3 %
4755 \ifnum\bb1@getluadir{page}=\bb1@getluadir{par}\else
4756 \shapemode\tw@
4757 \fi}}
4758 {}
4759 \IfBabelLayout{graphics}
4760 {\let\bb1@pictresetdir\relax
4761 \def\bb1@pictsetdir{%
4762 \ifcase\bb1@thetextdir
4763 \let\bb1@pictresetdir\relax
4764 \else
4765 \textdir TLT\relax
4766 \def\bb1@pictresetdir{\textdir TRT\relax}%
4767 \fi}%
4768 \let\bb1@OL@picture\@picture
4769 \let\bb1@OL@put\put
4770 \bb1@sreplace\@picture{\hskip-}{\bb1@pictsetdir\hskip-}%
4771 \def\put(#1,#2)#3{% Not easy to patch. Better redefine.
4772 \@killglue
4773 \raise#2\unitlength
4774 \hb@xt@\z@\{\kern#1\unitlength{\bb1@pictresetdir#3}\hss}}%
4775 \AtBeginDocument

```

```

4776     {\ifx\tikz@atbegin@node\@undefined\else
4777       \let\bbl@OL@pgfpicture\pgfpicture
4778       \bbl@sreplace\pgfpicture{\pgfpicturetrue}{\bbl@pictsetdir\pgfpicturetrue}%
4779       \bbl@add\pgfsys@beginpicture{\bbl@pictsetdir}%
4780       \bbl@add\tikz@atbegin@node{\bbl@pictresetdir}%
4781     \fi}}
4782 {}

```

Implicitly reverses sectioning labels in `bidi=basic-r`, because the full stop is not in contact with L numbers any more. I think there must be a better way. Assumes `bidi=basic`, but there are some additional readjustments for `bidi=default`.

```

4783 \IfBabelLayout{counters}%
4784 {\let\bbl@OL@@textsuperscript\textsuperscript
4785  \bbl@sreplace\textsuperscript{\m@th}{\m@th\mathdir\pagedir}%
4786  \let\bbl@latinarabic=\@arabic
4787  \let\bbl@OL@@arabic\@arabic
4788  \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
4789  \@ifpackagewith{babel}{bidi=default}%
4790  {\let\bbl@asciroman=\@roman
4791   \let\bbl@OL@@roman\@roman
4792   \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
4793   \let\bbl@asciiRoman=\@Roman
4794   \let\bbl@OL@@roman\@Roman
4795   \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}%
4796   \let\bbl@OL@labelenumii\labelenumii
4797   \def\labelenumii{}\theenumii{}}%
4798   \let\bbl@OL@p@enumiii\p@enumiii
4799   \def\p@enumiii{\p@enumii}\theenumii{}}{}{}
4800 \langle\langle Footnote changes\rangle\rangle
4801 \IfBabelLayout{footnotes}%
4802 {\let\bbl@OL@footnote\footnote
4803  \BabelFootnote\footnote\language\language{}}{}%
4804  \BabelFootnote\localfootnote\language\language{}}{}%
4805  \BabelFootnote\mainfootnote{}}{}{}
4806 {}

```

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

```

4807 \IfBabelLayout{extras}%
4808 {\let\bbl@OL@underline\underline
4809  \bbl@sreplace\underline{\$@@underline}{\bbl@nextfake\$@@underline}%
4810  \let\bbl@OL@LaTeX2e\LaTeX2e
4811  \DeclareRobustCommand{\LaTeXe}{\mbox{\m@th
4812   \if b\expandafter\@car\@series\@nil\boldmath\fi
4813   \babelsublr{%
4814     \LaTeX\kern.15em2\bbl@nextfake$_{\textstyle\varepsilon}$}}}
4815 {}
4816 \langle/luatex\rangle

```

15.8 Auto bidi with basic and basic-r

The file `babel-data-bidi.lua` currently only contains data. It is a large and boring file and it is not shown here (see the generated file), but here is a sample:

```

[0x25]={d='et'},
[0x26]={d='on'},
[0x27]={d='on'},

```

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

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):

Arrrrgh!! The UAX#9 algorithm is too deeply entrenched in the assumption of batch-style processing [...]. May the fleas of a thousand camels infest the armpits of those who design supposedly general-purpose algorithms by looking at their own implementations, and fail to consider other possible implementations!

Well, it took me some time to guess what the batch rules in UAX#9 actually mean (in other word, *what* they do and *why*, and not only *how*), but I think (or I hope) I've managed to understand them.

In some sense, there are two bidi modes, one for numbers, and the other for text. Furthermore, setting just the direction in R text is not enough, because there are actually *two* R modes (set explicitly in Unicode with RLM and ALM). In babel the dir is set by a higher protocol based on the language/script, which in turn sets the correct dir (<l>, <r> or <al>).

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

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

```
4817 (*basic-r)
4818 Babel = Babel or {}
4819
4820 Babel.bidi_enabled = true
4821
4822 require('babel-data-bidi.lua')
4823
4824 local characters = Babel.characters
4825 local ranges = Babel.ranges
4826
4827 local DIR = node.id("dir")
4828
4829 local function dir_mark(head, from, to, outer)
4830   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
4831   local d = node.new(DIR)
4832   d.dir = '+' .. dir
4833   node.insert_before(head, from, d)
4834   d = node.new(DIR)
4835   d.dir = '-' .. dir
4836   node.insert_after(head, to, d)
4837 end
```

```

4838
4839 function Babel.bidi(head, ispar)
4840   local first_n, last_n          -- first and last char with nums
4841   local last_es                  -- an auxiliary 'last' used with nums
4842   local first_d, last_d          -- first and last char in L/R block
4843   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):

```

4844   local strong = ('TRT' == tex.pardir) and 'r' or 'l'
4845   local strong_lr = (strong == 'l') and 'l' or 'r'
4846   local outer = strong
4847
4848   local new_dir = false
4849   local first_dir = false
4850   local inmath = false
4851
4852   local last_lr
4853
4854   local type_n = ''
4855
4856   for item in node.traverse(head) do
4857
4858     -- three cases: glyph, dir, otherwise
4859     if item.id == node.id'glyph'
4860       or (item.id == 7 and item.subtype == 2) then
4861
4862       local itemchar
4863       if item.id == 7 and item.subtype == 2 then
4864         itemchar = item.replace.char
4865       else
4866         itemchar = item.char
4867       end
4868       local chardata = characters[itemchar]
4869       dir = chardata and chardata.d or nil
4870       if not dir then
4871         for nn, et in ipairs(ranges) do
4872           if itemchar < et[1] then
4873             break
4874           elseif itemchar <= et[2] then
4875             dir = et[3]
4876             break
4877           end
4878         end
4879       end
4880       dir = dir or 'l'
4881       if inmath then dir = ('TRT' == tex.mathdir) and 'r' or 'l' end

```

Next is based on the assumption babel sets the language AND switches the script with its dir. We treat a language block as a separate Unicode sequence. The following piece of code is executed at the first glyph after a 'dir' node. We don't know the current language until then. This is not exactly true, as the math mode may insert explicit dirs in the node list, so, for the moment there is a hack by brute force (just above).

```

4882   if new_dir then
4883     attr_dir = 0
4884     for at in node.traverse(item.attr) do
4885       if at.number == luatexbase.registernumber'bbl@attr@dir' then

```

```

4886         attr_dir = at.value % 3
4887     end
4888 end
4889 if attr_dir == 1 then
4890     strong = 'r'
4891 elseif attr_dir == 2 then
4892     strong = 'al'
4893 else
4894     strong = 'l'
4895 end
4896 strong_lr = (strong == 'l') and 'l' or 'r'
4897 outer = strong_lr
4898 new_dir = false
4899 end
4900
4901 if dir == 'nsm' then dir = strong end -- W1

```

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

```

4902     dir_real = dir -- We need dir_real to set strong below
4903     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:

```

4904     if strong == 'al' then
4905         if dir == 'en' then dir = 'an' end -- W2
4906         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
4907         strong_lr = 'r' -- W3
4908     end

```

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

```

4909     elseif item.id == node.id'dir' and not inmath then
4910         new_dir = true
4911         dir = nil
4912     elseif item.id == node.id'math' then
4913         inmath = (item.subtype == 0)
4914     else
4915         dir = nil -- Not a char
4916     end

```

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

```

4917     if dir == 'en' or dir == 'an' or dir == 'et' then
4918         if dir ~= 'et' then
4919             type_n = dir
4920         end
4921         first_n = first_n or item
4922         last_n = last_es or item
4923         last_es = nil
4924     elseif dir == 'es' and last_n then -- W3+W6
4925         last_es = item
4926     elseif dir == 'cs' then -- it's right - do nothing
4927     elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
4928         if strong_lr == 'r' and type_n ~= '' then

```

```

4929     dir_mark(head, first_n, last_n, 'r')
4930   elseif strong_lr == 'l' and first_d and type_n == 'an' then
4931     dir_mark(head, first_n, last_n, 'r')
4932     dir_mark(head, first_d, last_d, outer)
4933     first_d, last_d = nil, nil
4934   elseif strong_lr == 'l' and type_n ~= '' then
4935     last_d = last_n
4936   end
4937   type_n = ''
4938   first_n, last_n = nil, nil
4939 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, whatsits, etc., are ignored:

```

4940   if dir == 'l' or dir == 'r' then
4941     if dir ~= outer then
4942       first_d = first_d or item
4943       last_d = item
4944     elseif first_d and dir ~= strong_lr then
4945       dir_mark(head, first_d, last_d, outer)
4946       first_d, last_d = nil, nil
4947     end
4948   end

```

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

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

```

4949   if dir and not last_lr and dir ~= 'l' and outer == 'r' then
4950     item.char = characters[item.char] and
4951       characters[item.char].m or item.char
4952   elseif (dir or new_dir) and last_lr ~= item then
4953     local mir = outer .. strong_lr .. (dir or outer)
4954     if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
4955       for ch in node.traverse(node.next(last_lr)) do
4956         if ch == item then break end
4957         if ch.id == node.id'glyph' and characters[ch.char] then
4958           ch.char = characters[ch.char].m or ch.char
4959         end
4960       end
4961     end
4962   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).

```

4963   if dir == 'l' or dir == 'r' then
4964     last_lr = item
4965     strong = dir_real          -- Don't search back - best save now
4966     strong_lr = (strong == 'l') and 'l' or 'r'
4967   elseif new_dir then
4968     last_lr = nil
4969   end
4970 end

```

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

```

4971 if last_lr and outer == 'r' then
4972     for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
4973         if characters[ch.char] then
4974             ch.char = characters[ch.char].m or ch.char
4975         end
4976     end
4977 end
4978 if first_n then
4979     dir_mark(head, first_n, last_n, outer)
4980 end
4981 if first_d then
4982     dir_mark(head, first_d, last_d, outer)
4983 end

```

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

```

4984 return node.prev(head) or head
4985 end
4986 </basic-r>

```

And here the Lua code for bidi=basic:

```

4987 <(*basic)
4988 Babel = Babel or {}
4989
4990 -- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
4991
4992 Babel.fontmap = Babel.fontmap or {}
4993 Babel.fontmap[0] = {}      -- l
4994 Babel.fontmap[1] = {}      -- r
4995 Babel.fontmap[2] = {}      -- al/an
4996
4997 Babel.bidi_enabled = true
4998 Babel.mirroring_enabled = true
4999
5000 -- Temporary:
5001
5002 if harf then
5003     Babel.mirroring_enabled = false
5004 end
5005
5006 require('babel-data-bidi.lua')
5007
5008 local characters = Babel.characters
5009 local ranges = Babel.ranges
5010
5011 local DIR = node.id('dir')
5012 local GLYPH = node.id('glyph')
5013
5014 local function insert_implicit(head, state, outer)
5015     local new_state = state
5016     if state.sim and state.eim and state.sim ~= state.eim then
5017         dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
5018         local d = node.new(DIR)
5019         d.dir = '+' .. dir
5020         node.insert_before(head, state.sim, d)
5021         local d = node.new(DIR)
5022         d.dir = '-' .. dir
5023         node.insert_after(head, state.eim, d)
5024     end

```

```

5025 new_state.sim, new_state.eim = nil, nil
5026 return head, new_state
5027 end
5028
5029 local function insert_numeric(head, state)
5030   local new
5031   local new_state = state
5032   if state.san and state.ean and state.san ~= state.ean then
5033     local d = node.new(DIR)
5034     d.dir = '+TLT'
5035     _, new = node.insert_before(head, state.san, d)
5036     if state.san == state.sim then state.sim = new end
5037     local d = node.new(DIR)
5038     d.dir = '-TLT'
5039     _, new = node.insert_after(head, state.ean, d)
5040     if state.ean == state.eim then state.eim = new end
5041   end
5042   new_state.san, new_state.ean = nil, nil
5043   return head, new_state
5044 end
5045
5046 -- TODO - \hbox with an explicit dir can lead to wrong results
5047 -- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>. A small attempt
5048 -- was made to improve the situation, but the problem is the 3-dir
5049 -- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
5050 -- well.
5051
5052 function Babel.bidi(head, ispar, hdir)
5053   local d -- d is used mainly for computations in a loop
5054   local prev_d = ''
5055   local new_d = false
5056
5057   local nodes = {}
5058   local outer_first = nil
5059   local inmath = false
5060
5061   local glue_d = nil
5062   local glue_i = nil
5063
5064   local has_en = false
5065   local first_et = nil
5066
5067   local ATDIR = luatexbase.registernumber'bbl@attr@dir'
5068
5069   local save_outer
5070   local temp = node.get_attribute(head, ATDIR)
5071   if temp then
5072     temp = temp % 3
5073     save_outer = (temp == 0 and 'l') or
5074                  (temp == 1 and 'r') or
5075                  (temp == 2 and 'al')
5076   elseif ispar then -- Or error? Shouldn't happen
5077     save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
5078   else -- Or error? Shouldn't happen
5079     save_outer = ('TRT' == hdir) and 'r' or 'l'
5080   end
5081   -- when the callback is called, we are just _after_ the box,
5082   -- and the textdir is that of the surrounding text
5083   -- if not ispar and hdir ~= tex.textdir then

```



```

5084 -- save_outer = ('TRT' == hdir) and 'r' or 'l'
5085 -- end
5086 local outer = save_outer
5087 local last = outer
5088 -- 'al' is only taken into account in the first, current loop
5089 if save_outer == 'al' then save_outer = 'r' end
5090
5091 local fontmap = Babel.fontmap
5092
5093 for item in node.traverse(head) do
5094
5095     -- In what follows, #node is the last (previous) node, because the
5096     -- current one is not added until we start processing the neutrals.
5097
5098     -- three cases: glyph, dir, otherwise
5099     if item.id == GLYPH
5100         or (item.id == 7 and item.subtype == 2) then
5101
5102         local d_font = nil
5103         local item_r
5104         if item.id == 7 and item.subtype == 2 then
5105             item_r = item.replace -- automatic discs have just 1 glyph
5106         else
5107             item_r = item
5108         end
5109         local chardata = characters[item_r.char]
5110         d = chardata and chardata.d or nil
5111         if not d or d == 'nsm' then
5112             for nn, et in ipairs(ranges) do
5113                 if item_r.char < et[1] then
5114                     break
5115                 elseif item_r.char <= et[2] then
5116                     if not d then d = et[3]
5117                     elseif d == 'nsm' then d_font = et[3]
5118                     end
5119                     break
5120                 end
5121             end
5122         end
5123         d = d or 'l'
5124
5125         -- A short 'pause' in bidi for mapfont
5126         d_font = d_font or d
5127         d_font = (d_font == 'l' and 0) or
5128             (d_font == 'nsm' and 0) or
5129             (d_font == 'r' and 1) or
5130             (d_font == 'al' and 2) or
5131             (d_font == 'an' and 2) or nil
5132         if d_font and fontmap and fontmap[d_font][item_r.font] then
5133             item_r.font = fontmap[d_font][item_r.font]
5134         end
5135
5136         if new_d then
5137             table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
5138             if inmath then
5139                 attr_d = 0
5140             else
5141                 attr_d = node.get_attribute(item, ATDIR)
5142                 attr_d = attr_d % 3

```

```

5143         end
5144         if attr_d == 1 then
5145             outer_first = 'r'
5146             last = 'r'
5147         elseif attr_d == 2 then
5148             outer_first = 'r'
5149             last = 'al'
5150         else
5151             outer_first = 'l'
5152             last = 'l'
5153         end
5154         outer = last
5155         has_en = false
5156         first_et = nil
5157         new_d = false
5158     end
5159
5160     if glue_d then
5161         if (d == 'l' and 'l' or 'r') ~= glue_d then
5162             table.insert(nodes, {glue_i, 'on', nil})
5163         end
5164         glue_d = nil
5165         glue_i = nil
5166     end
5167
5168     elseif item.id == DIR then
5169         d = nil
5170         new_d = true
5171
5172     elseif item.id == node.id'glue' and item.subtype == 13 then
5173         glue_d = d
5174         glue_i = item
5175         d = nil
5176
5177     elseif item.id == node.id'math' then
5178         inmath = (item.subtype == 0)
5179
5180     else
5181         d = nil
5182     end
5183
5184     -- AL <= EN/ET/ES      -- W2 + W3 + W6
5185     if last == 'al' and d == 'en' then
5186         d = 'an'          -- W3
5187     elseif last == 'al' and (d == 'et' or d == 'es') then
5188         d = 'on'          -- W6
5189     end
5190
5191     -- EN + CS/ES + EN      -- W4
5192     if d == 'en' and #nodes >= 2 then
5193         if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
5194             and nodes[#nodes-1][2] == 'en' then
5195             nodes[#nodes][2] = 'en'
5196         end
5197     end
5198
5199     -- AN + CS + AN         -- W4 too, because uax9 mixes both cases
5200     if d == 'an' and #nodes >= 2 then
5201         if (nodes[#nodes][2] == 'cs')

```

```

5202         and nodes[#nodes-1][2] == 'an' then
5203             nodes[#nodes][2] = 'an'
5204         end
5205     end
5206
5207     -- ET/EN          -- W5 + W7->l / W6->on
5208     if d == 'et' then
5209         first_et = first_et or (#nodes + 1)
5210     elseif d == 'en' then
5211         has_en = true
5212         first_et = first_et or (#nodes + 1)
5213     elseif first_et then      -- d may be nil here !
5214         if has_en then
5215             if last == 'l' then
5216                 temp = 'l'      -- W7
5217             else
5218                 temp = 'en'     -- W5
5219             end
5220         else
5221             temp = 'on'         -- W6
5222         end
5223         for e = first_et, #nodes do
5224             if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
5225         end
5226         first_et = nil
5227         has_en = false
5228     end
5229
5230     if d then
5231         if d == 'al' then
5232             d = 'r'
5233             last = 'al'
5234         elseif d == 'l' or d == 'r' then
5235             last = d
5236         end
5237         prev_d = d
5238         table.insert(nodes, {item, d, outer_first})
5239     end
5240
5241     outer_first = nil
5242
5243 end
5244
5245 -- TODO -- repeated here in case EN/ET is the last node. Find a
5246 -- better way of doing things:
5247 if first_et then      -- dir may be nil here !
5248     if has_en then
5249         if last == 'l' then
5250             temp = 'l'      -- W7
5251         else
5252             temp = 'en'     -- W5
5253         end
5254     else
5255         temp = 'on'         -- W6
5256     end
5257     for e = first_et, #nodes do
5258         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
5259     end
5260 end

```

```

5261
5262 -- dummy node, to close things
5263 table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
5264
5265 ----- NEUTRAL -----
5266
5267 outer = save_outer
5268 last = outer
5269
5270 local first_on = nil
5271
5272 for q = 1, #nodes do
5273     local item
5274
5275     local outer_first = nodes[q][3]
5276     outer = outer_first or outer
5277     last = outer_first or last
5278
5279     local d = nodes[q][2]
5280     if d == 'an' or d == 'en' then d = 'r' end
5281     if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
5282
5283     if d == 'on' then
5284         first_on = first_on or q
5285     elseif first_on then
5286         if last == d then
5287             temp = d
5288         else
5289             temp = outer
5290         end
5291         for r = first_on, q - 1 do
5292             nodes[r][2] = temp
5293             item = nodes[r][1] -- MIRRORING
5294             if Babel.mirroring_enabled and item.id == GLYPH
5295                 and temp == 'r' and characters[item.char] then
5296                 item.char = characters[item.char].m or item.char
5297             end
5298         end
5299         first_on = nil
5300     end
5301
5302     if d == 'r' or d == 'l' then last = d end
5303 end
5304
5305 ----- IMPLICIT, REORDER -----
5306
5307 outer = save_outer
5308 last = outer
5309
5310 local state = {}
5311 state.has_r = false
5312
5313 for q = 1, #nodes do
5314
5315     local item = nodes[q][1]
5316
5317     outer = nodes[q][3] or outer
5318
5319     local d = nodes[q][2]

```

```

5320
5321   if d == 'nsm' then d = last end           -- W1
5322   if d == 'en' then d = 'an' end
5323   local isdir = (d == 'r' or d == 'l')
5324
5325   if outer == 'l' and d == 'an' then
5326     state.san = state.san or item
5327     state.ean = item
5328   elseif state.san then
5329     head, state = insert_numeric(head, state)
5330   end
5331
5332   if outer == 'l' then
5333     if d == 'an' or d == 'r' then          -- im -> implicit
5334       if d == 'r' then state.has_r = true end
5335       state.sim = state.sim or item
5336       state.eim = item
5337     elseif d == 'l' and state.sim and state.has_r then
5338       head, state = insert_implicit(head, state, outer)
5339     elseif d == 'l' then
5340       state.sim, state.eim, state.has_r = nil, nil, false
5341     end
5342   else
5343     if d == 'an' or d == 'l' then
5344       if nodes[q][3] then -- nil except after an explicit dir
5345         state.sim = item -- so we move sim 'inside' the group
5346       else
5347         state.sim = state.sim or item
5348       end
5349       state.eim = item
5350     elseif d == 'r' and state.sim then
5351       head, state = insert_implicit(head, state, outer)
5352     elseif d == 'r' then
5353       state.sim, state.eim = nil, nil
5354     end
5355   end
5356
5357   if isdir then
5358     last = d           -- Don't search back - best save now
5359   elseif d == 'on' and state.san then
5360     state.san = state.san or item
5361     state.ean = item
5362   end
5363
5364 end
5365
5366 return node.prev(head) or head
5367 end
5368 /basic)

```

16 Data for CJK

It is a boring file and it is not shown here (see the generated file), but here is a sample:

```

[0x0021]={c='ex'},
[0x0024]={c='pr'},

```

```
[0x0025]={c='po'},
[0x0028]={c='op'},
[0x0029]={c='cp'},
[0x002B]={c='pr'},
```

For the meaning of these codes, see the Unicode standard.

17 The ‘nil’ language

This ‘language’ does nothing, except setting the hyphenation patterns to nohyphenation. For this language currently no special definitions are needed or available.

The macro `\LdfInit` takes care of preventing that this file is loaded more than once, checking the category code of the `@` sign, etc.

```
5369 ⟨*nil⟩
5370 \ProvidesLanguage{nil}[⟨⟨date⟩⟩ ⟨⟨version⟩⟩ Nil language]
5371 \LdfInit{nil}{datenil}
```

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

```
5372 \ifx\l@nil\undefined
5373   \newlanguage\l@nil
5374   \@namedef{bbl@hyphendata@the\l@nil}{}{}{}% Remove warning
5375   \let\bbl@elt\relax
5376   \edef\bbl@languages{% Add it to the list of languages
5377     \bbl@languages\bbl@elt{nil}{the\l@nil}{}{}}
5378 \fi
```

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

```
5379 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}
```

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

```
\captionnil
\datenil
5380 \let\captionnil\@empty
5381 \let\datenil\@empty
```

The macro `\ldf@finish` takes care of looking for a configuration file, setting the main language to be switched on at `\begin{document}` and resetting the category code of `@` to its original value.

```
5382 \ldf@finish{nil}
5383 ⟨/nil⟩
```

18 Support for Plain T_EX (plain.def)

18.1 Not renaming hyphen.tex

As Don Knuth has declared that the filename `hyphen.tex` may only be used to designate *his* version of the american English hyphenation patterns, a new solution has to be found in order to be able to load hyphenation patterns for other languages in a plain-based T_EX-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 `iniTEX`, you will get a file called either `bplain.fmt` or `blplain.fmt`, which you can use as replacements for `plain.fmt` and `lplain.fmt`. As these files are going to be read as the first thing `iniTEX` sees, we need to set some category codes just to be able to change the definition of `\input`

```
5384 <*bplain | blplain>
5385 \catcode`\{=1 % left brace is begin-group character
5386 \catcode`\}=2 % right brace is end-group character
5387 \catcode`\#=6 % hash mark is macro parameter character
```

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

```
5388 \openin 0 hyphen.cfg
```

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

```
5389 \ifeof0
5390 \else
```

When `hyphen.cfg` could be opened we make sure that *it* will be read instead of the file `hyphen.tex` which should (according to Don Knuth’s ruling) contain the american English hyphenation patterns and nothing else.

We do this by first saving the original meaning of `\input` (and I use a one letter control sequence for that so as not to waste multi-letter control sequence on this in the format).

```
5391 \let\input
```

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

```
5392 \def\input #1 {%
5393   \let\input\input
5394   \input hyphen.cfg
```

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

```
5395   \let\input\undefined
5396 }
5397 \fi
5398 </bplain | blplain>
```

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

```
5399 <bplain>\input plain.tex
5400 <blplain>\input 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.

```
5401 <bplain>\def\fmtname{babel-plain}
5402 <blplain>\def\fmtname{babel-lplain}
```

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

18.2 Emulating some L^AT_EX features

The following code duplicates or emulates parts of L^AT_EX 2_ε that are needed for babel.

```
5403 (*plain)
5404 \def\@empty{}
5405 \def\loadlocalcfg#1{%
5406   \openin0#1.cfg
5407   \ifeof0
5408     \closein0
5409   \else
5410     \closein0
5411     {\immediate\write16{*****}%
5412      \immediate\write16{* Local config file #1.cfg used}%
5413      \immediate\write16{*}%
5414     }
5415   \input #1.cfg\relax
5416   \fi
5417   \@endoflfd}
```

18.3 General tools

A number of L^AT_EX macro's that are needed later on.

```
5418 \long\def\@firstofone#1{#1}
5419 \long\def\@firstoftwo#1#2{#1}
5420 \long\def\@secondoftwo#1#2{#2}
5421 \def\@nnil{\@nil}
5422 \def\@gobbletwo#1#2{}
5423 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}{}}
5424 \def\@star@or@long#1{%
5425   \@ifstar
5426   {\let\l@ngrel@x\relax#1}%
5427   {\let\l@ngrel@x\long#1}}
5428 \let\l@ngrel@x\relax
5429 \def\@car#1#2\@nil{#1}
5430 \def\@cdr#1#2\@nil{#2}
5431 \let\@typeset@protect\relax
5432 \let\protected@edef\edef
5433 \long\def\@gobble#1{}
5434 \edef\@backslashchar{\expandafter\@gobble\string\}
5435 \def\strip@prefix#1>{}
5436 \def\g@addto@macro#1#2{%
5437   \toks@\expandafter{#1#2}%
5438   \xdef#1{\the\toks@}}
5439 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
5440 \def\@nameuse#1{\csname #1\endcsname}
5441 \def\@ifundefined#1{%
5442   \expandafter\ifx\csname#1\endcsname\relax
5443     \expandafter\@firstoftwo
5444   \else
5445     \expandafter\@secondoftwo
5446   \fi}
5447 \def\@expandtwoargs#1#2#3{%
5448   \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
5449 \def\zap@space#1 #2{%
5450   #1%
5451   \ifx#2\@empty\else\expandafter\zap@space\fi
5452   #2}
```


L^AT_EX 2_ε has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```
5453 \ifx\@preamblecmds\undefined
5454   \def\@preamblecmds{}
5455 \fi
5456 \def\@onlypreamble#1{%
5457   \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
5458     \@preamblecmds\do#1}}
5459 \@onlypreamble\@onlypreamble
```

Mimick L^AT_EX's `\AtBeginDocument`; for this to work the user needs to add `\begin{document}` to his file.

```
5460 \def\begin{document}{%
5461   \@begin{document}hook
5462   \global\let\@begin{document}hook\undefined
5463   \def\do##1{\global\let##1\undefined}%
5464   \@preamblecmds
5465   \global\let\do\noexpand}

5466 \ifx\@begin{document}hook\undefined
5467   \def\@begin{document}hook{}
5468 \fi
5469 \@onlypreamble\@begin{document}hook
5470 \def\AtBeginDocument{\g@addto@macro\@begin{document}hook}
```

We also have to mimick L^AT_EX's `\AtEndOfPackage`. Our replacement macro is much simpler; it stores its argument in `\@endoflfd`.

```
5471 \def\AtEndOfPackage#1{\g@addto@macro\@endoflfd{#1}}
5472 \@onlypreamble\AtEndOfPackage
5473 \def\@endoflfd{}
5474 \@onlypreamble\@endoflfd
5475 \let\bbl@afterlang\empty
5476 \chardef\bbl@opt@hyphenmap\z@
```

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

```
5477 \ifx\if@filesw\undefined
5478   \expandafter\let\csname if@filesw\expandafter\endcsname
5479     \csname iffalse\endcsname
5480 \fi
```

Mimick L^AT_EX's commands to define control sequences.

```
5481 \def\newcommand{\@star@or@long\new@command}
5482 \def\new@command#1{%
5483   \@testopt{\@newcommand#1}0}
5484 \def\@newcommand#1[#2]{%
5485   \@ifnextchar [{\@xargdef#1[#2]}%
5486     {\@argdef#1[#2]}}
5487 \long\def\@argdef#1[#2]#3{%
5488   \@yargdef#1\@ne{#2}{#3}}
5489 \long\def\@xargdef#1[#2][#3]#4{%
5490   \expandafter\def\expandafter#1\expandafter{%
5491     \expandafter\@protected@testopt\expandafter #1%
5492     \csname\string#1\expandafter\endcsname{#3}}%
5493   \expandafter\@yargdef\csname\string#1\endcsname
5494   \tw@{#2}{#4}}
5495 \long\def\@yargdef#1#2#3{%
5496   \@tempcnta#3\relax
5497   \advance \@tempcnta \@ne
```

```

5498 \let\@hash@\relax
5499 \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
5500 \@tempcntb #2%
5501 \@whilenum\@tempcntb <\@tempcnta
5502 \do{%
5503   \edef\reserved@a{\reserved@a\@hash@\the\@tempcntb}%
5504   \advance\@tempcntb \@ne}%
5505 \let\@hash@###
5506 \l@ngrel@x\expandafter\def\expandafter#1\reserved@a}
5507 \def\providecommand{\@star@or@long\provide@command}
5508 \def\provide@command#1{%
5509   \begingroup
5510     \escapechar\m@ne\xdef\@gtempa{\string#1}%
5511   \endgroup
5512   \expandafter\@ifundefined\@gtempa
5513     {\def\reserved@a{\new@command#1}}%
5514     {\let\reserved@a\relax
5515      \def\reserved@a{\new@command\reserved@a}}%
5516   \reserved@a}%

5517 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
5518 \def\declare@robustcommand#1{%
5519   \edef\reserved@a{\string#1}%
5520   \def\reserved@b{#1}%
5521   \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
5522   \edef#1{%
5523     \ifx\reserved@a\reserved@b
5524       \noexpand\x@protect
5525       \noexpand#1%
5526     \fi
5527     \noexpand\protect
5528     \expandafter\noexpand\csname
5529       \expandafter\@gobble\string#1 \endcsname
5530   }%
5531   \expandafter\new@command\csname
5532     \expandafter\@gobble\string#1 \endcsname
5533 }
5534 \def\x@protect#1{%
5535   \ifx\protect\@typeset@protect\else
5536     \@x@protect#1%
5537   \fi
5538 }
5539 \def\@x@protect#1\fi#2#3{%
5540   \fi\protect#1%
5541 }

```

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

```

5542 \def\bbl@tempa{\csname newif\endcsname\ifin@}
5543 \ifx\in@\@undefined
5544   \def\in@#1#2{%
5545     \def\in@@##1##2##3\in@@{%
5546       \ifx\in@@##2\in@false\else\in@true\fi}%
5547     \in@@#2#1\in@\in@@}
5548 \else
5549   \let\bbl@tempa\@empty
5550 \fi

```

5551 \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).

5552 \def\@ifpackagewith#1#2#3#4{#3}

The \LaTeX macro \@ifl@aded checks whether a file was loaded. This functionality is not needed for plain \TeX but we need the macro to be defined as a no-op.

5553 \def\@ifl@aded#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 2 ϵ versions; just enough to make things work in plain \TeX environments.

```
5554 \ifx\@tempcnta\@undefined
5555   \csname newcount\endcsname\@tempcnta\relax
5556 \fi
5557 \ifx\@tempcntb\@undefined
5558   \csname newcount\endcsname\@tempcntb\relax
5559 \fi
```

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

```
5560 \ifx\bye\@undefined
5561   \advance\count10 by -2\relax
5562 \fi
5563 \ifx\@ifnextchar\@undefined
5564   \def\@ifnextchar#1#2#3{%
5565     \let\reserved@d=#1%
5566     \def\reserved@a{#2}\def\reserved@b{#3}%
5567     \futurelet\@let@token\@ifnch}
5568   \def\@ifnch{%
5569     \ifx\@let@token\@sptoken
5570       \let\reserved@c\@xifnch
5571     \else
5572       \ifx\@let@token\reserved@d
5573         \let\reserved@c\reserved@a
5574       \else
5575         \let\reserved@c\reserved@b
5576       \fi
5577     \fi
5578     \reserved@c}
5579   \def\:{\let\@sptoken= } \: % this makes \@sptoken a space token
5580   \def\:\@xifnch\ \expandafter\def\:{\futurelet\@let@token\@ifnch}
5581 \fi
5582 \def\@testopt#1#2{%
5583   \@ifnextchar[#{1}{#1[#2]}}
5584 \def\@protected@testopt#1{%
5585   \ifx\protect\@typeset@protect
5586     \expandafter\@testopt
5587   \else
5588     \@x@protect#1%
5589   \fi}
5590 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
5591   #2\relax}\fi}
```

```

5592 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
5593         \else\expandafter\@gobble\fi{#1}}

```

18.4 Encoding related macros

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

```

5594 \def\DeclareTextCommand{%
5595     \@dec@text@cmd\providecommand
5596 }
5597 \def\ProvideTextCommand{%
5598     \@dec@text@cmd\providecommand
5599 }
5600 \def\DeclareTextSymbol#1#2#3{%
5601     \@dec@text@cmd\chardef#1{#2}#3\relax
5602 }
5603 \def\@dec@text@cmd#1#2#3{%
5604     \expandafter\def\expandafter#2%
5605         \expandafter{%
5606             \csname#3-cmd\expandafter\endcsname
5607             \expandafter#2%
5608             \csname#3\string#2\endcsname
5609         }%
5610 %   \let\@ifdefinable\@rc@ifdefinable
5611     \expandafter#1\csname#3\string#2\endcsname
5612 }
5613 \def\@current@cmd#1{%
5614     \ifx\protect\@typeset@protect\else
5615         \noexpand#1\expandafter\@gobble
5616     \fi
5617 }
5618 \def\@changed@cmd#1#2{%
5619     \ifx\protect\@typeset@protect
5620         \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
5621             \expandafter\ifx\csname ?\string#1\endcsname\relax
5622                 \expandafter\def\csname ?\string#1\endcsname{%
5623                     \@changed@x@err{#1}%
5624                 }%
5625             \fi
5626             \global\expandafter\let
5627                 \csname\cf@encoding \string#1\expandafter\endcsname
5628                 \csname ?\string#1\endcsname
5629             \fi
5630             \csname\cf@encoding\string#1%
5631                 \expandafter\endcsname
5632         \else
5633             \noexpand#1%
5634         \fi
5635 }
5636 \def\@changed@x@err#1{%
5637     \errhelp{Your command will be ignored, type <return> to proceed}%
5638     \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
5639 \def\DeclareTextCommandDefault#1{%
5640     \DeclareTextCommand#1?%
5641 }
5642 \def\ProvideTextCommandDefault#1{%
5643     \ProvideTextCommand#1?%
5644 }
5645 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd

```

```

5646 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
5647 \def\DeclareTextAccent#1#2#3{%
5648   \DeclareTextCommand#1{#2}[1]{\accent#3 #1}
5649 }
5650 \def\DeclareTextCompositeCommand#1#2#3#4{%
5651   \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
5652   \edef\reserved@b{\string#1}%
5653   \edef\reserved@c{%
5654     \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
5655   \ifx\reserved@b\reserved@c
5656     \expandafter\expandafter\expandafter\ifx
5657       \expandafter\@car\reserved@a\relax\relax\@nil
5658       \@text@composite
5659     \else
5660       \edef\reserved@b##1{%
5661         \def\expandafter\noexpand
5662           \csname#2\string#1\endcsname####1{%
5663             \noexpand\@text@composite
5664             \expandafter\noexpand\csname#2\string#1\endcsname
5665             ####1\noexpand\@empty\noexpand\@text@composite
5666             {##1}%
5667           }%
5668         }%
5669       \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
5670     \fi
5671     \expandafter\def\csname\expandafter\string\csname
5672       #2\endcsname\string#1-\string#3\endcsname{#4}
5673   \else
5674     \errhelp{Your command will be ignored, type <return> to proceed}%
5675     \errmessage{\string\DeclareTextCompositeCommand\space used on
5676       inappropriate command \protect#1}
5677   \fi
5678 }
5679 \def\@text@composite#1#2#3\@text@composite{%
5680   \expandafter\@text@composite@x
5681   \csname\string#1-\string#2\endcsname
5682 }
5683 \def\@text@composite@x#1#2{%
5684   \ifx#1\relax
5685     #2%
5686   \else
5687     #1%
5688   \fi
5689 }
5690 %
5691 \def\@strip@args#1:#2-#3\@strip@args{#2}
5692 \def\DeclareTextComposite#1#2#3#4{%
5693   \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
5694   \bgroup
5695     \lccode`\@=#4%
5696     \lowercase{%
5697       \egroup
5698       \reserved@a @%
5699     }%
5700 }
5701 %
5702 \def\UseTextSymbol#1#2{%
5703   \let\@curr@enc\cf@encoding
5704   \@use@text@encoding{#1}%

```

```

5705 #2%
5706 % \@use@text@encoding\@curr@enc
5707 }
5708 \def\UseTextAccent#1#2#3{%
5709 % \let\@curr@enc\cf@encoding
5710 % \@use@text@encoding{#1}%
5711 % #2{\@use@text@encoding\@curr@enc\selectfont#3}%
5712 % \@use@text@encoding\@curr@enc
5713 }
5714 \def\@use@text@encoding#1{%
5715 % \edef\font@name{#1}%
5716 % \xdef\font@name{%
5717 % \csname\curr@fontshape/\fontsize\endcsname
5718 % }%
5719 % \pickup@font
5720 % \font@name
5721 % \@enc@update
5722 }
5723 \def\DeclareTextSymbolDefault#1#2{%
5724 % \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
5725 }
5726 \def\DeclareTextAccentDefault#1#2{%
5727 % \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
5728 }
5729 \def\cf@encoding{OT1}

```

Currently we only use the $\LaTeX 2_{\epsilon}$ method for accents for those that are known to be made active in *some* language definition file.

```

5730 \DeclareTextAccent{"}{OT1}{127}
5731 \DeclareTextAccent{'}{OT1}{19}
5732 \DeclareTextAccent{^}{OT1}{94}
5733 \DeclareTextAccent`}{OT1}{18}
5734 \DeclareTextAccent~}{OT1}{126}

```

The following control sequences are used in `babel.def` but are not defined for PLAIN \TeX .

```

5735 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
5736 \DeclareTextSymbol{\textquotedblright}{OT1}{`\"}
5737 \DeclareTextSymbol{\textquoteleft}{OT1}{``}
5738 \DeclareTextSymbol{\textquoteright}{OT1}{``}
5739 \DeclareTextSymbol{\i}{OT1}{16}
5740 \DeclareTextSymbol{\ss}{OT1}{25}

```

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

```

5741 \ifx\scriptsize\@undefined
5742 % \let\scriptsize\sevenrm
5743 \fi
5744 </plain>

```

19 Acknowledgements

I would like to thank all who volunteered as β -testers for their time. Michel Goossens supplied contributions for most of the other languages. Nico Poppelier helped polish the text of the documentation and supplied parts of the macros for the Dutch language. Paul Wackers and Werenfried Spit helped find and repair bugs. During the further development of the babel system I received much help from Bernd Raichle, for which I am grateful.

References

- [1] Huda Smitshuijzen Abifares, *Arabic Typography*, Saqi, 2001.
- [2] Johannes Braams, Victor Eijkhout and Nico Poppelier, *The development of national \LaTeX styles*, *TUGboat* 10 (1989) #3, p. 401–406.
- [3] Yannis Haralambous, *Fonts & Encodings*, O'Reilly, 2007.
- [4] Donald E. Knuth, *The $T_{\text{E}}X$ book*, Addison-Wesley, 1986.
- [5] Jukka K. Korpela, *Unicode Explained*, O'Reilly, 2006.
- [6] Leslie Lamport, *\LaTeX , A document preparation System*, Addison-Wesley, 1986.
- [7] Leslie Lamport, in: $T_{\text{E}}X$ hax Digest, Volume 89, #13, 17 February 1989.
- [8] Ken Lunde, *CJKV Information Processing*, O'Reilly, 2nd ed., 2009.
- [9] Hubert Partl, *German $T_{\text{E}}X$* , *TUGboat* 9 (1988) #1, p. 70–72.
- [10] Joachim Schrod, *International \LaTeX is ready to use*, *TUGboat* 11 (1990) #1, p. 87–90.
- [11] Apostolos Syropoulos, Antonis Tsolomitis and Nick Sofroniu, *Digital typography using \LaTeX* , Springer, 2002, p. 301–373.
- [12] K.F. Treebus. *Tekstwijzer, een gids voor het grafisch verwerken van tekst*, SDU Uitgeverij ('s-Gravenhage, 1988).