

Babel

Version 3.23.1391

2018/09/06

Original author

Johannes L. Braams

Current maintainer

Javier Bezos

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

This manual describes babel, a package that makes use of the capabilities of \TeX version 3 and, to some extent, xetex and luatex, to provide an environment in which documents can be typeset in a language other than US English, or in more than one language or script.

Current development is focused on Unicode engines (Xe \TeX and Lua \TeX) and the so-called *complex scripts*. New features related to font selection, bidi writing and the like will be added incrementally.

Babel provides support (total or partial) for about 200 languages, either as a “classical” package option or as an ini file. Furthermore, new languages can be created from scratch easily.

Contents

I	User guide	4
1	The user interface	4
1.1	Monolingual documents	4
1.2	Multilingual documents	5
1.3	Modifiers	6
1.4	xelatex and lualatex	6
1.5	Troubleshooting	7
1.6	Plain	8
1.7	Basic language selectors	8
1.8	Auxiliary language selectors	9
1.9	More on selection	10
1.10	Shorthands	11
1.11	Package options	14
1.12	The base option	16
1.13	ini files	17
1.14	Selecting fonts	23
1.15	Modifying a language	25
1.16	Creating a language	25
1.17	Digits	27
1.18	Getting the current language name	28
1.19	Hyphenation tools	28
1.20	Selecting scripts	29
1.21	Selecting directions	30
1.22	Language attributes	34
1.23	Hooks	34
1.24	Languages supported by babel with ldf files	35
1.25	Tips, workarounds, know issues and notes	36
1.26	Current and future work	38
1.27	Tentative and experimental code	38
2	Loading languages with language.dat	39
2.1	Format	39
3	The interface between the core of babel and the language definition files	40
3.1	Guidelines for contributed languages	41
3.2	Basic macros	41
3.3	Skeleton	43
3.4	Support for active characters	43
3.5	Support for saving macro definitions	44
3.6	Support for extending macros	44
3.7	Macros common to a number of languages	44
3.8	Encoding-dependent strings	45
4	Changes	48
4.1	Changes in babel version 3.9	48
II	The code	49
5	Identification and loading of required files	49
6	locale directory	49

7	Tools	50
7.1	Multiple languages	53
8	The Package File (\LaTeX, babel.sty)	54
8.1	base	55
8.2	key=value options and other general option	56
8.3	Conditional loading of shorthands	58
8.4	Language options	59
9	The kernel of Babel (babel.def, common)	62
9.1	Tools	62
9.2	Hooks	64
9.3	Setting up language files	66
9.4	Shorthands	68
9.5	Language attributes	77
9.6	Support for saving macro definitions	79
9.7	Short tags	80
9.8	Hyphens	80
9.9	Interword spacing in Southeast Asian scripts	82
9.10	Multiencoding strings	82
9.11	Macros common to a number of languages	88
9.12	Making glyphs available	88
9.12.1	Quotation marks	88
9.12.2	Letters	90
9.12.3	Shorthands for quotation marks	91
9.12.4	Umlauts and tremas	91
9.13	Layout	93
9.14	Creating languages	94
10	The kernel of Babel (babel.def, only \LaTeX)	101
10.1	The redefinition of the style commands	101
10.2	Cross referencing macros	101
10.3	Marks	105
10.4	Preventing clashes with other packages	106
10.4.1	ifthen	106
10.4.2	varioref	107
10.4.3	hhline	107
10.4.4	hyperref	108
10.4.5	fancyhdr	108
10.5	Encoding and fonts	108
10.6	Basic bidi support	110
10.7	Local Language Configuration	113
11	Multiple languages (switch.def)	113
11.1	Selecting the language	114
11.2	Errors	122
12	Loading hyphenation patterns	123
13	Font handling with fontspec	128
14	Hooks for XeTeX and LuaTeX	131
14.1	XeTeX	131
14.2	Layout	133
14.3	LuaTeX	136

15 Southeast Asian scripts	141
15.1 Layout	142
15.2 Auto bidi with basic-r	144
16 The ‘nil’ language	155
17 Support for Plain T_EX (plain.def)	155
17.1 Not renaming hyphen.tex	155
17.2 Emulating some L ^A T _E X features	156
17.3 General tools	157
17.4 Encoding related macros	160
18 Acknowledgements	163

Troubleshooting

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

Part I

User guide

- This user guide focuses on \LaTeX . There are also some notes on its use with Plain \TeX .
- Changes and new features with relation to version 3.8 are highlighted with **New X.XX**. The most recent features could be still unstable. Please, report any issues you find on <https://github.com/latex3/latex2e/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 on <https://github.com/latex3/latex2e/tree/master/required/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 will *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.

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):

```
\documentclass{article}

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

\usepackage[french]{babel}

\begin{document}

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

\end{document}
```

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

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

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, T_EXLive, etc.) for further info about how to configure it.

1.2 Multilingual documents

In multilingual documents, just use several options. The last one is considered the main language, activated by default. Sometimes, the main language changes the document layout (eg, spanish and french).

EXAMPLE In L^AT_EX, the preamble of the document:

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

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

You can also set the main language explicitly:

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

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

```
\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.

```
\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}
```

1.3 Modifiers

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

1.4 xelatex and lualatex

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

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

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

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

```
\documentclass{article}

\usepackage[vietnamese,danish]{babel}

\begin{document}

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

\selectlanguage{vietnamese}

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

\end{document}
```

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

```
\documentclass{article}

\usepackage[russian]{babel}

\babelfont{rm}{DejaVu Serif}

\begin{document}

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

\end{document}
```

1.5 Troubleshooting

- Loading directly sty files in 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

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”.


```
! 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.

1.6 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

1.7 Basic language selectors

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

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

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

```
\selectlanguage{german}
```

This command can be used as environment, too.

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

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

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

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

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

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

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

1.8 Auxiliary language selectors

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

The environment `otherlanguage` does basically the same as `\selectlanguage`, except the 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}` `{\language} ... \end{hyphenrules}`

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

encodings of characters like, say, ' done by some languages (eg, italian, french, ukraineb). To set hyphenation exceptions, use `\babelhyphenation` (see below).

1.9 More on selection

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

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

EXAMPLE With

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

you can write

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

and

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

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

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

`\babelensure` $[\text{include}=\langle commands \rangle, \text{exclude}=\langle commands \rangle, \text{fontenc}=\langle encoding \rangle]\{\langle language \rangle\}$

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

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

Of course, $\text{T}_{\text{E}}\text{X}$ 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.

1.10 Shorthands

A *shorthand* is a sequence of one or two characters that expands to arbitrary \TeX code. Shorthands can be used for different kinds of things, as for example: (1) in some languages shorthands such as "a are defined to be able to hyphenate the word if the encoding is OT1; (2) in some languages shorthands such as ! are used to insert the right amount of white space; (3) several kinds of discretionary 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 `\kernbcode`, and `luatex` can manipulate the glyph list. Tools for point 3 can be still very useful in general. There are three levels of shorthands: *user*, *language*, and *system* (by order of precedence). Version 3.9 introduces the *language user* level on top of the user level, as described below. In most cases, you will use only shorthands provided by languages.

NOTE Note the following:

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

A typical error when using shorthands is the following:

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

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

`\shorthandon` $\{\langle shorthands-list \rangle\}$

\shorthandoff `*{\<shorthands-list>}`

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

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

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

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

\useshorthands `*{\<char>}`

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*{\<char>}` is provided, which makes sure shorthands are always activated.

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

\defineshorthand `[\<language>,\<language>,...]{\<shorthand>}{\<code>}`

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

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

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

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

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

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

⁵With it encoded string may not work as expected.

```
\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.

\aliasshorthand $\{\langle original \rangle\}\{\langle alias \rangle\}$

The command `\aliasshorthand` can be used to let another character perform the same functions as the default shorthand character. If one prefers for example to use the character / over " in typing Polish texts, this can be achieved by entering `\aliasshorthand{/}{/}`.

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

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

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

WARNING Shorthands remember somehow the original character, and the fallback value is that of the latter. So, in this example, if no shorthand 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.

\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 with, for example, `\usesshorthands`.)

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

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

`\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 you own user shorthands provided they do not overlap.)

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

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

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` $\{\langle character \rangle\}\{\langle true \rangle\}\{\langle false \rangle\}$

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

1.11 Package options

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

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

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

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.

⁷Thanks to Enrico Gregorio

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

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 \LaTeX before they are passed to the package and therefore they will not be recognized); however, t is provided for the common case of ~ (as well as c for not so common case of the comma). With shorthands=off no language shorthands are defined. As some languages use this mechanism for tools not available otherwise, a macro \babelshorthand is defined, which allows using them; see above.

safe= none | ref | bib

Some \LaTeX macros are redefined so that using shorthands is safe. With safe=bib only \nocite, \bibcite and \bibitem are redefined. With safe=ref only \newlabel, \ref and \pageref are redefined (as well as a few macros from varioref and ifthen). With safe=none no macro is redefined. This option is strongly recommended, because a good deal of incompatibilities and errors are related to these redefinitions – of course, in such a case you cannot use shorthands in these macros, but this is not a real problem (just use “allowed” characters).

math= active | normal

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

config= $\langle file \rangle$

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

main= $\langle language \rangle$

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

headfoot= $\langle language \rangle$

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

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

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

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

silent New 3.9l No warnings and no *infos* are written to the log file.⁹

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

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

New 3.9g Sets the behavior of case mapping for hyphenation, provided the language defines it.¹⁰ It can take the following values:

off deactivates this feature and no case mapping is applied;

first sets it at the first switching commands in the current or parent scope (typically, when the aux file is first read and at `\begin{document}`), but also the first `\selectlanguage` in the preamble), and it's the default if a single language option has been stated;¹¹

select sets it only at `\selectlanguage`;

other also sets it at `otherlanguage`;

other* also sets it at `otherlanguage*` as well as in heads and foots (if the option `headfoot` is used) and in auxiliary files (ie, at `\select@language`), and it's the default if several language options have been stated. The option `first` can be regarded as an optimized version of `other*` for monolingual documents.¹²

bidi= default | basic | basic-r

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

layout=

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

1.12 The base option

With this package option `babel` just loads some basic macros (those in `switch.def`), defines `\AfterBabelLanguage` and exits. It also selects the hyphenations 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 hyphenations patterns of a single language, too.

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

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 paragraph, but if either xetex or luatex change this behavior it might be added. On the other hand, `other` is provided even if I [JBL] think it isn't really useful, but who knows.

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

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

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

EXAMPLE Consider two languages `foo` and `bar` defining the same `\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}
```

1.13 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 development (in other words, `\babelprovide` is mainly intended 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.

```
\documentclass{book}

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

\babelfont{rm}{DejaVu Sans}

\begin{document}

\tableofcontents

\chapter{სამზარეულო და სუფრის ტრადიციები}

ქართული ტრადიციული სამზარეულო ერთ-ერთი უმდიდრესია მთელ მსოფლიოში.

\end{document}
```

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

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

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}
nd	North Ndebele	sr-Cyrl-XK	Serbian ^{ul}
ne	Nepali	sr-Cyrl	Serbian ^{ul}
nl	Dutch ^{ul}	sr-Latn-BA	Serbian ^{ul}
nmg	Kwasio	sr-Latn-ME	Serbian ^{ul}
nn	Norwegian Nynorsk ^{ul}	sr-Latn-XK	Serbian ^{ul}
nnh	Ngiemboon	sr-Latn	Serbian ^{ul}
nus	Nuer	sr	Serbian ^{ul}
nyn	Nyankole	sv	Swedish ^{ul}
om	Oromo	sw	Swahili
or	Odia	ta	Tamil ^u
os	Ossetic	te	Telugu ^{ul}
pa-Arab	Punjabi	teo	Teso
pa-Guru	Punjabi	th	Thai ^{ul}
pa	Punjabi	ti	Tigrinya

tk	Turkmen ^{ul}	wae	Walser
to	Tongan	xog	Soga
tr	Turkish ^{ul}	yav	Yangben
twq	Tasawaq	yi	Yiddish
tzm	Central Atlas Tamazight	yo	Yoruba
ug	Uyghur	yue	Cantonese
uk	Ukrainian ^{ul}	zgh	Standard Moroccan Tamazight
ur	Urdu ^{ul}		
uz-Arab	Uzbek	zh-Hans-HK	Chinese
uz-Cyrl	Uzbek	zh-Hans-MO	Chinese
uz-Latn	Uzbek	zh-Hans-SG	Chinese
uz	Uzbek	zh-Hans	Chinese
vai-Latn	Vai	zh-Hant-HK	Chinese
vai-Vaii	Vai	zh-Hant-MO	Chinese
vai	Vai	zh-Hant	Chinese
vi	Vietnamese ^{ul}	zh	Chinese
vun	Vunjo	zu	Zulu

In some contexts (currently `\babelfont`) an ini file may be loaded by its name. Here is the list of the names currently supported. With these languages, `\babelfont` 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 `\babelprovide` with a valueless `import`.

aghem	bemba
akan	ben
albanian	bengali
american	bodo
amharic	bosnian-cyrillic
arabic	bosnian-cyrl
arabic-algeria	bosnian-latin
arabic-DZ	bosnian-latn
arabic-morocco	bosnian
arabic-MA	brazilian
arabic-syria	breton
arabic-SY	british
armenian	bulgarian
assamese	burmese
asturian	canadian
asu	cantonese
australian	catalan
austrian	centralatlastamazight
azerbaijani-cyrillic	centralkurdish
azerbaijani-cyrl	chechen
azerbaijani-latin	cherokee
azerbaijani-latn	chiga
azerbaijani	chinese-hans-hk
bafia	chinese-hans-mo
bambara	chinese-hans-sg
basaa	chinese-hans
basque	chinese-hant-hk
belarusian	chinese-hant-mo

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
french-ch	lakota
french-lu	langi
french-luxembourg	lao
french-switzerland	latvian
french	lingala
friulian	lithuanian
fulah	lowersorbian
galician	lsorbian
ganda	lubakatanga
georgian	luo
german-at	luxembourgish
german-austria	luyia
german-ch	macedonian
german-switzerland	machame

makhuwameetto
makonde
malagasy
malay-bn
malay-brunei
malay-sg
malay-singapore
malay
malayalam
maltese
manx
marathi
masai
mazanderani
meru
meta
mexican
mongolian
morisyen
mundang
nama
nepali
newzealand
ngiemboon
ngomba
norsk
northernluri
northernsami
northndebele
norwegianbokmal
norwegiannynorsk
nswissgerman
nuer
nyankole
nynorsk
occitan
oriya
oromo
ossetic
pashto
persian
piedmontese
polish
portuguese-br
portuguese-brazil
portuguese-portugal
portuguese-pt
portuguese
punjabi-arab
punjabi-arabic
punjabi-gurmukhi
punjabi-guru
punjabi
quechua

romanian
romansh
rombo
rundi
russian
rwa
sakha
samburu
samin
sango
sangu
sanskrit-beng
sanskrit-bengali
sanskrit-deva
sanskrit-devanagari
sanskrit-gujarati
sanskrit-gujr
sanskrit-kannada
sanskrit-knda
sanskrit-malayalam
sanskrit-mlym
sanskrit-telu
sanskrit-telugu
sanskrit
scottishgaelic
sena
serbian-cyrillic-bosniaherzegovina
serbian-cyrillic-kosovo
serbian-cyrillic-montenegro
serbian-cyrillic
serbian-cyrl-ba
serbian-cyrl-me
serbian-cyrl-xk
serbian-cyrl
serbian-latin-bosniaherzegovina
serbian-latin-kosovo
serbian-latin-montenegro
serbian-latin
serbian-latn-ba
serbian-latn-me
serbian-latn-xk
serbian-latn
serbian
shambala
shona
sichuanyi
sinhala
slovak
slovene
slovenian
soga
somali
spanish-mexico
spanish-mx

spanish	usorbian
standardmoroccantamazight	uyghur
swahili	uzbek-arab
swedish	uzbek-arabic
swissgerman	uzbek-cyrillic
tachelhit-latin	uzbek-cyrl
tachelhit-latn	uzbek-latin
tachelhit-tfng	uzbek-latn
tachelhit-tifinagh	uzbek
tachelhit	vai-latin
taita	vai-latn
tamil	vai-vai
tasawaq	vai-vaii
telugu	vai
teso	vietnam
thai	vietnamese
tibetan	vunjo
tigrinya	walser
tongan	welsh
turkish	westernfrisian
turkmen	yangben
ukenglish	yiddish
ukrainian	yoruba
upporsorbian	zarma
urdu	zulu afrikaans
usenglish	

1.14 Selecting fonts

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

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

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

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

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

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

```
\documentclass{article}

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

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


```

\babelprovide[import]{hebrew}

\babelfont{rm}{FreeSerif}

\begin{document}

Svenska \foreignlanguage{hebrew}{עברית} svenska.

\end{document}

```

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

```

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

```

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

EXAMPLE Here is how to do it:

```

\babelfont{kai}{FandolKai}

```

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

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

```

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

```

This makes sure the OpenType script for Devanagari is `deva` and not `dev2` (luatex does not detect automatically the correct script¹⁴).

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 are passed. You must add them by hand. This is by design, for several reasons (for example, each font has its own set of features and a generic setting for several of them could be problematic, and also a “lower level” font selection is useful).

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

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

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).

1.15 Modifying a language

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

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

```
\addto\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\extrarussian{\mymacro}
```

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

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.

1.16 Creating a language

New 3.10 And what if there is no style for your language or none fits your needs? You may then define quickly a language with the help of the following macro in the preamble.

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

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

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

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

EXAMPLE If you need a language named arhinish:

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

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

import= *<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= \langle language-list \rangle

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 T_EX 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= \langle script-name \rangle

New 3.15 Sets the script name to be used by fontspec (eg, Devanagari). Overrides the value in the ini file. This value is particularly important because it sets the writing direction, so you must use it if for some reason the default value is wrong.

language= \langle language-name \rangle

New 3.15 Sets the language name to be used by fontspec (eg, Hindi). Overrides the value in the ini file. Not so important, but sometimes still relevant.

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

1.17 Digits

New 3.20 About thirty ini files define a field named digits.native. When it is present, two macros are created: \<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).

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*.

1.18 Getting the current language name

`\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` `{\language}{\true}{\false}`

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

WARNING The advice about `\language` also applies here – use `iflang` instead of `\iflanguage` if possible.

1.19 Hyphenation tools

`\babelhyphen` `*{\type}`

`\babelhyphen` `*{\text}`

New 3.9a It is customary to classify hyphens in two types: (1) *explicit* or *hard hyphens*, which in \TeX are entered as `-`, and (2) *optional* or *soft hyphens*, which are entered as `\-`. Strictly, a *soft hyphen* is not a hyphen, but just a breaking opportunity or, in \TeX terms, a “discretionary”; a *hard hyphen* is a hyphen with a breaking opportunity after it. A further type is a *non-breaking hyphen*, a hyphen without a breaking opportunity. In \TeX , `-` and `\-` forbid further breaking opportunities in the word. This is the desired behavior very often, but not always, and therefore many languages provide shorthands for these cases. Unfortunately, this has not been done consistently: for example, `-` in Dutch, Portugese, 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 provide 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, Portugese 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 enabling 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 \LaTeX : (1) the character used is that set for the current font, while in \LaTeX it is hardwired to `-` (a typical value); (2) the hyphen to be used in fonts with a negative `\hyphenchar` is `-`, like in \LaTeX , 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` [*$\langle language \rangle$* , *$\langle language \rangle$* , ...]{ *$\langle exceptions \rangle$* }

New 3.9a Sets hyphenation exceptions for the languages given or, without the optional argument, for *all* languages (eg, proper nouns or common loan words, and of course monolingual documents). Language exceptions take precedence over global ones. It can be used only in the preamble, and exceptions are set when the language is first selected, thus taking into account changes of `\lccodes`'s done in `\extras{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.

`\babelpatterns` [*$\langle language \rangle$* , *$\langle language \rangle$* , ...]{ *$\langle patterns \rangle$* }

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.

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.

1.20 Selecting scripts

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

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.

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.

encodings and fonts (for example, if the main latin encoding was LY1), and therefore it has been deprecated.¹⁷

`\ensureascii` $\langle text \rangle$

New 3.9i This macro makes sure $\langle text \rangle$ is typeset with a LICR-savvy encoding in the ASCII range. It is used to redefine \backslashTeX and \backslashLaTeX so that they are correctly typeset even with LGR or X2 (the complete list is stored in \backslashBabelNonASCII , 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 \backslashTeX and \backslashLaTeX 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 used for “ordinary” text, so they should be selected with some other encoding, but they are taken into account, just in case. The foregoing rules (which are applied “at begin document”) cover most of cases. No assumption is made on characters above 127, which may not follow the LICR conventions – the goal is just to ensure most of the ASCII letters and symbols are the right ones.

1.21 Selecting directions

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

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 very likely until (Northern) Winter. This applies to text, but **graphical** elements, including the picture environment and PDF or PS based graphics, are not yet correctly handled. Also, indexes and the like are under study, as well as math.

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

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 in typical cases.

New 3.19 Finally, `basic` supports both L and R text. (They are named `basic` mainly because they only consider the intrinsic direction of scripts and weak directionality.) There are samples on GitHub, under `/required/babel/samples`. See particularly `lua-bidibasic.tex` and `lua-secenum.tex`.

¹⁷But still defined for backwards compatibility.

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.¹⁸

```
\documentclass{article}

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

\babelprovide[import, main]{arabic}

\babelfont{rm}{FreeSerif}

\begin{document}

    وقد عرفت شبه جزيرة العرب طيلة العصر الهليني (الاجريقي) بـ
    Arabia أو Aravia (بالاغريقية Αραβία), 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 فصحى العصر \textit{fuṣḥā l-‘aṣr} (MSA) and
    فصحى التراث \textit{fuṣḥā t-turāth} (CA).

\end{document}
```

What `mapfont=direction` means is, ‘when a character has the same direction as the script for the “provided” language (arabic in this case), then change its font to that set for this language’ (here defined via `*arabic`, because `Crimson` does not provide Arabic letters). Note Hebrew and Arabic, following the Unicode rules, have different directions (internally ‘r’ and ‘al’, respectively).

¹⁸At the time of this writing some Arabic fonts are not rendered correctly by the default `luatex` font loader, with misplaced kerns inside some words, so double check the resulting text. Have a look at the workaround available on GitHub, under `/required/babel/samples`

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

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

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

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

New 3.16 *To be expanded.* Selects which layout elements are adapted in bidi documents, including some text elements. You may use several options with a comma-separated list (eg, `layout=counters.contents.sectioning`). This list will be expanded in future releases (tables, captions, etc.). 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.¹⁹

lists required in `xetex` and `pdftex`, but only in multilingual documents in `luatex`.

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 multilingual documents in all engines; you may use alternatively `\BabelFootnote` described below (what this options does exactly is also explained there).

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

tabular required in `luatex` for R tabular (it has been tested only with simple tables, so expect some readjustments in the future); ignored in `pdftex` or `xetex` (which will not support a similar option in the short term) **New 3.18** ,

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

`\babelsublr` $\{\langle lr\text{-}text\rangle\}$

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 $\{\langle lr\text{-}text\rangle\}$ in L mode if necessary. It's intended for what Unicode calls weak characters, because words are best set with the corresponding language. For this reason, there is no `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 behaviour 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` $\{\langle section\text{-}name\rangle\}$

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` $\{\langle cmd\rangle\}\{\langle local\text{-}language\rangle\}\{\langle before\rangle\}\{\langle after\rangle\}$

New 3.17 Something like:

```
\BabelFootnote{\parsfootnote}{\language}\{()\}
```

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

```
\footnote{(\foreignlanguage{\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}\{()\}%  
\BabelFootnote{\localfootnote}{\language}\{()\}%  
\BabelFootnote{\mainfootnote}\{()\}
```

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

¹⁹Next on the roadmap are counters and numeral systems in general. Expect some minor readjustments.

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

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

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

1.22 Language attributes

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

1.23 Hooks

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

\AddBabelHook `{\langle name \rangle}{\langle event \rangle}{\langle code \rangle}`

The same name can be applied to several events. Hooks may be enabled and disabled for all defined events with `\EnableBabelHook{\langle name \rangle}`, `\DisableBabelHook{\langle name \rangle}`. Names containing the string `babel` are reserved (they are used, for example, by `\useshortands*` to add a hook for the event `afterextras`). Current events are the following; in some of them you can use one to three T_EX parameters (`#1`, `#2`, `#3`), with the meaning given:

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

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

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

defaultcommands Used (locally) in `\StartBabelCommands`.

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

stopcommands Used to reset the the above, if necessary.

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

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

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

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

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

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

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

afterreset **New 3.9i** Executed when selecting a language just after `\originalTeX` is run and reset to its base value, before executing `\captions⟨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 files or to load nothing.

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

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

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

1.24 Languages supported by babel 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 upporsorbian
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). 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 *<file>*, which creates *<file>.tex*; you can then typeset the latter with \LaTeX .

NOTE Please, for info about the support in luatex for some complex scripts, see the wiki, on <https://github.com/latex3/latex2e/wiki/Babel:-Remarks-on-the-luatex-support-for-some-scripts>.

1.25 Tips, workarounds, know issues and notes

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

- Both `ltxdoc` and `babel` use `\AtBeginDocument` to change some catcodes, and `babel` reloads `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\extrarussian{\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.²⁰ 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 `\defineshortand`, 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`.

²⁰This 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.

1.26 Current and future work

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

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

Useful additions would be, for example, time, currency, addresses and personal names.²¹. But that is the easy part, because they don’t require modifying the \TeX internals.

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

1.27 Tentative and experimental code

Southeast Asian interword spacing

There is some rudimentary interword spacing for Thai, Lao and Khemer in luatex (provided there are hyphenation patterns). It is activated automatically if a language with one of these scripts are loaded with \babelprovide. They are just the first steps. See the sample on the babel repository.

Note you can easily add, modify and remove patterns with \babelpatterns. Even if there are no pattern for the language, you can add at least some typical cases.

Old 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}
```

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

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

²¹See for example POSIX, ISO 14652 and the Unicode Common Locale Data Repository (CLDR). Those system, however, have limited application to \TeX because their aim is just to display information and not fine typesetting.

2 Loading languages with `language.dat`

TeX and most engines based on it (pdfTeX, xetex, ϵ -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).²² 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).²³

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 stored²⁴. 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.²⁵ 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
```

²²This feature was added to 3.9o, but it was buggy. Both 3.9o and 3.9p are deprecated.

²³The 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`.

²⁴This is because different operating systems sometimes use very different file-naming conventions.

²⁵This 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 `\langle lang \rangle hyphenmins`, `\captions\langle lang \rangle`, `\date\langle lang \rangle`, `\extras\langle lang \rangle` and `\noextras\langle lang \rangle` (the last two may be left empty); where `\langle lang \rangle` is either the name of the language definition file or the name of the \LaTeX option that is to be used. These macros and their functions are discussed below. You must define all or none for a language (or a dialect); defining, say, `\date\langle lang \rangle` but not `\captions\langle lang \rangle` does not raise an error but can lead to unexpected results.
- When a language definition file is loaded, it can define `\l@\langle lang \rangle` to be a dialect of `\language0` when `\l@\langle lang \rangle` is undefined.
- Language names must be all lowercase. If an unknown language is selected, babel will attempt setting it after lowercasing its name.
- The semantics of modifiers is not defined (on purpose). In most cases, they will just be simple separated options (eg, `spanish`), but a language might require, say, a set of options organized as a tree with suboptions (in such a case, the recommended separator is `/`).

Some recommendations:

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

- Avoid adding things to `\noextras⟨lang⟩` except for `umlauthigh` and friends, `\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.²⁶
- 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 your need further assistance and technical advice in the development of language styles, I am willing to help you. And of course, you can make any suggestion you like.

3.2 Basic macros

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

- `\addlanguage` The macro `\addlanguage` is a non-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

²⁶But 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 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>

\bbbl@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>}
```

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`
`\bbl@deactivate`

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

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

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

`\bbl@add@special`
`\bbl@remove@special` The \TeX book states: “Plain \TeX includes a macro called `\dospecials` that is essentially a set macro, representing the set of all characters that have a special category code.” [2, p. 380] It is used to set text ‘verbatim’. To make this work if more characters get a special category code, you have to add this character to the macro `\dospecial`. \LaTeX adds another macro called `\@sanitize` representing the same character set, but without the curly braces. The macros `\bbl@add@special⟨char⟩` and `\bbl@remove@special⟨char⟩` add and remove the character `⟨char⟩` to these two sets.

3.5 Support for saving macro definitions

Language definition files may want to *redefine* macros that already exist. Therefore a mechanism for saving (and restoring) the original definition of those macros is provided. We provide two macros for this²⁷.

`\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`.

3.7 Macros common to a number of languages

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

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

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

`\set@low@box` For some languages, quotes need to be lowered to the baseline. For this purpose the macro

²⁷This mechanism was introduced by Bernd Raichle.

`\set@low@box` is available. It takes one argument and puts that argument in an `\hbox`, at the baseline. The result is available in `\box0` for further processing.

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

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

3.8 Encoding-dependent strings

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

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

It 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 take precedence (ie, it works much like `\providecommand`).

Encoding info is `charset=` followed by a `charset`, which if given sets how the strings should be traslated to the internal representation used by the engine, typically `utf8`, which is the only value supported currently (default is no traslations). Note `charset` is applied by `luatex` and `xetex` when reading the file, not when the macro or string is used in the document. A list of font encodings which the strings are expected to work with can be given after `fontenc=` (separated with spaces, if two or more) – recommended, but not mandatory, although blocks without this key are not taken into account if you have requested `strings=encoded`.

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

The $\langle category \rangle$ is either captions, date or extras. You must stick to these three categories, even if no error is raised when using other name.²⁸ It may be empty, too, but in such a case using $\backslash SetString$ is an error (but not $\backslash SetCase$).

```
\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\monthiname{Jänner}

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

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

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

\StartBabelCommands{german,austrian}{date}
\SetString\monthiiiname{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}
\SetString\today{\number\day.~%
\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 $\backslash \langle category \rangle \langle language \rangle$ 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).

²⁸In future releases further categories may be added.

\StartBabelCommands *{\language-list}{\category}{[\selector]}

The starred version just forces strings to take a value – if not set as package option, then the default for the engine is used. This is not done by default to prevent backward incompatibilities, but if you are creating a new language this version is better. It's up to the maintainers of the current languages to decide if using it is appropriate.²⁹

\EndBabelCommands Marks the end of the series of blocks.

\AfterBabelCommands {\code}

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

\SetString {\macro-name}{\string}

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

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

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

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

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

#1 is replaced by the roman numeral.

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

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

\StartBabelCommands{turkish}{}[
```

²⁹This replaces in 3.9g a short-lived \UseStrings which has been removed because it did not work.


```

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

\EndBabelCommands

```

(Note the mapping for OT1 is not complete.)

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

New 3.9g Case mapping serves in T_EX 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 T_EX 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 through 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 through 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 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

The code

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

5 Identification and loading of required files

Code documentation is still under revision.

The babel package after unpacking consists of the following files:

switch.def defines macros to set and switch languages.

babel.def defines the rest of macros. It has 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=3.23.1391>>
2 <<date=2018/09/06>>
```

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` 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`-statement³⁰. These macros will break if another `\if... \fi` statement appears in one of the arguments and it is not enclosed in braces.

`\bbl@afterfi`

```
26 \long\def\bbl@afterelse#1\else#2\fi{\fi#1}
27 \long\def\bbl@afterfi#1\fi{\fi#1}
```

`\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.

```
28 \def\bbl@tempa#1{%
29   \long\def\bbl@trim##1##2{%
30     \futurelet\bbl@trim@a\bbl@trim@c##2\@nil\@nil#1\@nil\relax{##1}}%
31   \def\bbl@trim@c{%
32     \ifx\bbl@trim@a\sptoken
33       \expandafter\bbl@trim@b
34     \else
35       \expandafter\bbl@trim@b\expandafter#1%
36     \fi}%
37   \long\def\bbl@trim@b#1##1 \@nil{\bbl@trim@i##1}}
38 \bbl@tempa{ }
39 \long\def\bbl@trim@i#1\@nil#2\relax#3{#3{#1}}
40 \long\def\bbl@trim@def#1{\bbl@trim{\def#1}}
```

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

```
41 \def\bbl@ifunset#1{%
42   \expandafter\ifx\csname#1\endcsname\relax
43     \expandafter\@firstoftwo
44   \else
45     \expandafter\@secondoftwo
46   \fi}
47 \bbl@ifunset{ifcsname}%
48 {}%
49 {\def\bbl@ifunset#1{%
50   \ifcsname#1\endcsname
51     \expandafter\ifx\csname#1\endcsname\relax
52       \bbl@afterelse\expandafter\@firstoftwo
53     \else
54       \bbl@afterfi\expandafter\@secondoftwo
55     \fi
56   \else
57     \expandafter\@firstoftwo
58   \fi}}
```

³⁰This code is based on code presented in TUGboat vol. 12, no2, June 1991 in “An expansion Power Lemma” by Sonja Maus.

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

```
59 \def\bbl@ifblank#1{%
60   \bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
61 \long\def\bbl@ifblank@i#1#2\@nil#3#4#5\@nil{#4}
```

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

```
62 \def\bbl@forkv#1#2{%
63   \def\bbl@kvcmd##1##2##3{#2}%
64   \bbl@kvnext#1,\@nil,}
65 \def\bbl@kvnext#1,{%
66   \ifx\@nil#1\relax\else
67     \bbl@ifblank{#1}{\bbl@forkv@eq#1=\@empty=\@nil{#1}}%
68     \expandafter\bbl@kvnext
69   \fi}
70 \def\bbl@forkv@eq#1=#2=#3\@nil#4{%
71   \bbl@trim@def\bbl@forkv@a{#1}%
72   \bbl@trim{\expandafter\bbl@kvcmd\expandafter{\bbl@forkv@a}}{#2}{#4}}
```

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

```
73 \def\bbl@vforeach#1#2{%
74   \def\bbl@forcmd##1{#2}%
75   \bbl@fornext#1,\@nil,}
76 \def\bbl@fornext#1,{%
77   \ifx\@nil#1\relax\else
78     \bbl@ifblank{#1}{\bbl@trim\bbl@forcmd{#1}}%
79     \expandafter\bbl@fornext
80   \fi}
81 \def\bbl@foreach#1{\expandafter\bbl@vforeach\expandafter{#1}}
```

`\bbl@replace`

```
82 \def\bbl@replace#1#2#3{% in #1 -> repl #2 by #3
83   \toks@{}}%
84 \def\bbl@replace@aux##1#2##2#2{%
85   \ifx\bbl@nil##2%
86     \toks@\expandafter{\the\toks@##1}%
87   \else
88     \toks@\expandafter{\the\toks@##1#3}%
89     \bbl@afterfi
90     \bbl@replace@aux##2#2%
91   \fi}%
92 \expandafter\bbl@replace@aux#1#2\bbl@nil#2%
93 \edef#1{\the\toks@}}
```

`\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.

```
94 \def\bbl@exp#1{%
95   \begingroup
96   \let\ \noexpand
97   \def\<##1>{\expandafter\noexpand\csname##1\endcsname}%
98   \edef\bbl@exp@aux{\endgroup#1}%
99   \bbl@exp@aux}
```

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.

```

100 \def\bbl@ifsamestring#1#2{%
101   \begingroup
102     \protected@edef\bbl@tempb{#1}%
103     \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
104     \protected@edef\bbl@tempc{#2}%
105     \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
106     \ifx\bbl@tempb\bbl@tempc
107       \aftergroup\@firstoftwo
108     \else
109       \aftergroup\@secondoftwo
110     \fi
111   \endgroup}
112 \chardef\bbl@engine=%
113 \ifx\directlua\@undefined
114   \ifx\XeTeXinputencoding\@undefined
115     \z@
116   \else
117     \tw@
118   \fi
119 \else
120   \@ne
121 \fi
122 <</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 .

```

123 <<*Make sure ProvidesFile is defined>> ≡
124 \ifx\ProvidesFile\@undefined
125   \def\ProvidesFile#1[#2 #3 #4]{%
126     \wlog{File: #1 #4 #3 <#2>}%
127     \let\ProvidesFile\@undefined}
128 \fi
129 <</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`.

```

130 <<*Load patterns in luatex>> ≡
131 \ifx\directlua\@undefined\else
132   \ifx\bbl@luapatterns\@undefined
133     \input luababel.def
134   \fi
135 \fi
136 <</Load patterns in luatex>>

```

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

```

137 <<*Load macros for plain if not LaTeX>> ≡
138 \ifx\AtBeginDocument\@undefined
139   \input plain.def\relax
140 \fi
141 <</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.

```
142 <<*Define core switching macros>> ≡
143 \ifx\language\undefined
144   \csname newcount\endcsname\language
145 \fi
146 <</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.

```
147 <<*Define core switching macros>> ≡
148 \ifx\newlanguage\undefined
149   \csname newcount\endcsname\last@language
150   \def\addlanguage#1{%
151     \global\advance\last@language\@ne
152     \ifnum\last@language<\@ccclvi
153       \else
154         \errmessage{No room for a new \string\language!}%
155       \fi
156       \global\chardef#1\last@language
157       \wlog{\string#1 = \string\language\the\last@language}}
158   \else
159     \countdef\last@language=19
160     \def\addlanguage{\alloc@9\language\chardef\@ccclvi}
161   \fi
162 <</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.

```

163 (*package)
164 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
165 \ProvidesPackage{babel}[\langle date \rangle \langle version \rangle The Babel package]
166 \@ifpackagewith{babel}{debug}
167   {\providecommand\bbl@trace[1]{\message{^^J[ #1 ]}}}%
168   \let\bbl@debug\@firstofone}
169   {\providecommand\bbl@trace[1]{}}%
170   \let\bbl@debug\gobble}
171 \ifx\bbl@switchflag@undefined % Prevent double input
172   \let\bbl@switchflag\relax
173   \input switch.def\relax
174 \fi
175 \langle Load patterns in luatex \rangle
176 \langle Basic macros \rangle
177 \def\AfterBabelLanguage#1{%
178   \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.

```

179 \ifx\bbl@languages\undefined\else
180   \begingroup
181     \catcode`\^^I=12
182     \@ifpackagewith{babel}{showlanguages}{%
183       \begingroup
184         \def\bbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}}%
185         \wlog{<*languages>}%
186         \bbl@languages
187         \wlog{</languages>}%
188       \endgroup}{%
189     \endgroup
190     \def\bbl@elt#1#2#3#4{%
191       \ifnum#2=\z@
192         \gdef\bbl@nulllanguage{#1}%
193         \def\bbl@elt##1##2##3##4{}%
194       \fi}%
195     \bbl@languages
196 \fi
197 \ifodd\bbl@engine
198   \let\bbl@tempa\relax
199   \@ifpackagewith{babel}{bidi=basic}%
200   {\def\bbl@tempa{basic}}%
201   {\@ifpackagewith{babel}{bidi=basic-r}%
202    {\def\bbl@tempa{basic-r}}}%
203   {}%
204 \ifx\bbl@tempa\relax\else
205   \let\bbl@beforeforeign\leavevmode

```



```

206 \AtEndOfPackage{\EnableBabelHook{babel-bidi}}%
207 \RequirePackage{luatexbase}%
208 \directlua{
209     require('babel-bidi.lua')
210     require('babel-bidi-\bbl@tempa.lua')
211     luatexbase.add_to_callback('pre_linebreak_filter',
212         Babel.pre_otfload_v,
213         'Babel.pre_otfload_v',
214         luatexbase.priority_in_callback('pre_linebreak_filter',
215             'luaotfload.node_processor') or nil)
216     luatexbase.add_to_callback('hpack_filter',
217         Babel.pre_otfload_h,
218         'Babel.pre_otfload_h',
219         luatexbase.priority_in_callback('hpack_filter',
220             'luaotfload.node_processor') or nil)
221 }
222 \fi
223 \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.

```

224 \bbl@trace{Defining option 'base'}
225 \@ifpackagewith{babel}{base}{%
226     \ifx\directlua\undefined
227         \DeclareOption*{\bbl@patterns{\CurrentOption}}%
228     \else
229         \DeclareOption*{\bbl@patterns@lua{\CurrentOption}}%
230     \fi
231     \DeclareOption{base}{}%
232     \DeclareOption{showlanguages}{}%
233     \ProcessOptions
234     \global\expandafter\let\csname opt@babel.sty\endcsname\relax
235     \global\expandafter\let\csname ver@babel.sty\endcsname\relax
236     \global\let@ifl@ter@\@ifl@ter
237     \def\@ifl@ter#1#2#3#4#5{\global\let@ifl@ter\@ifl@ter@@}%
238     \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.

```

239 \bbl@trace{key=value and another general options}
240 \bbl@csarg\let{tempa\expandafter}\csname opt@babel.sty\endcsname
241 \def\bbl@tempb#1.#2{%
242     #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%
243 \def\bbl@tempd#1.#2\@nnil{%
244     \ifx\@empty#2%
245         \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
246     \else
247         \in@{=}{#1}\ifin@
248         \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
249     \else
250         \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
251         \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
252     \fi

```

```

253 \fi}
254 \let\bbl@tempc\@empty
255 \bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty\@nnil}
256 \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.

```

257 \DeclareOption{KeepShorthandsActive}{}
258 \DeclareOption{activeacute}{}
259 \DeclareOption{activegrave}{}
260 \DeclareOption{debug}{}
261 \DeclareOption{noconfigs}{}
262 \DeclareOption{showlanguages}{}
263 \DeclareOption{silent}{}
264 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
265 <<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.

```

266 \let\bbl@opt@shorthands\@nnil
267 \let\bbl@opt@config\@nnil
268 \let\bbl@opt@main\@nnil
269 \let\bbl@opt@headfoot\@nnil
270 \let\bbl@opt@layout\@nnil

```

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

```

271 \def\bbl@tempa#1=#2\bbl@tempa{%
272 \bbl@csarg\ifx{opt@#1}\@nnil
273 \bbl@csarg\edef{opt@#1}{#2}%
274 \else
275 \bbl@error{%
276 Bad option `#1=#2'. Either you have misspelled the\\
277 key or there is a previous setting of `#1' {%
278 Valid keys are `shorthands', `config', `strings', `main', \\
279 `headfoot', `safe', `math', among others.}
280 \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.

```

281 \let\bbl@language@opts\@empty
282 \DeclareOption*{%
283 \bbl@xin@{\string=}{\CurrentOption}%
284 \ifin@
285 \expandafter\bbl@tempa\CurrentOption\bbl@tempa
286 \else
287 \bbl@add@list\bbl@language@opts{\CurrentOption}%
288 \fi}

```

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

```

289 \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=...`

```
290 \bbl@trace{Conditional loading of shorthands}
291 \def\bbl@sh@string#1{%
292   \ifx#1\@empty\else
293     \ifx#1t\string~%
294     \else\ifx#1c\string,%
295     \else\string#1%
296     \fi\fi
297   \expandafter\bbl@sh@string
298   \fi}
299 \ifx\bbl@opt@shorthands\@nnil
300   \def\bbl@ifshorthand#1#2#3{#2}%
301 \else\ifx\bbl@opt@shorthands\@empty
302   \def\bbl@ifshorthand#1#2#3{#3}%
303 \else
```

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

```
304 \def\bbl@ifshorthand#1{%
305   \bbl@xin@{\string#1}{\bbl@opt@shorthands}%
306   \ifin@
307     \expandafter\@firstoftwo
308   \else
309     \expandafter\@secondoftwo
310   \fi}
```

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

```
311 \edef\bbl@opt@shorthands{%
312   \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.

```
313 \bbl@ifshorthand{'}%
314   {\PassOptionsToPackage{activeacute}{babel}}{}
315 \bbl@ifshorthand{`}%
316   {\PassOptionsToPackage{activegrave}{babel}}{}
317 \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.

```
318 \ifx\bbl@opt@headfoot\@nnil\else
319   \g@addto@macro\@resetactivechars{%
320     \set@typeset@protect
321     \expandafter\select@language@x\expandafter{\bbl@opt@headfoot}%
322     \let\protect\noexpand}
323 \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.

```
324 \ifx\bbl@opt@safe\@undefined
325   \def\bbl@opt@safe{BR}
```

```

326 \fi
327 \ifx\bbbl@opt@main\@nnil\else
328   \edef\bbbl@language@opts{%
329     \ifx\bbbl@language@opts\@empty\else\bbbl@language@opts,\fi
330     \bbbl@opt@main}
331 \fi

```

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

```

332 \bbbl@trace{Defining IfBabelLayout}
333 \ifx\bbbl@opt@layout\@nnil
334   \newcommand\IfBabelLayout[3]{#3}%
335 \else
336   \newcommand\IfBabelLayout[1]{%
337     \@expandtwoargs\in@{.#1.}{.\bbbl@opt@layout.}%
338     \ifin@
339       \expandafter\@firstoftwo
340     \else
341       \expandafter\@secondoftwo
342     \fi}
343 \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).

```

344 \bbbl@trace{Language options}
345 \let\bbbl@afterlang\relax
346 \let\BabelModifiers\relax
347 \let\bbbl@loaded\@empty
348 \def\bbbl@load@language#1{%
349   \InputIfFileExists{#1.ldf}%
350   {\edef\bbbl@loaded{\CurrentOption
351     \ifx\bbbl@loaded\@empty\else,\bbbl@loaded\fi}%
352     \expandafter\let\expandafter\bbbl@afterlang
353       \csname\CurrentOption.ldf-h@@k\endcsname
354     \expandafter\let\expandafter\BabelModifiers
355       \csname bbl@mod@\CurrentOption\endcsname}%
356   {\bbbl@error{%
357     Unknown option '\CurrentOption'. Either you misspelled it\\%
358     or the language definition file \CurrentOption.ldf was not found}{%
359     Valid options are: shorthands=, KeepShorthandsActive,\\%
360     activeacute, activegrave, noconfigs, safe=, main=, math=\\%
361     headfoot=, strings=, config=, hyphenmap=, or a language name.}}}

```

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

```

362 \def\bbbl@try@load@lang#1#2#3{%
363   \IfFileExists{\CurrentOption.ldf}%
364   {\bbbl@load@language{\CurrentOption}}%
365   {#1\bbbl@load@language{#2}#3}}
366 \DeclareOption{afrikaans}{\bbbl@try@load@lang{}{dutch}}
367 \DeclareOption{brazil}{\bbbl@try@load@lang{}{portuges}}
368 \DeclareOption{brazilian}{\bbbl@try@load@lang{}{portuges}}
369 \DeclareOption{hebrew}{%
370   \input{rlbabel.def}%
371   \bbbl@load@language{hebrew}}
372 \DeclareOption{hungarian}{\bbbl@try@load@lang{}{magyar}}

```

```

373 \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}{}}
374 \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}{}}
375 \DeclareOption{polutonikogreek}{%
376   \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}}
377 \DeclareOption{portuguese}{\bbl@try@load@lang{}{portuges}{}}
378 \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}{}}
379 \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}{}}
380 \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.

```

381 \ifx\bbl@opt@config\@nnil
382   \@ifpackagewith{babel}{noconfigs}{}%
383   {\InputIfFileExists{bblopts.cfg}%
384     {\typeout{*****^J%
385               * Local config file bblopts.cfg used^^J%
386               *}}%
387     {}}%
388 \else
389   \InputIfFileExists{\bbl@opt@config.cfg}%
390   {\typeout{*****^J%
391             * Local config file \bbl@opt@config.cfg used^^J%
392             *}}%
393   {\bbl@error{%
394     Local config file '\bbl@opt@config.cfg' not found}%
395     Perhaps you misspelled it.}%
396 \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 name of the option and the file are the same.

```

397 \bbl@for\bbl@tempa\bbl@language@opts{%
398   \bbl@ifunset{ds@\bbl@tempa}%
399   {\edef\bbl@tempb{%
400     \noexpand\DeclareOption
401     {\bbl@tempa}%
402     {\noexpand\bbl@load@language{\bbl@tempa}}}%
403     \bbl@tempb}%
404   \@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.

```

405 \bbl@foreach\@classoptionslist{%
406   \bbl@ifunset{ds@#1}%
407   {\IfFileExists{#1.ldf}%
408     {\DeclareOption{#1}{\bbl@load@language{#1}}}%
409     {}}%
410   {}}

```

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

```

411 \ifx\bbl@opt@main\@nnil\else
412   \expandafter

```

```

413 \let\expandafter\bbl@loadmain\csname ds@\bbl@opt@main\endcsname
414 \DeclareOption{\bbl@opt@main}{\fi}
415 \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):

```

416 \def\AfterBabelLanguage#1{%
417   \bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang}{\fi}
418 \DeclareOption*{}
419 \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`.

```

420 \ifx\bbl@opt@main\@nnil
421   \edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
422   \let\bbl@tempc\@empty
423   \bbl@for\bbl@tempb\bbl@tempa{%
424     \bbl@xin@{\bbl@tempb,}{\bbl@loaded,}%
425     \ifin@{\edef\bbl@tempc{\bbl@tempb}\fi}
426   \def\bbl@tempa#1,#2\@nnil{\def\bbl@tempb{#1}}
427   \expandafter\bbl@tempa\bbl@loaded,\@nnil
428   \ifx\bbl@tempb\bbl@tempc\else
429     \bbl@warning{%
430       Last declared language option is '\bbl@tempc',\%
431       but the last processed one was '\bbl@tempb'.\%
432       The main language cannot be set as both a global\%
433       and a package option. Use 'main=\bbl@tempc' as\%
434       option. Reported}%
435   \fi
436 \else
437   \DeclareOption{\bbl@opt@main}{\bbl@loadmain}
438   \ExecuteOptions{\bbl@opt@main}
439   \DeclareOption*{}
440   \ProcessOptions*
441 \fi
442 \def\AfterBabelLanguage{%
443   \bbl@error
444   {Too late for \string\AfterBabelLanguage}%
445   {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.

```

446 \ifx\bbl@main@language\@undefined
447   \bbl@info{%
448     You haven't specified a language. I'll use 'nil'\%
449     as the main language. Reported}
450   \bbl@load@language{nil}
451 \fi
452 \end{package}
453 \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 L^AT_EX-specific stuff. Because plain T_EX users might want to use some of the features of the babel system too, care has to be taken that plain T_EX 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 T_EX and L^AT_EX, some of it is for the L^AT_EX 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

```
454 \ifx\ldf@quit\@undefined
455 \else
456   \expandafter\endinput
457 \fi
458 <<Make sure ProvidesFile is defined>>
459 \ProvidesFile{babel.def}[\<date>] <<version>> Babel common definitions]
460 <<Load macros for plain if not LaTeX>>
```

The file babel.def expects some definitions made in the L^AT_EX 2_ε style file. So, In L^AT_EX 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 in no package options, and therefore and alternative mechanism is provided. For the moment, only \babeloptionstrings and \babeloptionmath are provided, which can be defined before loading babel. \BabelModifiers can be set too (but not sure it works).

```
461 \ifx\bbl@ifshorthand\@undefined
462   \let\bbl@opt@shorthands\@nnil
463   \def\bbl@ifshorthand#1#2#3{#2}%
464   \let\bbl@language@opts\@empty
465   \ifx\babeloptionstrings\@undefined
466     \let\bbl@opt@strings\@nnil
467   \else
468     \let\bbl@opt@strings\babeloptionstrings
469   \fi
470   \def\BabelStringsDefault{generic}
471   \def\bbl@tempa{normal}
472   \ifx\babeloptionmath\bbl@tempa
473     \def\bbl@mathnormal{\noexpand\textormath}
474   \fi
475   \def\AfterBabelLanguage#1#2{}
476   \ifx\BabelModifiers\@undefined\let\BabelModifiers\relax\fi
477   \let\bbl@afterlang\relax
478   \def\bbl@opt@safe{BR}
479   \ifx\@uclclist\@undefined\let\@uclclist\@empty\fi
480   \ifx\bbl@trace\@undefined\def\bbl@trace#1{}\fi
481 \fi
```

And continue.

```
482 \ifx\bbl@switchflag\@undefined % Prevent double input
483   \let\bbl@switchflag\relax
```

```

484 \input switch.def\relax
485 \fi
486 \bbl@trace{Compatibility with language.def}
487 \ifx\bbl@languages\@undefined
488 \ifx\directlua\@undefined
489 \openin1 = language.def
490 \ifeof1
491 \closein1
492 \message{I couldn't find the file language.def}
493 \else
494 \closein1
495 \beginngroup
496 \def\addlanguage#1#2#3#4#5{%
497 \expandafter\ifx\csname lang@#1\endcsname\relax\else
498 \global\expandafter\let\csname l@#1\expandafter\endcsname
499 \csname lang@#1\endcsname
500 \fi}%
501 \def\uselanguage#1{%
502 \input language.def
503 \endgroup
504 \fi
505 \fi
506 \chardef\l@english\z@
507 \fi
508 <<Load patterns in luatex>>
509 <<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 T_EX-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 T_EX-code to be added. Finally the *<control sequence>* is redefined, using the contents of the token register.

```

510 \def\addto#1#2{%
511 \ifx#1\@undefined
512 \def#1{#2}%
513 \else
514 \ifx#1\relax
515 \def#1{#2}%
516 \else
517 {\toks@\expandafter{#1#2}%
518 \xdef#1{\the\toks@}}%
519 \fi
520 \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.

```

521 \def\bbl@withactive#1#2{%
522 \beginngroup
523 \lccode`~=#2\relax
524 \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@...`

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

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

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

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

```
530 \def\bbl@redefine@long#1{%
531   \edef\bbl@tempa{\bbl@stripslash#1}%
532   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
533   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
534 \@onlypreamble\bbl@redefine@long
```

`\bbl@redefineroobust` 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_`.

```
535 \def\bbl@redefineroobust#1{%
536   \edef\bbl@tempa{\bbl@stripslash#1}%
537   \bbl@ifunset{\bbl@tempa\space}%
538   {\expandafter\let\csname org@\bbl@tempa\endcsname#1%
539     \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
540   {\bbl@exp{\let\<org@\bbl@tempa>\<\bbl@tempa\space>}}}%
541   \@namedef{\bbl@tempa\space}}
```

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

```
542 \@onlypreamble\bbl@redefineroobust
```

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.

```
543 \bbl@trace{Hooks}
544 \def\AddBabelHook#1#2{%
545   \bbl@ifunset{bbl@hk@#1}{\EnableBabelHook{#1}}}%
546   \def\bbl@tempa##1,#2=##2,##3\@empty{\def\bbl@tempb{##2}}%
547   \expandafter\bbl@tempa\bbl@evargs,#2=,\@empty
548   \bbl@ifunset{bbl@ev@#1@#2}%
549     {\bbl@csarg\bbl@add{ev@#2}{\bbl@elt{#1}}}%
550     \bbl@csarg\newcommand}%
551     {\bbl@csarg\let{ev@#1@#2}\relax
552     \bbl@csarg\newcommand}%
553     {ev@#1@#2}[\bbl@tempb]}
554 \def\EnableBabelHook#1{\bbl@csarg\let{hk@#1}\@firstofone}
555 \def\DisableBabelHook#1{\bbl@csarg\let{hk@#1}\@gobble}
556 \def\bbl@usehooks#1#2{%
```

```

557 \def\bbl@elt##1{%
558   \@nameuse{bbl@hk@##1}\@nameuse{bbl@ev@##1@#1}#2}}%
559 \@nameuse{bbl@ev@#1}}

```

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).

```

560 \def\bbl@evargs{% <- don't delete this comma
561   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
562   adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
563   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
564   hyphenation=2,initiateactive=3,afterreset=0,foreign=0,foreign*=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.

```

565 \bbl@trace{Defining babelensure}
566 \newcommand\babelensure[2][{}]{% TODO - revise test files
567   \AddBabelHook{babel-ensure}{afterextras}{%
568     \ifcase\bbl@select@type
569       \@nameuse{bbl@e@\languageame}%
570     \fi}%
571   \begingroup
572     \let\bbl@ens@include\@empty
573     \let\bbl@ens@exclude\@empty
574     \def\bbl@ens@fontenc{\relax}%
575     \def\bbl@tempb##1{%
576       \ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}%
577     \edef\bbl@tempa{\bbl@tempb#1\@empty}%
578     \def\bbl@tempb##1=##2\@{\@namedef{bbl@ens@##1}{##2}}%
579     \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
580     \def\bbl@tempc{\bbl@ensure}%
581     \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
582       \expandafter{\bbl@ens@include}}%
583     \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
584       \expandafter{\bbl@ens@exclude}}%
585     \toks@\expandafter{\bbl@tempc}%
586     \bbl@exp{%
587   \endgroup
588   \def\<bbl@e@#2>{\the\toks@{\bbl@ens@fontenc}}}%
589 \def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
590   \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
591     \ifx##1\@empty\else
592       \in@{##1}{#2}%
593     \ifin\else
594       \bbl@ifunset{bbl@ensure@\languageame}%
595       {\bbl@exp{%
596         \\\DeclareRobustCommand\<bbl@ensure@\languageame>[1]{%
597           \\\foreignlanguage{\languageame}%
598           {\ifx\relax#3\else

```

```

599          \\\fontencoding{#3}\\\selectfont
600          \fi
601          #####1}}}%
602          {}%
603          \toks@\expandafter{##1}%
604          \edef##1{%
605              \bbl@csarg\noexpand{ensure@\language\name}%
606              {\the\toks@}}%
607          \fi
608          \expandafter\bbl@tempb
609          \fi}%
610          \expandafter\bbl@tempb\bbl@captionslist\today\@empty
611          \def\bbl@tempa##1{% elt for include list
612              \ifx##1\@empty\else
613                  \bbl@csarg\in@{ensure@\language\name\expandafter}\expandafter{##1}%
614                  \ifin@else
615                      \bbl@tempb##1\@empty
616                  \fi
617                  \expandafter\bbl@tempa
618                  \fi}%
619          \bbl@tempa#1\@empty}
620          \def\bbl@captionslist{%
621              \prefacename\refname\abstractname\bibname\chaptername\appendixname
622              \contentsname\listfigurename\listtablename\indexname\figurename
623              \tablename\partname\encname\ccname\headtoname\pagename\seename
624              \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`.

```

625 \bbl@trace{Macros for setting language files up}
626 \def\bbl@ldfinit{%
627     \let\bbl@screset\@empty
628     \let\BabelStrings\bbl@opt@string
629     \let\BabelOptions\@empty
630     \let\BabelLanguages\relax
631     \ifx\originalTeX\@undefined

```

```

632 \let\originalTeX\@empty
633 \else
634 \originalTeX
635 \fi}
636 \def\LdfInit#1#2{%
637 \chardef\atcatcode=\catcode`\@
638 \catcode`\@=11\relax
639 \chardef\eqcatcode=\catcode`\=
640 \catcode`\==12\relax
641 \expandafter\if\expandafter\@backslashchar
642 \expandafter\@car\string#2\@nil
643 \ifx#2\@undefined\else
644 \ldf@quit{#1}%
645 \fi
646 \else
647 \expandafter\ifx\csname#2\endcsname\relax\else
648 \ldf@quit{#1}%
649 \fi
650 \fi
651 \bbl@ldfinit}

```

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

```

652 \def\ldf@quit#1{%
653 \expandafter\main@language\expandafter{#1}%
654 \catcode`\@=\atcatcode \let\atcatcode\relax
655 \catcode`\==\eqcatcode \let\eqcatcode\relax
656 \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.

```

657 \def\bbl@afterldf#1{%
658 \bbl@afterlang
659 \let\bbl@afterlang\relax
660 \let\BabelModifiers\relax
661 \let\bbl@screset\relax}%
662 \def\ldf@finish#1{%
663 \loadlocalcfg{#1}%
664 \bbl@afterldf{#1}%
665 \expandafter\main@language\expandafter{#1}%
666 \catcode`\@=\atcatcode \let\atcatcode\relax
667 \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 .

```

668 \@onlypreamble\LdfInit
669 \@onlypreamble\ldf@quit
670 \@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.

```

671 \def\main@language#1{%
672 \def\bbl@main@language{#1}%
673 \let\languagename\bbl@main@language
674 \bbl@patterns{\languagename}}

```

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

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

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

```
678 \def\select@language@x#1{%
679   \ifcase\bbl@select@type
680     \bbl@ifsamestring\languagename{#1}{\select@language{#1}}%
681   \else
682     \select@language{#1}%
683   \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.

```
684 \bbl@trace{Shorhands}
685 \def\bbl@add@special#1{% 1:a macro like "\, \?, etc.
686   \bbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
687   \bbl@ifunset{@sanitize}{\bbl@add\@sanitize{\@makeother#1}}%
688   \ifx\nfss@catcodes\@undefined\else % TODO - same for above
689     \begingroup
690       \catcode`#1\active
691       \nfss@catcodes
692       \ifnum\catcode`#1=\active
693         \endgroup
694         \bbl@add\nfss@catcodes{\@makeother#1}%
695       \else
696         \endgroup
697       \fi
698   \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.

```
699 \def\bbl@remove@special#1{%
700   \begingroup
701     \def\x##1##2{\ifnum`#1=`##2\noexpand\@empty
702       \else\noexpand##1\noexpand##2\fi}%
703     \def\do{\x\do}%
704     \def\@makeother{\x\@makeother}%
705   \edef\x{\endgroup
706     \def\noexpand\dospecials{\dospecials}%
707     \expandafter\ifx\csname @sanitize\endcsname\relax\else
708       \def\noexpand\@sanitize{\@sanitize}%
709     \fi}%
710   \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).

```

711 \def\bbl@active@def#1#2#3#4{%
712   \@namedef{#3#1}{%
713     \expandafter\ifx\csname#2@sh@#1@\endcsname\relax
714       \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
715     \else
716       \bbl@afterfi\csname#2@sh@#1@\endcsname
717     \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.

```

718   \long\@namedef{#3@arg#1}##1{%
719     \expandafter\ifx\csname#2@sh@#1@\string##1@\endcsname\relax
720       \bbl@afterelse\csname#4#1\endcsname##1%
721     \else
722       \bbl@afterfi\csname#2@sh@#1@\string##1@\endcsname
723     \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.

```

724 \def\@initiate@active@char#1{%
725   \bbl@ifunset{active@char\string#1}%
726   {\bbl@withactive
727     {\expandafter\@initiate@active@char\expandafter}#1\string#1}%
728   {}}

```

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`).

```

729 \def\@initiate@active@char#1#2#3{%
730   \bbl@csarg\edef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%
731   \ifx#1\@undefined
732     \bbl@csarg\edef{oridef@#2}{\let\noexpand#1\noexpand\@undefined}%
733   \else
734     \bbl@csarg\let{oridef@#2}#1%
735     \bbl@csarg\edef{oridef@#2}{%
736       \let\noexpand#1%
737       \expandafter\noexpand\csname\bbl@oridef@@#2\endcsname}%
738   \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`).

```

739 \ifx#1#3\relax
740   \expandafter\let\csname normal@char#2\endcsname#3%
741 \else
742   \bbl@info{Making #2 an active character}%
743   \ifnum\mathcode`#2="8000
744     \@namedef{normal@char#2}{%
745       \textormath{#3}{\csname bbl@oridef@#2\endcsname}}%
746   \else
747     \@namedef{normal@char#2}{#3}%
748   \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).

```

749   \bbl@restoreactive{#2}%
750   \AtBeginDocument{%
751     \catcode`#2\active
752     \if@filesw
753       \immediate\write\@mainaux{\catcode`\string#2\active}%
754     \fi}%
755   \expandafter\bbl@add@special\csname#2\endcsname
756   \catcode`#2\active
757 \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⟩`).

```

758 \let\bbl@tempa\@firstoftwo
759 \if\string^#2%
760   \def\bbl@tempa{\noexpand\textormath}%
761 \else
762   \ifx\bbl@mathnormal\@undefined\else
763     \let\bbl@tempa\bbl@mathnormal
764   \fi
765 \fi
766 \expandafter\edef\csname active@char#2\endcsname{%
767   \bbl@tempa
768     {\noexpand\if@safe@actives
769       \noexpand\expandafter
770       \expandafter\noexpand\csname normal@char#2\endcsname
771     \noexpand\else
772       \noexpand\expandafter
773       \expandafter\noexpand\csname bbl@doactive#2\endcsname
774     \noexpand\fi}%
775   {\expandafter\noexpand\csname normal@char#2\endcsname}}%

```

```

776 \bbl@csarg\edef{doactive#2}{%
777   \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

$$\backslash\mathrm{active@prefix}\langle char\rangle\backslash\mathrm{normal@char}\langle char\rangle$$

(where $\backslash\mathrm{active@char}\langle char\rangle$ is *one* control sequence!).

```

778 \bbl@csarg\edef{active@#2}{%
779   \noexpand\active@prefix\noexpand#1%
780   \expandafter\noexpand\csname active@char#2\endcsname}%
781 \bbl@csarg\edef{normal@#2}{%
782   \noexpand\active@prefix\noexpand#1%
783   \expandafter\noexpand\csname normal@char#2\endcsname}%
784 \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.

```

785 \bbl@active@def#2\user@group{user@active}{language@active}%
786 \bbl@active@def#2\language@group{language@active}{system@active}%
787 \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 $\mathrm{T}_\mathrm{E}\mathrm{X}$ would see $\backslash\mathrm{protect}'\backslash\mathrm{protect}'$. To prevent this from happening a couple of shorthand needs to be defined at user level.

```

788 \expandafter\edef\csname\user@group @sh@#2@@\endcsname
789   {\expandafter\noexpand\csname normal@char#2\endcsname}%
790 \expandafter\edef\csname\user@group @sh@#2@\string\protect@\endcsname
791   {\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 $\backslash\mathrm{pr@m@s}$ 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.

```

792 \if\string'#2%
793   \let\prim@s\bbl@prim@s
794   \let\active@math@prime#1%
795 \fi
796 \bbl@usehooks{initiateactive}{\#1}{\#2}{\#3}}

```

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

```

797 <<(*More package options)>> \equiv
798 \DeclareOption{math=active}{}
799 \DeclareOption{math=normal}{{\def\bbl@mathnormal{\noexpand\textormath}}}
800 <</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*.

```

801 \@ifpackagewith{babel}{KeepShorthandsActive}%
802   {\let\bbl@restoreactive\@gobble}%
803   {\def\bbl@restoreactive#1{%
804     \bbl@exp{%

```



```

805      \\AfterBabelLanguage\\CurrentOption
806      {\catcode`#1=\the\catcode`#1\relax}%
807      \\AtEndOfPackage
808      {\catcode`#1=\the\catcode`#1\relax}}}%
809      \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.

```

810 \def\bbl@sh@select#1#2{%
811   \expandafter\ifx\csname#1@sh#2@sel\endcsname\relax
812     \bbl@afterelse\bbl@scndcs
813   \else
814     \bbl@afterfi\csname#1@sh#2@sel\endcsname
815   \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`.

```

816 \def\active@prefix#1{%
817   \ifx\protect\@typeset@protect
818     \else

```

When `\protect` is set to `\@unexpandable@protect` we make sure that the active character is also *not* expanded by inserting `\noexpand` in front of it. The `\@gobble` is needed to remove a token such as `\activechar:` (when the double colon was the active character to be dealt with).

```

819     \ifx\protect\@unexpandable@protect
820       \noexpand#1%
821     \else
822       \protect#1%
823     \fi
824     \expandafter\@gobble
825   \fi}

```

`\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⟩`.

```

826 \newif\if@safe@actives
827 \@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.

```

828 \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`.

```

829 \def\bbl@activate#1{%
830   \bbl@withactive{\expandafter\let\expandafter}#1%
831   \csname bbl@active@\string#1\endcsname}
832 \def\bbl@deactivate#1{%

```

```

833 \bbl@withactive{\expandafter\let\expandafter}#1%
834 \csname bbl@normal@\string#1\endcsname}

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

835 \def\bbl@firstcs#1#2{\csname#1\endcsname}
836 \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.

837 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil}
838 \def\@decl@short#1#2#3\@nil#4{%
839 \def\bbl@tempa{#3}%
840 \ifx\bbl@tempa\@empty
841 \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs
842 \bbl@ifunset{#1@sh@\string#2@}{}%
843 {\def\bbl@tempa{#4}%
844 \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa
845 \else
846 \bbl@info
847 {Redefining #1 shorthand \string#2\\%
848 in language \CurrentOption}%
849 \fi}%
850 \@namedef{#1@sh@\string#2@}{#4}%
851 \else
852 \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs
853 \bbl@ifunset{#1@sh@\string#2@\string#3@}{}%
854 {\def\bbl@tempa{#4}%
855 \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
856 \else
857 \bbl@info
858 {Redefining #1 shorthand \string#2\string#3\\%
859 in language \CurrentOption}%
860 \fi}%
861 \@namedef{#1@sh@\string#2@\string#3@}{#4}%
862 \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.

863 \def\textormath{%
864 \ifmmode
865 \expandafter\@secondoftwo
866 \else
867 \expandafter\@firstoftwo
868 \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’.

869 \def\user@group{user}
870 \def\language@group{english}
871 \def\system@group{system}

```

`\useshortands` 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.

```

872 \def\useshortands{%
873   \@ifstar\bb1@usesesh@s{\bb1@usesesh@x{}}
874 \def\bb1@usesesh@s#1{%
875   \bb1@usesesh@x
876   {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bb1@activate{#1}}}%
877   {#1}}
878 \def\bb1@usesesh@x#1#2{%
879   \bb1@ifshorthand{#2}%
880   {\def\user@group{user}%
881     \initiate@active@char{#2}%
882     #1%
883     \bb1@activate{#2}}%
884   {\bb1@error
885     {Cannot declare a shorthand turned off (\string#2)}
886     {Sorry, but you cannot use shorthands which have been\%
887       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 `\bb1@set@user@generic`); we make also sure `{}` and `\protect` are taken into account in this new top level.

```

888 \def\user@language@group{user@\language@group}
889 \def\bb1@set@user@generic#1#2{%
890   \bb1@ifunset{user@generic@active#1}%
891   {\bb1@active@def#1\user@language@group{user@active}{user@generic@active}%
892     \bb1@active@def#1\user@group{user@generic@active}{language@active}%
893     \expandafter\edef\csname#2@sh@#1@@\endcsname{%
894       \expandafter\noexpand\csname normal@char#1\endcsname}%
895     \expandafter\edef\csname#2@sh@#1@\string\protect\endcsname{%
896       \expandafter\noexpand\csname user@active#1\endcsname}}%
897   \@empty}
898 \newcommand\defineshorthand[3][user]{%
899   \edef\bb1@tempa{\zap@space#1 \@empty}%
900   \bb1@for\bb1@tempb\bb1@tempa{%
901     \if*\expandafter\car\bb1@tempb\@nil
902       \edef\bb1@tempb{user\expandafter@gobble\bb1@tempb}%
903       \expandafter\twoargs
904       \bb1@set@user@generic{\expandafter\string\car#2\@nil}\bb1@tempb
905     \fi
906     \declare@shorthand{\bb1@tempb}{#2}{#3}}}

```

`\languageshortands` 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.

```

907 \def\languageshortands#1{\def\language@group{#1}}

```

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

```

908 \def\aliasshorthand#1#2{%
909   \bb1@ifshorthand{#2}%
910   {\expandafter\ifx\csname active@char\string#2\endcsname\relax
911     \ifx\document\@notprerr

```

```

912      \@notshorthand{#2}%
913      \else
914      \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`.

```

915      \expandafter\let\csname active@char\string#2\expandafter\endcsname
916      \csname active@char\string#1\endcsname
917      \expandafter\let\csname normal@char\string#2\expandafter\endcsname
918      \csname normal@char\string#1\endcsname
919      \bbl@activate{#2}%
920      \fi
921    \fi}%
922    {\bbl@error
923      {Cannot declare a shorthand turned off (\string#2)}
924      {Sorry, but you cannot use shorthands which have been\\%
925        turned off in the package options}}}

```

`\@notshorthand`

```

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

```

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

`\shorthandoff`

```

933 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
934 \DeclareRobustCommand*\shorthandoff{%
935   \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
936 \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.

```

937 \def\bbl@switch@sh#1#2{%
938   \ifx#2\@nnil\else
939     \bbl@ifunset{\bbl@active@\string#2}%
940     {\bbl@error
941       {I cannot switch '\string#2' on or off--not a shorthand}%
942       {This character is not a shorthand. Maybe you made\\%
943         a typing mistake? I will ignore your instruction}}%
944     {\ifcase#1%
945       \catcode`#2\relax
946       \or
947       \catcode`#2\active
948       \or
949       \csname bbl@oricat@\string#2\endcsname

```

```

950      \csname bbl@oridef@\string#2\endcsname
951      \fi}%
952      \bbl@afterfi\bbl@switch@sh#1%
953      \fi}

```

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

```

954 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
955 \def\bbl@putsh#1{%
956   \bbl@ifunset{\bbl@active@\string#1}%
957   {\bbl@putsh@i#1\@empty\@nnil}%
958   {\csname bbl@active@\string#1\endcsname}}
959 \def\bbl@putsh@i#1#2\@nnil{%
960   \csname\language\sh@\string#1@%
961     \ifx\@empty#2\else\string#2@\fi\endcsname}
962 \ifx\bbl@opt@shorthands\@nnil\else
963   \let\bbl@s@initiate@active@char\initiate@active@char
964   \def\initiate@active@char#1{%
965     \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
966   \let\bbl@s@switch@sh\bbl@switch@sh
967   \def\bbl@switch@sh#1#2{%
968     \ifx#2\@nnil\else
969       \bbl@afterfi
970       \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
971       \fi}
972   \let\bbl@s@activate\bbl@activate
973   \def\bbl@activate#1{%
974     \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
975   \let\bbl@s@deactivate\bbl@deactivate
976   \def\bbl@deactivate#1{%
977     \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
978   \fi

```

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

```

979 \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.

```

980 \def\bbl@prim@s{%
981   \prime\futurelet\@let@token\bbl@pr@m@s}
982 \def\bbl@if@primes#1#2{%
983   \ifx#1\@let@token
984     \expandafter\@firstoftwo
985   \else\ifx#2\@let@token
986     \bbl@afterelse\expandafter\@firstoftwo
987   \else
988     \bbl@afterfi\expandafter\@secondoftwo
989   \fi\fi}
990 \begingroup
991   \catcode`\^=7 \catcode`\*=\active \lccode`\*=\^
992   \catcode`\'=12 \catcode`\"=\active \lccode`\"=\'
993   \lowercase{%
994     \gdef\bbl@pr@m@s{%
995       \bbl@if@primes" '%
996       \pr@@@s

```

```

997      {\bbl@if@primes*^{\pr@@@t\egroup}}
998 \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).

```

999 \initiate@active@char{~}
1000 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
1001 \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.

```

1002 \expandafter\def\csname OT1dqpos\endcsname{127}
1003 \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

```

1004 \ifx\f@encoding\@undefined
1005   \def\f@encoding{OT1}
1006 \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.

```

1007 \bbl@trace{Language attributes}
1008 \newcommand\languageattribute[2]{%
1009   \def\bbl@tempc{#1}%
1010   \bbl@fixname\bbl@tempc
1011   \bbl@iflanguage\bbl@tempc{%
1012     \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.

```

1013     \ifx\bbl@known@attribs\@undefined
1014       \in@false
1015     \else

```

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

```

1016       \bbl@xin@{\bbl@tempc-##1,}{\bbl@known@attribs,}%
1017     \fi

```

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

```

1018     \ifin@
1019       \bbl@warning{%
1020         You have more than once selected the attribute '##1'\%
1021         for language #1. Reported}%
1022     \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.

```

1023      \bbl@exp{%
1024        \\bbl@add@list\\bbl@known@attribs{\bbl@tempc-##1}}%
1025      \edef\bbl@tempa{\bbl@tempc-##1}%
1026      \expandafter\bbl@ifknown@ttrib\expandafter{\bbl@tempa}\bbl@attributes%
1027      {\csname\bbl@tempc @attr@##1\endcsname}%
1028      {\@attrerr{\bbl@tempc}{##1}}%
1029      \fi}}

```

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

```
1030 \@onlypreamble\languageattribute
```

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

```

1031 \newcommand*{\@attrerr}[2]{%
1032   \bbl@error
1033   {The attribute #2 is unknown for language #1.}%
1034   {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}`.

```

1035 \def\bbl@declare@ttribute#1#2#3{%
1036   \bbl@xin@{,#2,}{,\BabelModifiers,}%
1037   \ifin@
1038     \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
1039   \fi
1040   \bbl@add@list\bbl@attributes{#1-#2}%
1041   \expandafter\def\csname#1@attr@#2\endcsname{#3}}

```

`\bbl@ifattributeset` This internal macro has 4 arguments. It can be used to interpret \TeX code based on whether a certain attribute was set. This command should appear inside the argument to `\AtBeginDocument` because the attributes are set in the document preamble, *after* babel is loaded.

The first argument is the language, the second argument the attribute being checked, and the third and fourth arguments are the true and false clauses.

```
1042 \def\bbl@ifattributeset#1#2#3#4{%
```

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

```

1043   \ifx\bbl@known@attribs\@undefined
1044     \in@false
1045     \else

```

The we need to check the list of known attributes.

```

1046     \bbl@xin@{,#1-#2,}{,\bbl@known@attribs,}%
1047     \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`'.

```

1048   \ifin@
1049     \bbl@afterelse#3%
1050   \else
1051     \bbl@afterfi#4%
1052   \fi
1053 }

```

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

```
1054 \def\bbl@ifknown@ttrib#1#2{%
```

We first assume the attribute is unknown.

```
1055 \let\bbl@tempa\@secondoftwo
```

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

```
1056 \bbl@loopx\bbl@tempb{#2}{%
```

```
1057 \expandafter\in@\expandafter{\expandafter,\bbl@tempb,}{, #1,}%
```

```
1058 \ifin@
```

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

```
1059 \let\bbl@tempa\@firstoftwo
```

```
1060 \else
```

```
1061 \fi}%
```

Finally we execute `\bbl@tempa`.

```
1062 \bbl@tempa
```

```
1063 }
```

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

```
1064 \def\bbl@clear@ttribs{%
```

```
1065 \ifx\bbl@attributes\undefined\else
```

```
1066 \bbl@loopx\bbl@tempa{\bbl@attributes}{%
```

```
1067 \expandafter\bbl@clear@ttrib\bbl@tempa.
```

```
1068 }%
```

```
1069 \let\bbl@attributes\undefined
```

```
1070 \fi}
```

```
1071 \def\bbl@clear@ttrib#1-#2.{%
```

```
1072 \expandafter\let\csname#1@attr#2\endcsname\undefined}
```

```
1073 \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 1074 \bbl@trace{Macros for saving definitions}
```

```
1075 \def\babel@beginsave{\babel@savecnt\z@}
```

Before it's forgotten, allocate the counter and initialize all.

```
1076 \newcount\babel@savecnt
```

```
1077 \babel@beginsave
```

`\babel@save` The macro `\babel@save<csname>` saves the current meaning of the control sequence `<csname>` to `\originalTeX`³¹. To do this, we let the current meaning to a temporary control

³¹`\originalTeX` has to be expandable, i.e. you shouldn't let it to `\relax`.

sequence, the restore commands are appended to `\originalTeX` and the counter is incremented.

```
1078 \def\babel@save#1{%
1079   \expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax
1080   \toks@\expandafter{\originalTeX\let#1=}%
1081   \bbl@exp{%
1082     \def\\originalTeX{\the\toks@<\babel@number\babel@savecnt>\relax}}%
1083   \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.

```
1084 \def\babel@savevariable#1{%
1085   \toks@\expandafter{\originalTeX #1=}%
1086   \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.

```
1087 \def\bbl@frenchspacing{%
1088   \ifnum\the\sfcodes\@m
1089     \let\bbl@nonfrenchspacing\relax
1090   \else
1091     \frenchspacing
1092     \let\bbl@nonfrenchspacing\nonfrenchspacing
1093   \fi}
1094 \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.

```
1095 \bbl@trace{Short tags}
1096 \def\babeltags#1{%
1097   \edef\bbl@tempa{\zap@space#1 \@empty}%
1098   \def\bbl@tempb##1=##2\@{#}%
1099   \edef\bbl@tempc{%
1100     \noexpand\newcommand
1101     \expandafter\noexpand\csname ##1\endcsname{%
1102       \noexpand\protect
1103       \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
1104     \noexpand\newcommand
1105     \expandafter\noexpand\csname text##1\endcsname{%
1106       \noexpand\foreignlanguage{##2}}
1107   \bbl@tempc}%
1108   \bbl@for\bbl@tempa\bbl@tempa{%
1109     \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.

```
1110 \bbl@trace{Hyphens}
```

```

1111 \@onlypreamble\babelhyphenation
1112 \AtEndOfPackage{%
1113   \newcommand\babelhyphenation[2][\@empty]{%
1114     \ifx\bbbl@hyphenation@relax
1115       \let\bbbl@hyphenation@\@empty
1116     \fi
1117     \ifx\bbbl@hyphlist@\@empty\else
1118       \bbbl@warning{%
1119         You must not intermingle \string\selectlanguage\space and\\%
1120         \string\babelhyphenation\space or some exceptions will not\\%
1121         be taken into account. Reported}%
1122     \fi
1123     \ifx\@empty#1%
1124       \protected@edef\bbbl@hyphenation@{\bbbl@hyphenation@\space#2}%
1125     \else
1126       \bbbl@vforeach{#1}{%
1127         \def\bbbl@tempa{##1}%
1128         \bbbl@fixname\bbbl@tempa
1129         \bbbl@iflanguage\bbbl@tempa{%
1130           \bbbl@csarg\protected@edef{hyphenation@\bbbl@tempa}{%
1131             \bbbl@ifunset{bbbl@hyphenation@\bbbl@tempa}%
1132             \@empty
1133             {\csname bbl@hyphenation@\bbbl@tempa\endcsname\space}%
1134             #2}}}%
1135       \fi}}

```

`\bbbl@allowhyphens` This macro makes hyphenation possible. Basically its definition is nothing more than `\nobreak \hskip 0pt plus 0pt`³².

```

1136 \def\bbbl@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}
1137 \def\bbbl@t@one{T1}
1138 \def\allowhyphens{\ifx\cf@encoding\bbbl@t@one\else\bbbl@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`.

```

1139 \newcommand\babelnullhyphen{\char\hyphenchar\font}
1140 \def\babelhyphen{\active@prefix\babelhyphen\bbbl@hyphen}
1141 \def\bbbl@hyphen{%
1142   \@ifstar{\bbbl@hyphen@i @}{\bbbl@hyphen@i \@empty}}
1143 \def\bbbl@hyphen@i#1#2{%
1144   \bbbl@ifunset{bbbl@hy#1#2\@empty}%
1145   {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{-}{#2}}}%
1146   {\csname bbl@hy#1#2\@empty\endcsname}}

```

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

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.

```

1147 \def\bbbl@usehyphen#1{%
1148   \leavevmode
1149   \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
1150   \nobreak\hskip\z@skip}
1151 \def\bbbl@@usehyphen#1{%
1152   \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

³² $\mathrm{T}_{\mathrm{E}}\mathrm{X}$ begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.

The following macro inserts the hyphen char.

```
1153 \def\bbl@hyphenchar{%
1154   \ifnum\hyphenchar\font=\m@ne
1155     \babeinullhyphen
1156   \else
1157     \char\hyphenchar\font
1158   \fi}
```

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

```
1159 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1160 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1161 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
1162 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
1163 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}
1164 \def\bbl@hy@nobreak{\mbox{\bbl@hyphenchar}}
1165 \def\bbl@hy@repeat{%
1166   \bbl@usehyphen{%
1167     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
1168 \def\bbl@hy@repeat{%
1169   \bbl@usehyphen{%
1170     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
1171 \def\bbl@hy@empty{\hskip\z@skip}
1172 \def\bbl@hy@@empty{\discretionary{}{}{}}
```

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

```
1173 \def\bbl@disc#1#2{\nobreak\discretionary{#2-}{#1}\bbl@allowhyphens}
```

9.9 Interword spacing in Southeast Asian scripts

At work.

```
1174 \def\bbl@seascripts{,Thai,Lao,Khmer,} % Malayalam?
1175 \newcommand\babelinterword[3]{%
1176   \def\BabelInterword{#1}%
1177   \def\BabelInterwordStretch{#2}%
1178   \def\BabelInterwordShrink{#3}}
1179 \babelinterword{0}{.1}{0}
```

9.10 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.

```
1180 \bbl@trace{Multiencoding strings}
1181 \def\bbl@tglobal#1{\global\let#1#1}
1182 \def\bbl@recatcode#1{%
1183   \@tempcnta="7F
1184   \def\bbl@tempa{%
1185     \ifnum\@tempcnta>"FF\else
1186       \catcode\@tempcnta=#1\relax
1187       \advance\@tempcnta\@ne
```

```

1188     \expandafter\bb1@tempa
1189     \fi}%
1190 \bb1@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 `\bb1@uclc`. The parser is restarted inside `\(lang)\bb1@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\bb1@tolower\@empty\bb1@toupper\@empty
```

and starts over (and similarly when lowercasing).

```

1191 \@ifpackagewith{babel}{nocase}%
1192   {\let\bb1@patchuclc\relax}%
1193   {\def\bb1@patchuclc{%
1194     \global\let\bb1@patchuclc\relax
1195     \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bb1@uclc}}%
1196     \gdef\bb1@uclc##1{%
1197       \let\bb1@encoded\bb1@encoded@uclc
1198       \bb1@ifunset{\language @bb1@uclc}% and resumes it
1199       {##1}%
1200       {\let\bb1@tempa##1\relax % Used by LANG@bb1@uclc
1201         \csname\language @bb1@uclc\endcsname}%
1202       {\bb1@tolower\@empty}{\bb1@toupper\@empty}}%
1203     \gdef\bb1@tolower{\csname\language @bb1@lc\endcsname}%
1204     \gdef\bb1@toupper{\csname\language @bb1@uc\endcsname}}}
1205 <<(*More package options)>> ≡
1206 \DeclareOption{nocase}{}
1207 <</More package options>>

```

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

```

1208 <<(*More package options)>> ≡
1209 \let\bb1@opt@strings\@nnil % accept strings=value
1210 \DeclareOption{strings}{\def\bb1@opt@strings{\BabelStringsDefault}}
1211 \DeclareOption{strings=encoded}{\let\bb1@opt@strings\relax}
1212 \def\BabelStringsDefault{generic}
1213 <</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.

```

1214 \@onlypreamble\StartBabelCommands
1215 \def\StartBabelCommands{%
1216   \begingroup
1217   \bb1@recatcode{11}%
1218   <<Macros local to BabelCommands>>
1219   \def\bb1@provstring##1##2{%
1220     \providecommand##1{##2}%
1221     \bb1@tglobal##1}%
1222   \global\let\bb1@scafter\@empty
1223   \let\StartBabelCommands\bb1@startcmds
1224   \ifx\BabelLanguages\relax
1225     \let\BabelLanguages\CurrentOption

```

```

1226 \fi
1227 \begingroup
1228 \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
1229 \StartBabelCommands}
1230 \def\bbl@startcmds{%
1231 \ifx\bbl@screset\@nnil\else
1232 \bbl@usehooks{stopcommands}{}%
1233 \fi
1234 \endgroup
1235 \begingroup
1236 \@ifstar
1237 {\ifx\bbl@opt@strings\@nnil
1238 \let\bbl@opt@strings\BabelStringsDefault
1239 \fi
1240 \bbl@startcmds@i}%
1241 \bbl@startcmds@i}
1242 \def\bbl@startcmds@i#1#2{%
1243 \edef\bbl@L{\zap@space#1 \@empty}%
1244 \edef\bbl@G{\zap@space#2 \@empty}%
1245 \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.

```

1246 \newcommand\bbl@startcmds@ii[1][\@empty]{%
1247 \let\SetString@gobbletwo
1248 \let\bbl@stringdef@gobbletwo
1249 \let\AfterBabelCommands@gobble
1250 \ifx\@empty#1%
1251 \def\bbl@sc@label{generic}%
1252 \def\bbl@encstring##1##2{%
1253 \ProvideTextCommandDefault##1{##2}%
1254 \bbl@toggle##1%
1255 \expandafter\bbl@toggle\csname\string?\string##1\endcsname}%
1256 \let\bbl@sctest\in@true
1257 \else
1258 \let\bbl@sc@charset\space % <- zapped below
1259 \let\bbl@sc@fontenc\space % <- " "
1260 \def\bbl@tempa##1=##2\@nil{%
1261 \bbl@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
1262 \bbl@vforeach{label=#1}{\bbl@tempa##1\@nil}%
1263 \def\bbl@tempa##1 ##2{% space -> comma
1264 ##1%
1265 \ifx\@empty##2\else\ifx,##1,\else,\fi\bbl@afterfi\bbl@tempa##2\fi}%
1266 \edef\bbl@sc@fontenc{\expandafter\bbl@tempa\bbl@sc@fontenc\@empty}%
1267 \edef\bbl@sc@label{\expandafter\zap@space\bbl@sc@label\@empty}%
1268 \edef\bbl@sc@charset{\expandafter\zap@space\bbl@sc@charset\@empty}%
1269 \def\bbl@encstring##1##2{%
1270 \bbl@foreach\bbl@sc@fontenc{%
1271 \bbl@ifunset{T####1}%
1272 }%

```

```

1273         {\ProvideTextCommand###1{####1}{##2}%
1274         \bbl@toglobal##1%
1275         \expandafter
1276         \bbl@toglobal\csname####1\string##1\endcsname}}}%
1277     \def\bbl@sctest{%
1278         \bbl@xin@{\bbl@opt@strings,}{\bbl@sc@label,\bbl@sc@fontenc,}}%
1279     \fi
1280     \ifx\bbl@opt@strings\@nnil      % ie, no strings key -> defaults
1281     \else\ifx\bbl@opt@strings\relax  % ie, strings=encoded
1282         \let\AfterBabelCommands\bbl@aftercmds
1283         \let\SetString\bbl@setstring
1284         \let\bbl@stringdef\bbl@encstring
1285     \else      % ie, strings=value
1286     \bbl@sctest
1287     \fin@
1288         \let\AfterBabelCommands\bbl@aftercmds
1289         \let\SetString\bbl@setstring
1290         \let\bbl@stringdef\bbl@provstring
1291     \fi\fi\fi
1292     \bbl@scswitch
1293     \ifx\bbl@G\@empty
1294         \def\SetString##1##2{%
1295             \bbl@error{Missing group for string \string##1}%
1296             {You must assign strings to some category, typically\\%
1297             captions or extras, but you set none}}%
1298     \fi
1299     \ifx\@empty#1%
1300         \bbl@usehooks{defaultcommands}}}%
1301     \else
1302         \@expandtwoargs
1303         \bbl@usehooks{encodedcommands}{\bbl@sc@charset}{\bbl@sc@fontenc}}%
1304     \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).

```

1305 \def\bbl@forlang#1#2{%
1306     \bbl@for#1\bbl@L{%
1307         \bbl@xin@{, #1,}{\BabelLanguages,}%
1308         \ifin@#2\relax\fi}}
1309 \def\bbl@scswitch{%
1310     \bbl@forlang\bbl@tempa{%
1311         \ifx\bbl@G\@empty\else
1312             \ifx\SetString\@gobbletwo\else
1313                 \edef\bbl@GL{\bbl@G\bbl@tempa}%
1314                 \bbl@xin@{\bbl@GL,}{\bbl@screset,}%
1315             \fin@
1316             \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
1317             \xdef\bbl@screset{\bbl@screset,\bbl@GL}%
1318         \fi
1319     \fi
1320     \fi}}
1321 \AtEndOfPackage{%

```

```

1322 \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{\#2}}}%
1323 \let\bbl@scswitch\relax}
1324 \@onlypreamble\EndBabelCommands
1325 \def\EndBabelCommands{%
1326 \bbl@usehooks{stopcommands}{}%
1327 \endgroup
1328 \endgroup
1329 \bbl@scafter}

```

Now we define commands to be used inside \StartBabelCommands.

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

```

1330 \def\bbl@setstring#1#2{%
1331 \bbl@forlang\bbl@tempa{%
1332 \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
1333 \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
1334 {\global\expandafter % TODO - con \bbl@exp ?
1335 \bbl@add\csname\bbl@G\bbl@tempa\expandafter\endcsname\expandafter
1336 {\expandafter\bbl@scset\expandafter#1\csname\bbl@LC\endcsname}}}%
1337 }%
1338 \def\BabelString{\#2}%
1339 \bbl@usehooks{stringprocess}{}%
1340 \expandafter\bbl@stringdef
1341 \csname\bbl@LC\expandafter\endcsname\expandafter{\BabelString}}

```

Now, some additional stuff to be used when encoded strings are used. Captions then include \bbl@encoded for string to be expanded in case transformations. It is \relax by default, but in \MakeUppercase and \MakeLowercase its value is a modified expandable \@changed@cmd.

```

1342 \ifx\bbl@opt@strings\relax
1343 \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
1344 \bbl@patchuclc
1345 \let\bbl@encoded\relax
1346 \def\bbl@encoded@uclc#1{%
1347 \@inmathwarn#1%
1348 \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
1349 \expandafter\ifx\csname ?\string#1\endcsname\relax
1350 \TextSymbolUnavailable#1%
1351 \else
1352 \csname ?\string#1\endcsname
1353 \fi
1354 \else
1355 \csname\cf@encoding\string#1\endcsname
1356 \fi}
1357 \else
1358 \def\bbl@scset#1#2{\def#1{\#2}}
1359 \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.

```

1360 <<*Macros local to BabelCommands>> ≡
1361 \def\SetStringLoop##1##2{%

```

```

1362 \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
1363 \count@ \z@
1364 \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
1365 \advance\count@\@ne
1366 \toks@\expandafter{\bbl@tempa}%
1367 \bbl@exp{%
1368 \\\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
1369 \count@=\the\count@\relax}}}%
1370 <</Macros local to BabelCommands>>

```

Delaying code Now the definition of `\AfterBabelCommands` when it is activated.

```

1371 \def\bbl@aftercmds#1{%
1372 \toks@\expandafter{\bbl@scafter#1}%
1373 \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.

```

1374 <<*Macros local to BabelCommands>> ≡
1375 \newcommand\SetCase[3][]{%
1376 \bbl@patchuclc
1377 \bbl@forlang\bbl@tempa{%
1378 \expandafter\bbl@encstring
1379 \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
1380 \expandafter\bbl@encstring
1381 \csname\bbl@tempa @bbl@uc\endcsname{##2}%
1382 \expandafter\bbl@encstring
1383 \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
1384 <</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.

```

1385 <<*Macros local to BabelCommands>> ≡
1386 \newcommand\SetHyphenMap[1]{%
1387 \bbl@forlang\bbl@tempa{%
1388 \expandafter\bbl@stringdef
1389 \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}
1390 <</Macros local to BabelCommands>>

```

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

```

1391 \newcommand\BabelLower[2]{% one to one.
1392 \ifnum\lccode#1=#2\else
1393 \babel@savevariable{\lccode#1}%
1394 \lccode#1=#2\relax
1395 \fi}
1396 \newcommand\BabelLowerMM[4]{% many-to-many
1397 \@tempcnta=#1\relax
1398 \@tempcntb=#4\relax
1399 \def\bbl@tempa{%
1400 \ifnum\@tempcnta>#2\else
1401 \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
1402 \advance\@tempcnta#3\relax
1403 \advance\@tempcntb#3\relax
1404 \expandafter\bbl@tempa
1405 \fi}%
1406 \bbl@tempa}

```



```

1407 \newcommand\BabelLowerMO[4]{% many-to-one
1408   \@tempcnta=#1\relax
1409   \def\bbl@tempa{%
1410     \ifnum\@tempcnta>#2\else
1411       \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
1412       \advance\@tempcnta#3
1413       \expandafter\bbl@tempa
1414     \fi}%
1415   \bbl@tempa}

```

The following package options control the behavior of hyphenation mapping.

```

1416 <<(*More package options)> \equiv
1417 \DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
1418 \DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\@ne}
1419 \DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
1420 \DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@@}
1421 \DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}
1422 <</More package options>>

```

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

```

1423 \AtEndOfPackage{%
1424   \ifx\bbl@opt@hyphenmap\undefined
1425     \bbl@xin@{,}{\bbl@language@opts}%
1426     \chardef\bbl@opt@hyphenmap\ifin4\else\@ne\fi
1427   \fi}

```

9.11 Macros common to a number of languages

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

```

1428 \bbl@trace{Macros related to glyphs}
1429 \def\set@low@box#1{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}%
1430   \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@%
1431   \setbox\z@\hbox{\lower\dimen\z@ \box\z@}\ht\z@\ht\tw@ \dp\z@\dp\tw@}

```

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

```

1432 \def\save@sfc@q#1{\leavevmode
1433   \begingroup
1434     \edef\@SF{\spacefactor\the\spacefactor}\@SF
1435   \endgroup}

```

9.12 Making glyphs available

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

9.12.1 Quotation marks

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

```

1436 \ProvideTextCommand{\quotedblbase}{OT1}{%
1437   \save@sfc@q{\set@low@box{\textquotedblright\}%
1438     \box\z@\kern-.04em\bbl@allowhyphens}}

```

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

```
1439 \ProvideTextCommandDefault{\quotedblbase}{%
1440   \UseTextSymbol{OT1}{\quotedblbase}}
```

`\quotesinglbase` We also need the single quote character at the baseline.

```
1441 \ProvideTextCommand{\quotesinglbase}{OT1}{%
1442   \save@sf@q{\set@low@box{\textquoteright\}%
1443     \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.

```
1444 \ProvideTextCommandDefault{\quotesinglbase}{%
1445   \UseTextSymbol{OT1}{\quotesinglbase}}
```

`\guillemotleft` The guillemet characters are not available in OT1 encoding. They are faked.

```
\guillemotright 1446 \ProvideTextCommand{\guillemotleft}{OT1}{%
1447   \ifmmode
1448     \ll
1449   \else
1450     \save@sf@q{\nobreak
1451       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bb1@allowhyphens}%
1452   \fi}
1453 \ProvideTextCommand{\guillemotright}{OT1}{%
1454   \ifmmode
1455     \gg
1456   \else
1457     \save@sf@q{\nobreak
1458       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bb1@allowhyphens}%
1459   \fi}
```

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

```
1460 \ProvideTextCommandDefault{\guillemotleft}{%
1461   \UseTextSymbol{OT1}{\guillemotleft}}
1462 \ProvideTextCommandDefault{\guillemotright}{%
1463   \UseTextSymbol{OT1}{\guillemotright}}
```

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

```
\guilsinglright 1464 \ProvideTextCommand{\guilsinglleft}{OT1}{%
1465   \ifmmode
1466     <%
1467   \else
1468     \save@sf@q{\nobreak
1469       \raise.2ex\hbox{$\scriptscriptstyle<$}\bb1@allowhyphens}%
1470   \fi}
1471 \ProvideTextCommand{\guilsinglright}{OT1}{%
1472   \ifmmode
1473     >%
1474   \else
1475     \save@sf@q{\nobreak
1476       \raise.2ex\hbox{$\scriptscriptstyle>$}\bb1@allowhyphens}%
1477   \fi}
```

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

```
1478 \ProvideTextCommandDefault{\guilsinglleft}{%
1479   \UseTextSymbol{OT1}{\guilsinglleft}}
```

```

1479 \UseTextSymbol{OT1}{\guilsinglleft}}
1480 \ProvideTextCommandDefault{\guilsinglright}{%
1481 \UseTextSymbol{OT1}{\guilsinglright}}

```

9.12.2 Letters

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

```

1482 \DeclareTextCommand{\ij}{OT1}{%
1483 i\kern-0.02em\bbl@allowhyphens j}
1484 \DeclareTextCommand{\IJ}{OT1}{%
1485 I\kern-0.02em\bbl@allowhyphens J}
1486 \DeclareTextCommand{\ij}{T1}{\char188}
1487 \DeclareTextCommand{\IJ}{T1}{\char156}

```

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

```

1488 \ProvideTextCommandDefault{\ij}{%
1489 \UseTextSymbol{OT1}{\ij}}
1490 \ProvideTextCommandDefault{\IJ}{%
1491 \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 Stipcevic Mario, (stipcevic@olimp.irb.hr).

```

1492 \def\crrtic@{\hrule height0.1ex width0.3em}
1493 \def\crttic@{\hrule height0.1ex width0.33em}
1494 \def\ddj@{%
1495 \setbox0\hbox{d}\dimen@=\ht0
1496 \advance\dimen@1ex
1497 \dimen@.45\dimen@
1498 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1499 \advance\dimen@ii.5ex
1500 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
1501 \def\DDJ@{%
1502 \setbox0\hbox{D}\dimen@=.55\ht0
1503 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1504 \advance\dimen@ii.15ex % correction for the dash position
1505 \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
1506 \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
1507 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
1508 %
1509 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
1510 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}

```

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

```

1511 \ProvideTextCommandDefault{\dj}{%
1512 \UseTextSymbol{OT1}{\dj}}
1513 \ProvideTextCommandDefault{\DJ}{%
1514 \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.

```

1515 \DeclareTextCommand{\SS}{OT1}{SS}
1516 \ProvideTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}

```

9.12.3 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 1517 \ProvideTextCommandDefault{\glq}{%
      1518 \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}
```

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

```
1519 \ProvideTextCommand{\grq}{T1}{%
1520 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}}
1521 \ProvideTextCommand{\grq}{TU}{%
1522 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}}
1523 \ProvideTextCommand{\grq}{OT1}{%
1524 \save@sf@q{\kern-.0125em
1525 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
1526 \kern.07em\relax}}
1527 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}
```

`\glqq` The ‘german’ double quotes.

```
\grqq 1528 \ProvideTextCommandDefault{\glqq}{%
      1529 \textormath{\quotedblbase}{\mbox{\quotedblbase}}}
```

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

```
1530 \ProvideTextCommand{\grqq}{T1}{%
1531 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}}
1532 \ProvideTextCommand{\grqq}{TU}{%
1533 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}}
1534 \ProvideTextCommand{\grqq}{OT1}{%
1535 \save@sf@q{\kern-.07em
1536 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
1537 \kern.07em\relax}}
1538 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}
```

`\flq` The ‘french’ single guillemets.

```
\frq 1539 \ProvideTextCommandDefault{\flq}{%
      1540 \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}}
1541 \ProvideTextCommandDefault{\frq}{%
      1542 \textormath{\guilsinglright}{\mbox{\guilsinglright}}}}
```

`\flqq` The ‘french’ double guillemets.

```
\frqq 1543 \ProvideTextCommandDefault{\flqq}{%
      1544 \textormath{\guillemotleft}{\mbox{\guillemotleft}}}}
1545 \ProvideTextCommandDefault{\frqq}{%
      1546 \textormath{\guillemotright}{\mbox{\guillemotright}}}}
```

9.12.4 Umlauts and tremas

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

`\umlauthigh` 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).

```
1547 \def\umlauthigh{%
1548   \def\bbl@umlauta##1{\leavevmode\bgroup%
1549     \expandafter\accent\csname\fontencoding dqpos\endcsname
1550     ##1\bbl@allowhyphens\egroup}%
1551   \let\bbl@umlaute\bbl@umlauta}
1552 \def\umlautlow{%
1553   \def\bbl@umlauta{\protect\lower@umlaut}}
1554 \def\umlautelow{%
1555   \def\bbl@umlaute{\protect\lower@umlaut}}
1556 \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.

```
1557 \expandafter\ifx\csname U@D\endcsname\relax
1558   \csname newdimen\endcsname\U@D
1559 \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.

```
1560 \def\lower@umlaut#1{%
1561   \leavevmode\bgroup
1562   \U@D 1ex%
1563   {\setbox\z@\hbox{%
1564     \expandafter\char\csname\fontencoding dqpos\endcsname}%
1565     \dimen@ -.45ex\advance\dimen@\ht\z@
1566     \ifdim 1ex<\dimen@ \fontdimen5\fontdimen@ \fi}%
1567   \expandafter\accent\csname\fontencoding dqpos\endcsname
1568   \fontdimen5\font\U@D #1%
1569   \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).

```
1570 \AtBeginDocument{%
1571   \DeclareTextCompositeCommand{\}{OT1}{a}{\bbl@umlauta{a}}%
1572   \DeclareTextCompositeCommand{\}{OT1}{e}{\bbl@umlaute{e}}%
1573   \DeclareTextCompositeCommand{\}{OT1}{i}{\bbl@umlaute{\i}}%
1574   \DeclareTextCompositeCommand{\}{OT1}{\i}{\bbl@umlaute{\i}}%
1575   \DeclareTextCompositeCommand{\}{OT1}{o}{\bbl@umlauta{o}}%
1576   \DeclareTextCompositeCommand{\}{OT1}{u}{\bbl@umlauta{u}}%
1577   \DeclareTextCompositeCommand{\}{OT1}{A}{\bbl@umlauta{A}}%
1578   \DeclareTextCompositeCommand{\}{OT1}{E}{\bbl@umlaute{E}}%
```

```

1579 \DeclareTextCompositeCommand{"}{OT1}{I}{\bbl@umlaut{I}}%
1580 \DeclareTextCompositeCommand{"}{OT1}{O}{\bbl@umlaut{O}}%
1581 \DeclareTextCompositeCommand{"}{OT1}{U}{\bbl@umlaut{U}}%
1582 }

```

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

```

1583 \ifx\l@english\undefined
1584 \chardef\l@english\z@
1585 \fi
1586 \main@language{english}

```

9.13 Layout

Work in progress.

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

```

1587 \bbl@trace{Bidi layout}
1588 \providecommand\IfBabelLayout[3]{#3}%
1589 \newcommand\BabelPatchSection[1]{%
1590   \@ifundefined{#1}{}{%
1591     \bbl@exp{\let\bbl@ss@#1>\<#1>}%
1592     \@namedef{#1}{%
1593       \@ifstar{\bbl@presec@s{#1}}%
1594       {\@dblarg{\bbl@presec@x{#1}}}}%
1595 \def\bbl@presec@x#1[#2]#3{%
1596   \bbl@exp{%
1597     \\select@language@x{\bbl@main@language}%
1598     \\@nameuse{bbl@sspre@#1}%
1599     \\@nameuse{bbl@ss@#1}%
1600     [\\foreignlanguage{\languagename}{\unexpanded{#2}}}%
1601     {\\foreignlanguage{\languagename}{\unexpanded{#3}}}%
1602     \\select@language@x{\languagename}}%
1603 \def\bbl@presec@s#1#2{%
1604   \bbl@exp{%
1605     \\select@language@x{\bbl@main@language}%
1606     \\@nameuse{bbl@sspre@#1}%
1607     \\@nameuse{bbl@ss@#1}%
1608     {\\foreignlanguage{\languagename}{\unexpanded{#2}}}%
1609     \\select@language@x{\languagename}}%
1610 \IfBabelLayout{sectioning}%
1611   {\BabelPatchSection{part}%
1612    \BabelPatchSection{chapter}%
1613    \BabelPatchSection{section}%
1614    \BabelPatchSection{subsection}%
1615    \BabelPatchSection{subsubsection}%
1616    \BabelPatchSection{paragraph}%
1617    \BabelPatchSection{subparagraph}%
1618    \def\babel@toc#1{%
1619      \select@language@x{\bbl@main@language}}}%
1620 \IfBabelLayout{captions}%
1621   {\BabelPatchSection{caption}}%

```

Now we load definition files for engines.

```

1622 \bbl@trace{Input engine specific macros}
1623 \ifcase\bbl@engine
1624   \input txtbabel.def
1625 \or
1626   \input luababel.def

```

```

1627 \or
1628 \input xebabel.def
1629 \fi

```

9.14 Creating languages

`\babelprovide` is a general purpose tool for creating languages. Currently it just creates the language infrastructure, but in the future it will be able to read data from ini files, as well as to create variants. Unlike the nil pseudo-language, captions are defined, but with a warning to invite the user to provide the real string.

```

1630 \bbl@trace{Creating languages and reading ini files}
1631 \newcommand\babelprovide[2][{}%
1632   \let\bbl@savelangname\language
1633   \def\language{#2}%
1634   \let\bbl@KVP@captions\@nil
1635   \let\bbl@KVP@import\@nil
1636   \let\bbl@KVP@main\@nil
1637   \let\bbl@KVP@script\@nil
1638   \let\bbl@KVP@language\@nil
1639   \let\bbl@KVP@dir\@nil
1640   \let\bbl@KVP@hyphenrules\@nil
1641   \let\bbl@KVP@mapfont\@nil
1642   \let\bbl@KVP@maparabic\@nil
1643   \bbl@forkv{#1}{\bbl@csarg\def{KVP@##1}{##2}}% TODO - error handling
1644   \ifx\bbl@KVP@import\@nil\else
1645     \bbl@exp{\bbl@ifblank{\bbl@KVP@import}}%
1646     {\begingroup
1647       \def\BabelBeforeIni##1##2{\gdef\bbl@KVP@import{##1}\endinput}%
1648       \InputIfFileExists{babel-#2.tex}{}{}%
1649       \endgroup}%
1650     {}%
1651   \fi
1652   \ifx\bbl@KVP@captions\@nil
1653     \let\bbl@KVP@captions\bbl@KVP@import
1654   \fi
1655   \bbl@ifunset{date#2}%
1656   {\bbl@provide@new{#2}}%
1657   {\bbl@ifblank{#1}%
1658     {\bbl@error
1659       {If you want to modify `#2' you must tell how in\\
1660        the optional argument. Currently there are three\\
1661        options: captions=lang-tag, hyphenrules=lang-list\\
1662        import=lang-tag}%
1663       {Use this macro as documented}}%
1664     {\bbl@provide@renew{#2}}}%
1665   \bbl@exp{\bbl@babelensure[exclude=\today]{#2}}%
1666   \bbl@ifunset{\bbl@ensure@\language}%
1667   {\bbl@exp{%
1668     \DeclareRobustCommand\<bbl@ensure@\language>[1]{%
1669       \foreignlanguage{\language}%
1670       {###1}}}%
1671   }%
1672   \ifx\bbl@KVP@script\@nil\else
1673     \bbl@csarg\edef{sname#2}{\bbl@KVP@script}%
1674   \fi
1675   \ifx\bbl@KVP@language\@nil\else
1676     \bbl@csarg\edef{lname#2}{\bbl@KVP@language}%
1677   \fi

```

```

1678 \ifx\bb1@KVP@mapfont\@nil\else
1679 \bb1@ifsamestring{\bb1@KVP@mapfont}{direction}{}%
1680 {\bb1@error{Option '\bb1@KVP@mapfont' unknown for\%
1681 mapfont. Use 'direction'.%
1682 {See the manual for details.}}}%
1683 \bb1@ifunset{\bb1@lsys@\language}\bb1@provide@lsys{\language}{}%
1684 \bb1@ifunset{\bb1@wdir@\language}\bb1@provide@dirs{\language}{}%
1685 \ifx\bb1@mapselect\@undefined
1686 \AtBeginDocument{%
1687 \expandafter\bb1@add\csname selectfont \endcsname{\bb1@mapselect}%
1688 {\selectfont}}%
1689 \def\bb1@mapselect{%
1690 \let\bb1@mapselect\relax
1691 \edef\bb1@prefontid{\fontid\font}}%
1692 \def\bb1@mapdir##1{%
1693 {\def\language{##1}\bb1@switchfont
1694 \directlua{Babel.fontmap
1695 [\the\csname \bb1@wdir##1\endcsname]%
1696 [\bb1@prefontid]=\fontid\font}}}%
1697 \fi
1698 \bb1@exp{\bb1@add\bb1@mapselect{\bb1@mapdir{\language}}}%
1699 \fi
1700 \ifcase\bb1@engine\else
1701 \bb1@ifunset{\bb1@dgnat@\language}{}%
1702 {\expandafter\ifx\csname \bb1@dgnat@\language\endcsname\@empty\else
1703 \expandafter\expandafter\expandafter
1704 \bb1@setdigits\csname \bb1@dgnat@\language\endcsname
1705 \ifx\bb1@KVP@maparabic\@nil\else
1706 \ifx\bb1@latinarabic\@undefined
1707 \expandafter\let\expandafter\@arabic
1708 \csname \bb1@counter@\language\endcsname
1709 \else % ie, if layout=counters, which redefines \@arabic
1710 \expandafter\let\expandafter\bb1@latinarabic
1711 \csname \bb1@counter@\language\endcsname
1712 \fi
1713 \fi
1714 \fi}%
1715 \fi
1716 \let\language\bb1@savelangname}
1717 \def\bb1@setdigits#1#2#3#4#5{%
1718 \bb1@exp{%
1719 \def\<\language digits>####1{% ie, \langdigits
1720 \<\bb1@digits@\language>####1\@nil}%
1721 \def\<\language counter>####1{% ie, \langcounter
1722 \expandafter\<\bb1@counter@\language>%
1723 \csname c@####1\endcsname}%
1724 \def\<\bb1@counter@\language>####1{% ie, \bb1@counter@lang
1725 \expandafter\<\bb1@digits@\language>%
1726 \number####1\@nil}}%
1727 \def\bb1@tempa##1##2##3##4##5{%
1728 \bb1@exp{% Wow, quite a lot of hashes! :-(
1729 \def\<\bb1@digits@\language>#####1{%
1730 \ifx#####1\@nil % ie, \bb1@digits@lang
1731 \else
1732 \ifx0#####1#1%
1733 \else\ifx1#####1#2%
1734 \else\ifx2#####1#3%
1735 \else\ifx3#####1#4%
1736 \else\ifx4#####1#5%

```



```

1793 \bbl@savetoday
1794 \bbl@savedate
1795 \EndBabelCommands
1796 \fi
1797 \bbl@provide@hyphens{#1}}

```

The hyphenrules option is handled with an auxiliary macro.

```

1798 \def\bbl@provide@hyphens#1{%
1799 \let\bbl@tempa\relax
1800 \ifx\bbl@KVP@hyphenrules\@nil\else
1801 \bbl@replace\bbl@KVP@hyphenrules{ }{,}%
1802 \bbl@foreach\bbl@KVP@hyphenrules{%
1803 \ifx\bbl@tempa\relax % if not yet found
1804 \bbl@ifsamestring{##1}{+}%
1805 {{\bbl@exp{\\addlanguage\<l@##1>}}}%
1806 {}}%
1807 \bbl@ifunset{l@##1}%
1808 {}%
1809 {\bbl@exp{\let\bbl@tempa\<l@##1>}}}%
1810 \fi}%
1811 \fi
1812 \ifx\bbl@tempa\relax % if no opt or no language in opt found
1813 \ifx\bbl@KVP@import\@nil\else % if importing
1814 \bbl@exp{%
1815 \\bbl@ifblank{\@nameuse{bbl@hyphr@#1}}%
1816 {}}%
1817 {\let\\bbl@tempa\<l@\@nameuse{bbl@hyphr@\language}\>}}%
1818 \fi
1819 \fi
1820 \bbl@ifunset{bbl@tempa}% ie, relax or undefined
1821 {\bbl@ifunset{l@#1}% no hyphenrules found - fallback
1822 {\bbl@exp{\\adddialect\<l@#1>\language}}%
1823 {}}% so, l@<lang> is ok - nothing to do
1824 {\bbl@exp{\\adddialect\<l@#1>\bbl@tempa}}}% found in opt list or ini

```

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. *TODO - Work in progress.*

```

1825 \def\bbl@read@ini#1{%
1826 \openin1=babel-#1.ini
1827 \ifeof1
1828 \bbl@error
1829 {There is no ini file for the requested language\\%
1830 (#1). Perhaps you misspelled it or your installation\\%
1831 is not complete.}%
1832 {Fix the name or reinstall babel.}%
1833 \else
1834 \let\bbl@section\@empty
1835 \let\bbl@savestrings\@empty
1836 \let\bbl@savetoday\@empty
1837 \let\bbl@savedate\@empty
1838 \let\bbl@inireader\bbl@iniskip
1839 \bbl@info{Importing data from babel-#1.ini for \language}%
1840 \loop
1841 \if T\ifeof1F\fi T\relax % Trick, because inside \loop
1842 \endlinechar\m@ne
1843 \read1 to \bbl@line
1844 \endlinechar\^^M
1845 \ifx\bbl@line\@empty\else
1846 \expandafter\bbl@iniline\bbl@line\bbl@iniline

```

```

1847     \fi
1848     \repeat
1849   \fi}
1850 \def\bbl@inline#1\bbl@inline{%
1851   \@ifnextchar[\bbl@inisec{\@ifnextchar\bbl@iniskip\bbl@inireader}#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.

```

1852 \def\bbl@iniskip#1\@{%      if starts with ;
1853 \def\bbl@inisec[#1]#2\@{%   if starts with opening bracket
1854   \@nameuse{bbl@secpost@bbl@section}% ends previous section
1855   \def\bbl@section{#1}%
1856   \@nameuse{bbl@secpre@bbl@section}%  starts current section
1857   \bbl@ifunset{bbl@secline@#1}%
1858   {\let\bbl@inireader\bbl@iniskip}%
1859   {\bbl@exp{\let\bbl@inireader<bbl@secline@#1>}}}

```

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

```

1860 \def\bbl@inikv#1=#2\@{%    key=value
1861   \bbl@trim@def\bbl@tempa{#1}%
1862   \bbl@trim\toks@{#2}%
1863   \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.

```

1864 \def\bbl@exportkey#1#2#3{%
1865   \bbl@ifunset{bbl@kv@#2}%
1866   {\bbl@csarg\gdef{#1@language}{#3}}%
1867   {\expandafter\ifx\csname bbl@kv@#2\endcsname\empty
1868     \bbl@csarg\gdef{#1@language}{#3}}%
1869   \else
1870     \bbl@exp{\global\let\bbl@#1@language>\bbl@kv@#2>}%
1871   \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.

```

1872 \let\bbl@secline@identification\bbl@inikv
1873 \def\bbl@secpost@identification{%
1874   \bbl@exportkey{lname}{identification.name.english}{}%
1875   \bbl@exportkey{lbcpr}{identification.tag.bcp47}{}%
1876   \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
1877   \bbl@exportkey{sname}{identification.script.name}{}%
1878   \bbl@exportkey{sbcpr}{identification.script.tag.bcp47}{}%
1879   \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}}
1880 \let\bbl@secline@typography\bbl@inikv
1881 \let\bbl@secline@numbers\bbl@inikv
1882 \def\bbl@after@ini{%
1883   \bbl@exportkey{lftm}{typography.lefthyphenmin}{2}%
1884   \bbl@exportkey{rgtm}{typography.righthyphenmin}{3}%
1885   \bbl@exportkey{hyphr}{typography.hyphenrules}{}%
1886   \bbl@exportkey{dgndat}{numbers.digits.native}{}%
1887   \bbl@xin@{0.5}{\@nameuse{bbl@kv@identification.version}}%
1888   \ifin@
1889     \bbl@warning{%
1890       There are neither captions nor date in `language'.\%
1891       It may not be suitable for proper typesetting, and it\%
1892       could change. Reported}%
1893   \fi

```

```

1894 \bbl@xin@{0.9}{\@nameuse{\bbl@kv@identification.version}}%
1895 \ifin@
1896 \bbl@warning{%
1897     The '\language' date format may not be suitable\\%
1898     for proper typesetting, and therefore it very likely will\\%
1899     change in a future release. Reported}%
1900 \fi
1901 \bbl@tglobal\bbl@savetoday
1902 \bbl@tglobal\bbl@savestate

```

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

```

1903 \ifcase\bbl@engine
1904 \bbl@csarg\def{secline@captions.licr}#1=#2\@@{%
1905     \bbl@ini@captions@aux{#1}{#2}}
1906 \bbl@csarg\def{secline@date.gregorian}#1=#2\@@{%           for defaults
1907     \bbl@ini@dategreg#1...\relax{#2}}
1908 \bbl@csarg\def{secline@date.gregorian.licr}#1=#2\@@{%      override
1909     \bbl@ini@dategreg#1...\relax{#2}}
1910 \else
1911 \def\bbl@secline@captions#1=#2\@@{%
1912     \bbl@ini@captions@aux{#1}{#2}}
1913 \bbl@csarg\def{secline@date.gregorian}#1=#2\@@{%
1914     \bbl@ini@dategreg#1...\relax{#2}}
1915 \fi

```

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

```

1916 \def\bbl@ini@captions@aux#1#2{%
1917     \bbl@trim\def\bbl@tempa{#1}%
1918     \bbl@ifblank{#2}%
1919     {\bbl@exp{%
1920         \toks@{\bbl@nocaption{\bbl@tempa}{\language\bbl@tempa name}}}%
1921     {\bbl@trim\toks@{#2}}}%
1922 \bbl@exp{%
1923     \bbl@add\bbl@savestrings{%
1924         \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.

```

1925 \bbl@csarg\def{secpredate.gregorian.licr}{%
1926     \ifcase\bbl@engine\let\bbl@savestate\empty\fi}
1927 \def\bbl@ini@dategreg#1.#2.#3.#4\relax#5{% TODO - ignore with 'captions'
1928     \bbl@trim\def\bbl@tempa{#1.#2}%
1929     \bbl@ifsamestring{\bbl@tempa}{months.wide}%
1930     {\bbl@trim\def\bbl@tempa{#3}%
1931         \bbl@trim\toks@{#5}%
1932         \bbl@exp{%
1933             \bbl@add\bbl@savestate{%
1934                 \SetString\<month\romannumeral\bbl@tempa name>{\the\toks@}}}%
1935     {\bbl@ifsamestring{\bbl@tempa}{date.long}%
1936         {\bbl@trim\def\bbl@toreplace{#5}%
1937             \bbl@TG@date
1938             \global\bbl@csarg\let{date@\language}\bbl@toreplace
1939             \bbl@exp{%
1940                 \gdef\<\language date>{\protect\<\language date >}}%
1941                 \gdef\<\language date >####1####2####3{%
1942                     \bbl@usedategroupttrue
1943                     \<\bbl@ensure@\language>{%
1944                         \<\bbl@date@\language>{####1}{####2}{####3}}}%

```

```

1945      \\bbl@add\\bbl@savetoday{%
1946      \\SetString\\today{%
1947      \\<\\language\\name date>{\\the\\year}{\\the\\month}{\\the\\day}}}%
1948      {}

```

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.

```

1949 \\newcommand\\BabelDateSpace{\\nobreakspace}
1950 \\newcommand\\BabelDateDot{.\\@}
1951 \\newcommand\\BabelDated[1]{\\number#1}
1952 \\newcommand\\BabelDatedd[1]{\\ifnum#1<10 0\\fi\\number#1}
1953 \\newcommand\\BabelDateM[1]{\\number#1}
1954 \\newcommand\\BabelDateMM[1]{\\ifnum#1<10 0\\fi\\number#1}
1955 \\newcommand\\BabelDateMMMM[1]{%
1956   \\csname month\\romannumeral#1name\\endcsname}%
1957 \\newcommand\\BabelDatey[1]{\\number#1}%
1958 \\newcommand\\BabelDateyy[1]{%
1959   \\ifnum#1<10 0\\number#1 %
1960   \\else\\ifnum#1<100 \\number#1 %
1961   \\else\\ifnum#1<1000 \\expandafter\\@gobble\\number#1 %
1962   \\else\\ifnum#1<10000 \\expandafter\\@gobbletwo\\number#1 %
1963   \\else
1964     \\bbl@error
1965     {Currently two-digit years are restricted to the\\
1966     range 0-9999.}%
1967     {There is little you can do. Sorry.}%
1968     \\fi\\fi\\fi\\fi}
1969 \\newcommand\\BabelDateyyyy[1]{\\number#1}
1970 \\def\\bbl@replace@finish@iii#1{%
1971   \\bbl@exp{\\def\\#1####1####2####3{\\the\\toks@}}
1972   \\def\\bbl@TG@date{%
1973     \\bbl@replace\\bbl@toreplace{[ ]}{\\BabelDateSpace}}%
1974     \\bbl@replace\\bbl@toreplace{[. ]}{\\BabelDateDot}}%
1975     \\bbl@replace\\bbl@toreplace{[d]}{\\BabelDated{####3}}%
1976     \\bbl@replace\\bbl@toreplace{[dd]}{\\BabelDatedd{####3}}%
1977     \\bbl@replace\\bbl@toreplace{[M]}{\\BabelDateM{####2}}%
1978     \\bbl@replace\\bbl@toreplace{[MM]}{\\BabelDateMM{####2}}%
1979     \\bbl@replace\\bbl@toreplace{[MMMM]}{\\BabelDateMMMM{####2}}%
1980     \\bbl@replace\\bbl@toreplace{[y]}{\\BabelDatey{####1}}%
1981     \\bbl@replace\\bbl@toreplace{[yy]}{\\BabelDateyy{####1}}%
1982     \\bbl@replace\\bbl@toreplace{[yyyy]}{\\BabelDateyyyy{####1}}%
1983 % Note after \\bbl@replace \\toks@ contains the resulting string.
1984 % TODO - Using this implicit behavior doesn't seem a good idea.
1985     \\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. It also activates interword spacing in Southeast Asian scripts if not done before.

```

1986 \\def\\bbl@provide@lsys#1{%
1987   \\bbl@ifunset{\\bbl@lname@#1}%
1988   {\\bbl@ini@ids{#1}}%
1989   {}%
1990   \\bbl@csarg\\let{\\lsys@#1}\\@empty
1991   \\bbl@ifunset{\\bbl@sname@#1}{\\bbl@csarg\\gdef{sname@#1}{Default}}}%
1992   \\bbl@ifunset{\\bbl@sotf@#1}{\\bbl@csarg\\gdef{sotf@#1}{DFLT}}}%
1993   \\bbl@csarg\\bbl@add@list{\\lsys@#1}{Script=\\bbl@cs{sname@#1}}%
1994   \\bbl@ifunset{\\bbl@lname@#1}%
1995   {\\bbl@csarg\\bbl@add@list{\\lsys@#1}{Language=\\bbl@cs{\\lname@#1}}}%

```

```

1996 \bbl@csarg\bbl@toglobal{lsys@#1}%
1997 \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@seascripts}%
1998 \ifin@
1999 \@nameuse{bbl@seainterword}%
2000 \fi
2001 \bbl@exp{% TODO - should be global, but even local does its job
2002           % I'm still not sure -- must investigate
2003   \<keys_if_exist:nnF>{fontspec-opentype}{Script/\bbl@cs{sname@#1}}%
2004   {\newfontscript{\bbl@cs{sname@#1}}{\bbl@cs{sotf@#1}}}%
2005   \<keys_if_exist:nnF>{fontspec-opentype}{Language/\bbl@cs{lname@#1}}%
2006   {\newfontlanguage{\bbl@cs{lname@#1}}{\bbl@cs{lotf@#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.

```

2007 \def\bbl@ini@ids#1{%
2008   \def\BabelBeforeIni##1##2{%
2009     \begingroup
2010       \bbl@add\bbl@secpost@identification{\closein1 }%
2011       \catcode`\[=12 \catcode`\]=12 \catcode`\==12 %
2012       \bbl@read@ini{##1}%
2013     \endgroup}% boxed, to avoid extra spaces:
2014   {\setbox\z@\hbox{\InputIfFileExists{babel-#1.tex}{}}{}}}

```

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

10.1 The redefinition of the style commands

The rest of the code in this file can only be processed by \LaTeX , so we check the current format. If it is plain \TeX , processing should stop here. But, because of the need to limit the scope of the definition of `\format`, a macro that is used locally in the following `\if` statement, this comparison is done inside a group. To prevent \TeX from complaining about an unclosed group, the processing of the command `\endinput` is deferred until after the group is closed. This is accomplished by the command `\aftergroup`.

```

2015 {\def\format{lpain}}
2016 \ifx\fmtname\format
2017 \else
2018   \def\format{LaTeX2e}
2019   \ifx\fmtname\format
2020   \else
2021     \aftergroup\endinput
2022   \fi
2023 \fi}

```

10.2 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 *T_EXbook* [2] (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.

```
2024 %\bbl@redefine\newlabel#1#2{%
2025 % \@safe@activestrue\org@newlabel{#1}{#2}\@safe@activfalse}
```

`\@newl@bel` We need to change the definition of the *T_EX*-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.

```
2026 <<*More package options>> ≡
2027 \DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
2028 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
2029 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
2030 <</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.

```
2031 \bbl@trace{Cross referencing macros}
2032 \ifx\bbl@opt@safe\@empty\else
2033 \def\@newl@bel#1#2#3{%
2034   {\@safe@activestrue
2035     \bbl@ifunset{#1@#2}%
2036     \relax
2037     {\gdef\@multiplelabels{%
2038       \@latex@warning@no@line{There were multiply-defined labels}}}%
2039     \@latex@warning@no@line{Label `#2' multiply defined}}%
2040   \global\@namedef{#1@#2}{#3}}}
```

`\@testdef` An internal *T_EX* 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 *T_EX* keeps reporting that the labels may have changed.

```
2041 \CheckCommand*\@testdef[3]{%
2042   \def\reserved@a{#3}%
2043   \expandafter\ifx\csname#1@#2\endcsname\reserved@a
2044   \else
2045     \@tempwattrue
2046   \fi}
```

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

```
2047 \def\@testdef#1#2#3{%
2048   \@safe@activestrue
```

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

```
2049 \expandafter\let\expandafter\bbl@tempa\csname #1@#2\endcsname
```

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

```
2050 \def\bbl@tempb{#3}%
2051 \@safe@activesfalse
```

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

```
2052 \ifx\bbl@tempa\relax
2053 \else
2054 \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
2055 \fi
```

We do the same for `\bbl@tempb`.

```
2056 \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
```

If the label didn't change, `\bbl@tempa` and `\bbl@tempb` should be identical macros.

```
2057 \ifx\bbl@tempa\bbl@tempb
2058 \else
2059 \@tempswatrue
2060 \fi}
2061 \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.

```
2062 \bbl@xin@{R}\bbl@opt@safe
2063 \ifin@
2064 \bbl@redefineroobust\ref{#1}{%
2065 \@safe@activetrue\org@ref{#1}\@safe@activesfalse}
2066 \bbl@redefineroobust\pageref{#1}{%
2067 \@safe@activetrue\org@pageref{#1}\@safe@activesfalse}
2068 \else
2069 \let\org@ref\ref
2070 \let\org@pageref\pageref
2071 \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.

```
2072 \bbl@xin@{B}\bbl@opt@safe
2073 \ifin@
2074 \bbl@redefine\@citex[#1]#2{%
2075 \@safe@activetrue\edef\@tempa{#2}\@safe@activesfalse
2076 \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.

```
2077 \AtBeginDocument{%
2078 \@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.)


```

2079 \def\@citex[#1][#2]#3{%
2080   \@safe@activestrue\edef\@tempa{#3}\@safe@activesfalse
2081   \org@citex[#1][#2]{\@tempa}}%
2082 }{}}

```

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

```

2083 \AtBeginDocument{%
2084   \@ifpackageloaded{cite}{%
2085     \def\@citex[#1]#2{%
2086       \@safe@activestrue\org@citex[#1][#2]\@safe@activesfalse}%
2087     }{}}

```

\nocite The macro \nocite which is used to instruct BiBTeX to extract uncited references from the database.

```

2088 \bbl@redefine\nocite#1{%
2089   \@safe@activestrue\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@activestrue is in effect. This switch needs to be reset inside the \hbox which contains the citation label. In order to determine during .aux file processing which definition of \bibcite is needed we define \bibcite in such a way that it redefines itself with the proper definition.

```

2090 \bbl@redefine\bibcite{%

```

We call \bbl@cite@choice to select the proper definition for \bibcite. This new definition is then activated.

```

2091   \bbl@cite@choice
2092   \bibcite}

```

\bbl@bibcite The macro \bbl@bibcite holds the definition of \bibcite needed when neither natbib nor cite is loaded.

```

2093 \def\bbl@bibcite#1#2{%
2094   \org@bibcite{#1}{\@safe@activesfalse#2}}

```

\bbl@cite@choice The macro \bbl@cite@choice determines which definition of \bibcite is needed.

```

2095 \def\bbl@cite@choice{%

```

First we give \bibcite its default definition.

```

2096   \global\let\bibcite\bbl@bibcite

```

Then, when natbib is loaded we restore the original definition of \bibcite.

```

2097   \@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{}%

```

For cite we do the same.

```

2098   \@ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{}%

```

Make sure this only happens once.

```

2099   \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.

```

2100 \AtBeginDocument{\bbl@cite@choice}

```

`\@bibitem` One of the two internal \LaTeX macros called by `\bibitem` that write the citation label on the .aux file.

```
2101 \bbl@redefine\@bibitem#1{%
2102   \@safe@activetrue\org@bibitem{#1}\@safe@activesfalse}
2103 \else
2104   \let\org@nocite\nocite
2105   \let\org@@citex\@citex
2106   \let\org@bibcite\@bibcite
2107   \let\org@bibitem\@bibitem
2108 \fi
```

10.3 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.

```
2109 \bbl@trace{Marks}
2110 \IfBabelLayout{sectioning}
2111   {\ifx\bbl@opt@headfoot\@nnil
2112     \g@addto@macro\@resetactivechars{%
2113       \set@typeset@protect
2114       \expandafter\select@language@x\expandafter{\bbl@main@language}%
2115       \let\protect\@noexpand
2116       \edef\thepage{%
2117         \noexpand\babelsublr{\unexpanded\expandafter{\thepage}}}%
2118     \fi}
2119   {\bbl@redefine\markright#1{%
2120     \bbl@ifblank{#1}%
2121     {\org@markright{}}%
2122     {\toks@{#1}%
2123       \bbl@exp{%
2124         \\org@markright{\\protect\\foreignlanguage{\language}%
2125           {\\\protect\\bbl@restore@actives\the\toks@}}}%

```

`\markboth` The definition of `\markboth` is equivalent to that of `\markright`, except that we need two token registers. The 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`.

```
2126 \ifx\@mkboth\markboth
2127   \def\bbl@tempc{\let\@mkboth\markboth}
2128 \else
2129   \def\bbl@tempc{}
2130 \fi
```

Now we can start the new definition of `\markboth`

```
2131 \bbl@redefine\markboth#1#2{%
2132   \protected@edef\bbl@tempb##1{%
2133     \protect\foreignlanguage
2134     {\language}\protect\bbl@restore@actives##1}%

```

```

2135 \bbl@ifblank{#1}%
2136 {\toks@{}}%
2137 {\toks@expandafter{\bbl@tempb{#1}}}%
2138 \bbl@ifblank{#2}%
2139 {\@temptokena{}}%
2140 {\@temptokenaexpandafter{\bbl@tempb{#2}}}%
2141 \bbl@exp{\@org@markboth{\the\toks@}{\the\@temptokena}}

```

and copy it to \@mkboth if necessary.

```

2142 \bbl@tempc} % end \IfBabelLayout

```

10.4 Preventing clashes with other packages

10.4.1 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.

```

2143 \bbl@trace{Preventing clashes with other packages}
2144 \bbl@xin@{R}\bbl@opt@safe
2145 \ifin@
2146 \AtBeginDocument{%
2147 \ifpackageloaded{ifthen}{%

```

Then we can redefine `\ifthenelse`:

```

2148 \bbl@redefine@long\ifthenelse#1#2#3{%

```

We want to revert the definition of `\pageref` and `\ref` to their original definition for the first argument of `\ifthenelse`, so we first need to store their current meanings.

```

2149 \let\bbl@temp@pref\pageref
2150 \let\pageref\org@pageref
2151 \let\bbl@temp@ref\ref
2152 \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.

```

2153 \@safe@activetrue
2154 \org@ifthenelse{#1}%
2155 {\let\pageref\bbl@temp@pref
2156 \let\ref\bbl@temp@ref
2157 \@safe@activesfalse
2158 #2}%
2159 {\let\pageref\bbl@temp@pref
2160 \let\ref\bbl@temp@ref
2161 \@safe@activesfalse
2162 #3}%

```

```

2163     }%
2164   }{}%
2165 }

```

10.4.2 varioref

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

```

\Ref 2166 \AtBeginDocument{%
2167   \ifpackageloaded{varioref}{%
2168     \bbl@redefine\@@vpageref#1[#2]#3{%
2169       \@safe@activestrue
2170       \org@@vpageref{#1}{#2}{#3}%
2171       \@safe@activesfalse}%

```

The same needs to happen for `\vrefpagenum`.

```

2172   \bbl@redefine\vrefpagenum#1#2{%
2173     \@safe@activestrue
2174     \org@vrefpagenum{#1}{#2}%
2175     \@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.

```

2176   \expandafter\def\csname Ref \endcsname#1{%
2177     \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
2178   }{}%
2179 }
2180 \fi

```

10.4.3 hline

`\hline` Delaying the activation of the shorthand characters has introduced a problem with the `hline` 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 `hline` is loaded.

```

2181 \AtEndOfPackage{%
2182   \AtBeginDocument{%
2183     \ifpackageloaded{hline}%

```

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

```

2184     {\expandafter\ifx\csname normal@char:string\endcsname\relax
2185       \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.

```

2186       \makeatletter
2187       \def\currname{hline}\input{hline.sty}\makeatother
2188       \fi}%
2189     {}}}

```

10.4.4 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.

```
2190 \AtBeginDocument{%
2191   \ifx\pdfstringdefDisableCommands\@undefined\else
2192     \pdfstringdefDisableCommands{\languageshortands{system}}%
2193   \fi}
```

10.4.5 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 `\MakeUpper` case. To prevent unexpected results we need to define `\FOREIGNLANGUAGE` here.

```
2194 \DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
2195   \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.

```
2196 \def\substitutefontfamily#1#2#3{%
2197   \lowercase{\immediate\openout15=#1#2.fd\relax}%
2198   \immediate\write15{%
2199     \string\ProvidesFile{#1#2.fd}%
2200     [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
2201     \space generated font description file]^{}
2202     \string\DeclareFontFamily{#1}{#2}{}^{}
2203     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{}^{}
2204     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{}^{}
2205     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{}^{}
2206     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{}^{}
2207     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{}^{}
2208     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{}^{}
2209     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{}^{}
2210     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{}^{}
2211   }%
2212   \closeout15
2213 }
```

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

```
2214 \@onlypreamble\substitutefontfamily
```

10.5 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>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
2215 \bbl@trace{Encoding and fonts}
```

```

2216 \newcommand\BabelNonASCII
2217 {LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU,TS1,T3,TS3}
2218 \let\org@TeX\TeX
2219 \let\org@LaTeX\LaTeX
2220 \let\ensureascii\@firstofone
2221 \AtBeginDocument{%
2222   \in@false
2223   \bbl@foreach\BabelNonASCII{% is there a non-ascii enc?
2224     \ifin@false
2225       \lowercase{\bbl@xin@{,#1enc.def,},{,\@filelist,}}%
2226     \fi}%
2227   \ifin@ % if a non-ascii has been loaded
2228     \def\ensureascii#1{{\fontencoding{OT1}\selectfont#1}}%
2229     \DeclareTextCommandDefault{\TeX}{\org@TeX}%
2230     \DeclareTextCommandDefault{\LaTeX}{\org@LaTeX}%
2231     \def\bbl@tempb#1\@@{\uppercase{\bbl@tempc#1}ENC.DEF\@empty\@@}%
2232     \def\bbl@tempc#1ENC.DEF#2\@@{%
2233       \ifx\@empty#2\else
2234         \bbl@ifunset{T#1}%
2235         {}%
2236         {\bbl@xin@{,#1,},{,\BabelNonASCII,}}%
2237       \ifin@
2238         \DeclareTextCommand{\TeX}{#1}{\ensureascii{\org@TeX}}%
2239         \DeclareTextCommand{\LaTeX}{#1}{\ensureascii{\org@LaTeX}}%
2240       \else
2241         \def\ensureascii#1{{\fontencoding{#1}\selectfont#1}}%
2242       \fi}%
2243     \fi}%
2244   \bbl@foreach\@filelist{\bbl@tempb#1\@@}% TODO - \@@ de mas??
2245   \bbl@xin@{,\cf@encoding,},{,\BabelNonASCII,}%
2246   \ifin@false
2247     \edef\ensureascii#1{%
2248       \noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}}%
2249   \fi
2250 \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.

```

2251 \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.

```

2252 \AtBeginDocument{%
2253   \@ifpackageloaded{fontspec}%
2254   {\xdef\latinencoding%
2255     \ifx\UTFencname\@undefined
2256       EU\ifcase\bbl@engine\or2\or1\fi
2257     \else
2258       \UTFencname
2259     \fi}}%
2260   {\gdef\latinencoding{OT1}}%

```

```

2261 \ifx\cf@encoding\bbl@t@one
2262 \xdef\latinencoding{\bbl@t@one}%
2263 \else
2264 \@ifl@aded{def}{t1enc}{\xdef\latinencoding{\bbl@t@one}}{}%
2265 \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.

```

2266 \DeclareRobustCommand{\latintext}{%
2267 \fontencoding{\latinencoding}\selectfont
2268 \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.

```

2269 \ifx\@undefined\DeclareTextFontCommand
2270 \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
2271 \else
2272 \DeclareTextFontCommand{\textlatin}{\latintext}
2273 \fi

```

10.6 Basic bidi support

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

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

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

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

```

2274 \bbl@trace{Basic (internal) bidi support}
2275 \def\bbl@alscripts{,Arabic,Syriac,Thaana,}
2276 \def\bbl@rscripts{%
2277   ,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
2278   Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaeen,%
2279   Manichaeen,Meroitic Cursive,Meroitic,Old North Arabian,%
2280   Nabataean,N'Ko,Orkhon,Palmyrene,Inscriptional Pahlavi,%
2281   Psalter Pahlavi,Phoenician,Inscriptional Parthian,Samaritan,%
2282   Old South Arabian,}%

```

```

2283 \def\bbl@provide@dirs#1{%
2284   \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts\bbl@rscripts}%
2285   \ifin@
2286     \global\bbl@csarg\chardef{wdir@#1}\@ne
2287     \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts}%
2288     \ifin@
2289       \global\bbl@csarg\chardef{wdir@#1}\tw@ % useless in xetex
2290       \fi
2291     \else
2292       \global\bbl@csarg\chardef{wdir@#1}\z@
2293     \fi}
2294 \def\bbl@switchdir{%
2295   \bbl@ifunset{bbl@lsys@\language}{\bbl@provide@lsys{\language}}{}%
2296   \bbl@ifunset{bbl@wdir@\language}{\bbl@provide@dirs{\language}}{}%
2297   \bbl@exp{\bbl@setdirs\bbl@cs{wdir@\language}}%
2298 \def\bbl@setdirs#1{% TODO - math
2299   \ifcase\bbl@select@type % TODO - strictly, not the right test
2300     \bbl@bodydir{#1}%
2301     \bbl@pardir{#1}%
2302   \fi
2303   \bbl@textdir{#1}}
2304 \ifodd\bbl@engine % luatex=1
2305   \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
2306   \DisableBabelHook{babel-bidi}
2307   \chardef\bbl@thepardir\z@
2308 \def\bbl@getluadir#1{%
2309   \directlua{
2310     if tex.#1dir == 'TLT' then
2311       tex.sprint('0')
2312     elseif tex.#1dir == 'TRT' then
2313       tex.sprint('1')
2314     end}}
2315 \def\bbl@setluadir#1#2#3{% 1=text/par.. 2=\textdir.. 3=0 lr/1 rl
2316   \ifcase#3\relax
2317     \ifcase\bbl@getluadir{#1}\relax\else
2318       #2 TLT\relax
2319     \fi
2320   \else
2321     \ifcase\bbl@getluadir{#1}\relax
2322       #2 TRT\relax
2323     \fi
2324   \fi}
2325 \def\bbl@textdir#1{%
2326   \bbl@setluadir{text}\textdir{#1}% TODO - ?\linedir
2327   \setattribute\bbl@attr@dir{\numexpr\bbl@thepardir*3+#1}}
2328 \def\bbl@pardir#1{\bbl@setluadir{par}\pardir{#1}%
2329   \chardef\bbl@thepardir#1\relax}
2330 \def\bbl@bodydir{\bbl@setluadir{body}\bodydir}
2331 \def\bbl@pagedir{\bbl@setluadir{page}\pagedir}
2332 \def\bbl@dirparastext{\pardir\the\textdir\relax}% %%%
2333 \else % pdftex=0, xetex=2
2334   \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
2335   \DisableBabelHook{babel-bidi}
2336   \newcount\bbl@dirlevel
2337   \chardef\bbl@thetextdir\z@
2338   \chardef\bbl@thepardir\z@
2339   \def\bbl@textdir#1{%
2340     \ifcase#1\relax
2341       \chardef\bbl@thetextdir\z@

```



```

2342     \bbl@textdir@i\beginL\endL
2343   \else
2344     \chardef\bbl@thetextdir\@ne
2345     \bbl@textdir@i\beginR\endR
2346   \fi}
2347 \def\bbl@textdir@i#1#2{%
2348   \ifhmode
2349     \ifnum\currentgrouplevel>\z@
2350       \ifnum\currentgrouplevel=\bbl@dirlevel
2351         \bbl@error{Multiple bidi settings inside a group}%
2352         {I'll insert a new group, but expect wrong results.}%
2353         \bgroup\aftergroup#2\aftergroup\egroup
2354       \else
2355         \ifcase\currentgrouptype\or % 0 bottom
2356           \aftergroup#2% 1 simple {}
2357         \or
2358           \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
2359         \or
2360           \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
2361         \or\or\or % vbox vtop align
2362         \or
2363           \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
2364         \or\or\or\or\or\or % output math disc insert vcent mathchoice
2365         \or
2366           \aftergroup#2% 14 \begingroup
2367         \else
2368           \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
2369         \fi
2370       \fi
2371       \bbl@dirlevel\currentgrouplevel
2372     \fi
2373     #1%
2374   \fi}
2375 \def\bbl@pardir#1{\chardef\bbl@thepardir#1\relax}
2376 \let\bbl@bodydir\@gobble
2377 \let\bbl@pagedir\@gobble
2378 \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).

```

2379 \def\bbl@xebidipar{%
2380   \let\bbl@xebidipar\relax
2381   \TeXeTstate\@ne
2382   \def\bbl@xeverypar{%
2383     \ifcase\bbl@thepardir
2384       \ifcase\bbl@thetextdir\else\beginR\fi
2385     \else
2386       {\setbox\z@\lastbox\beginR\box\z@}%
2387     \fi}%
2388   \let\bbl@severypar\everypar
2389   \newtoks\everypar
2390   \everypar=\bbl@severypar
2391   \bbl@severypar{\bbl@xeverypar\the\everypar}}
2392 \fi

```

A tool for weak L (mainly digits).

```

2393 \DeclareRobustCommand\babelsublr[1]{\leavevmode{\bbl@textdir\z@#1}}

```

10.7 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`.

```
2394 \bbl@trace{Local Language Configuration}
2395 \ifx\loadlocalcfg\@undefined
2396   \@ifpackagewith{babel}{noconfigs}%
2397     {\let\loadlocalcfg\@gobble}%
2398     {\def\loadlocalcfg#1{%
2399       \InputIfFileExists{#1.cfg}%
2400       {\typeout{*****^J%
2401                * Local config file #1.cfg used^^J%
2402                *}}%
2403       \@empty}}
2404 \fi
```

Just to be compatible with \TeX 2.09 we add a few more lines of code:

```
2405 \ifx\@unexpandable@protect\@undefined
2406   \def\@unexpandable@protect{\noexpand\protect\noexpand}
2407   \long\def\protected@write#1#2#3{%
2408     \begingroup
2409       \let\thepage\relax
2410       #2%
2411       \let\protect\@unexpandable@protect
2412       \edef\reserved@a{\write#1{#3}}%
2413       \reserved@a
2414     \endgroup
2415     \if@nobreak\ifvmode\nobreak\fi\fi}
2416 \fi
2417 </core>
2418 <*kernel>
```

11 Multiple languages (switch.def)

Plain \TeX version 3.0 provides the primitive `\language` that is used to store the current language. When used with a pre-3.0 version this function has to be implemented by allocating a counter.

```
2419 <<Make sure ProvidesFile is defined>>
2420 \ProvidesFile{switch.def}[\<date>] [\<version>] Babel switching mechanism]
2421 <<Load macros for plain if not LaTeX>>
2422 <<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.

```
2423 \def\bbl@version{\<version>}
2424 \def\bbl@date{\<date>}
2425 \def\adddialect#1#2{%
2426   \global\chardef#1#2\relax
2427   \bbl@usehooks{adddialect}{\#1}{\#2}%
2428   \wlog{\string#1 = a dialect from \string\language#2}}
```

`\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.

```

2429 \def\bbl@fixname#1{%
2430   \begingroup
2431   \def\bbl@tempe{l@}%
2432   \edef\bbl@tempd{\noexpand\@ifundefined{\noexpand\bbl@tempe#1}}%
2433   \bbl@tempd
2434     {\lowercase\expandafter{\bbl@tempd}%
2435      {\uppercase\expandafter{\bbl@tempd}}%
2436      \@empty
2437      {\edef\bbl@tempd{\def\noexpand#1{#1}}%
2438       \uppercase\expandafter{\bbl@tempd}}}%
2439   {\edef\bbl@tempd{\def\noexpand#1{#1}}%
2440    \lowercase\expandafter{\bbl@tempd}}}%
2441   \@empty
2442   \edef\bbl@tempd{\endgroup\def\noexpand#1{#1}}%
2443   \bbl@tempd}
2444 \def\bbl@iflanguage#1{%
2445   \@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.

```

2446 \def\iflanguage#1{%
2447   \bbl@iflanguage{#1}{%
2448     \ifnum\csname l@#1\endcsname=\language
2449       \expandafter\@firstoftwo
2450     \else
2451       \expandafter\@secondoftwo
2452     \fi}}

```

11.1 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.

```
2453 \let\bbl@select@type\z@
2454 \edef\selectlanguage{%
2455   \noexpand\protect
2456   \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`.

```
2457 \ifx\@undefined\protect\let\protect\relax\fi
```

As \TeX 2.09 writes to files *expanded* whereas \TeX 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.

```
2458 \ifx\documentclass\@undefined
2459   \def\xstring{\string\string\string}
2460 \else
2461   \let\xstring\string
2462 \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.

```
2463 \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`

```
2464 \def\bbl@push@language{%
2465   \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.

```
2466 \def\bbl@pop@lang#1+#2-#3{%
2467   \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.

```
2468 \let\bbl@ifrestoring\@secondoftwo
2469 \def\bbl@pop@language{%
2470   \expandafter\bbl@pop@lang\bbl@language@stack-\bbl@language@stack
2471   \let\bbl@ifrestoring\@firstoftwo
2472   \expandafter\bbl@set@language\expandafter{\language}%
2473   \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.

```
2474 \expandafter\def\csname selectlanguage \endcsname#1{%
2475   \ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel\tw\fi
2476   \bbl@push@language
2477   \aftergroup\bbl@pop@language
2478   \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 not well defined. The list of auxiliary files can be extended by redefining \BabelContentsFiles, but make sure they are loaded inside a group (as aux, toc, lof, and lot do) or the last language of the document will remain active afterwards.

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

```
2479 \def\BabelContentsFiles{toc,lof,lot}
2480 \def\bbl@set@language#1{%
2481   \edef\language{%
2482     \ifnum\escapechar=\expandafter`\string#1\@empty
2483     \else\string#1\@empty\fi}%
2484   \select@language{\language}%
2485   \expandafter\ifx\csname date\language\endcsname\relax\else
2486     \if@filesw
2487       \protected@write\@auxout{{}\string\babel@aux{\language}}}%
2488       \bbl@usehooks{write}}}%
2489   \fi
2490 \fi}
2491 \def\select@language#1{%
2492   \ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi
2493   \edef\language{#1}%
2494   \bbl@fixname\language
2495   \bbl@iflanguage\language{%
2496     \expandafter\ifx\csname date\language\endcsname\relax
2497       \bbl@error
2498       {Unknown language `#1'. Either you have\\%
2499       misspelled its name, it has not been installed,\\%
2500       or you requested it in a previous run. Fix its name,\\%
2501       install it or just rerun the file, respectively. In\\%
2502       some cases, you may need to remove the aux file}%
2503       {You may proceed, but expect wrong results}%
2504   \else
2505     \let\bbl@select@type\z@
2506     \expandafter\bbl@switch\expandafter{\language}%
2507   \fi}}
2508 \def\babel@aux#1#2{%
2509   \expandafter\ifx\csname date#1\endcsname\relax
```

```

2510 \expandafter\ifx\csname bbl@auxwarn@#1\endcsname\relax
2511 \namedef{bbl@auxwarn@#1}{}%
2512 \bbl@warning
2513 {Unknown language `#1'. Very likely you\\%
2514 requested it in a previous run. Expect some\\%
2515 wrong results in this run, which should vanish\\%
2516 in the next one. Reported}%
2517 \fi
2518 \else
2519 \select@language{#1}%
2520 \bbl@foreach\BabelContentsFiles{%
2521 \@writefile{##1}{\babel@toc{#1}{#2}}}% % TODO - ok in plain?
2522 \fi}
2523 \def\babel@toc#1#2{%
2524 \select@language{#1}}

```

A bit of optimization. Select in heads/foots the language only if necessary. The real thing is in `babel.def`.

```

2525 \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.

```

2526 \newif\ifbbl@usedategroup
2527 \def\bbl@switch#1{%
2528 \originalTeX
2529 \expandafter\def\expandafter\originalTeX\expandafter{%
2530 \csname noextras#1\endcsname
2531 \let\originalTeX\@empty
2532 \babel@beginsave}%
2533 \bbl@usehooks{afterreset}}}%
2534 \languageshorthands{none}%
2535 \ifcase\bbl@select@type
2536 \ifhmode
2537 \hskip\z@skip % trick to ignore spaces
2538 \csname captions#1\endcsname\relax
2539 \csname date#1\endcsname\relax
2540 \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
2541 \else
2542 \csname captions#1\endcsname\relax
2543 \csname date#1\endcsname\relax
2544 \fi
2545 \else\ifbbl@usedategroup
2546 \bbl@usedategroupfalse
2547 \ifhmode
2548 \hskip\z@skip % trick to ignore spaces

```

```

2549 \csname date#1\endcsname\relax
2550 \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
2551 \else
2552 \csname date#1\endcsname\relax
2553 \fi
2554 \fi\fi
2555 \bbl@usehooks{beforeextras}{}%
2556 \csname extras#1\endcsname\relax
2557 \bbl@usehooks{afterextras}{}%
2558 \ifcase\bbl@opt@hyphenmap\or
2559 \def\BabelLower##1##2{\lccode##1=##2\relax}%
2560 \ifnum\bbl@hymapsel>4\else
2561 \csname\language @bbl@hyphenmap\endcsname
2562 \fi
2563 \chardef\bbl@opt@hyphenmap\z@
2564 \else
2565 \ifnum\bbl@hymapsel>\bbl@opt@hyphenmap\else
2566 \csname\language @bbl@hyphenmap\endcsname
2567 \fi
2568 \fi
2569 \global\let\bbl@hymapsel\@cclv
2570 \bbl@patterns{#1}%
2571 \babel@savevariable\lefthyphenmin
2572 \babel@savevariable\righthyphenmin
2573 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2574 \set@hyphenmins\tw@\thr@\relax
2575 \else
2576 \expandafter\expandafter\expandafter\set@hyphenmins
2577 \csname #1hyphenmins\endcsname\relax
2578 \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.

```

2579 \long\def\otherlanguage#1{%
2580 \ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel\thr@\fi
2581 \csname selectlanguage \endcsname{#1}%
2582 \ignorespaces}

```

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

```

2583 \long\def\endotherlanguage{%
2584 \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`.

```

2585 \expandafter\def\csname otherlanguage*\endcsname#1{%
2586 \ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi
2587 \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”.

```

2588 \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.

```

2589 \providecommand\bbl@beforeforeign{}
2590 \edef\foreignlanguage{%
2591   \noexpand\protect
2592   \expandafter\noexpand\csname foreignlanguage \endcsname}
2593 \expandafter\def\csname foreignlanguage \endcsname{%
2594   \@ifstar\bbl@foreign@s\bbl@foreign@x}
2595 \def\bbl@foreign@x#1#2{%
2596   \begingroup
2597     \let\BabelText\@firstofone
2598     \bbl@beforeforeign
2599     \foreign@language{#1}%
2600     \bbl@usehooks{foreign}{}%
2601     \BabelText{#2}% Now in horizontal mode!
2602   \endgroup}
2603 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \setpar, ?\@@par
2604   \begingroup
2605     {\par}%
2606     \let\BabelText\@firstofone
2607     \foreign@language{#1}%
2608     \bbl@usehooks{foreign*}{}%
2609     \bbl@dirparastext
2610     \BabelText{#2}% Still in vertical mode!
2611     {\par}%
2612   \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`.

```

2613 \def\foreign@language#1{%
2614   \edef\language#1%
2615   \bbl@fixname\language
2616   \bbl@iflanguage\language%
2617     \expandafter\ifx\csname date\language\endcsname\relax
2618       \bbl@warning

```



```

2619      {Unknown language `#1'. Either you have\\%
2620      misspelled its name, it has not been installed,\\%
2621      or you requested it in a previous run. Fix its name,\\%
2622      install it or just rerun the file, respectively.\\%
2623      I'll proceed, but expect wrong results.\\%
2624      Reported}%
2625  \fi
2626  \let\bbl@select@type@ne
2627  \expandafter\bbl@switch\expandafter{\language}

```

\bbl@patterns This macro selects the hyphenation patterns by changing the `\language` register. If special hyphenation patterns are available specifically for the current font encoding, use them instead of the default.

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

```

2628 \let\bbl@hyphlist@empty
2629 \let\bbl@hyphenation@relax
2630 \let\bbl@pttnlist@empty
2631 \let\bbl@patterns@relax
2632 \let\bbl@hymapsel=\ccclv
2633 \def\bbl@patterns#1{%
2634   \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
2635     \csname l@#1\endcsname
2636     \edef\bbl@tempa{#1}%
2637   \else
2638     \csname l@#1:\f@encoding\endcsname
2639     \edef\bbl@tempa{#1:\f@encoding}%
2640   \fi
2641   \@expandtwoargs\bbl@usehooks{patterns}{#1}{\bbl@tempa}%
2642   \@ifundefined{bbl@hyphenation@}{#1}{% Can be \relax!
2643     \begingroup
2644       \bbl@xin@{,\number\language,}{,\bbl@hyphlist}%
2645       \ifin@else
2646         \@expandtwoargs\bbl@usehooks{hyphenation}{#1}{\bbl@tempa}%
2647         \hyphenation{
2648           \bbl@hyphenation@
2649           \@ifundefined{bbl@hyphenation@#1}{
2650             \@empty
2651             {\space\csname bbl@hyphenation@#1\endcsname}}%
2652         \xdef\bbl@hyphlist{\bbl@hyphlist\number\language,}%
2653       \fi
2654     \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*.

```

2655 \def\hyphenrules#1{%
2656   \edef\bbl@tempf{#1}%
2657   \bbl@fixname\bbl@tempf
2658   \bbl@iflanguage\bbl@tempf{%
2659     \expandafter\bbl@patterns\expandafter{\bbl@tempf}%
2660     \language\shorthands{none}%

```

```

2661 \expandafter\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
2662 \set@hyphenmins\tw@\thr@@\relax
2663 \else
2664 \expandafter\expandafter\expandafter\set@hyphenmins
2665 \csname\bbl@tempf hyphenmins\endcsname\relax
2666 \fi}}
2667 \let\endhyphenrules\@empty

```

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

```

2668 \def\providehyphenmins#1#2{%
2669 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2670 \@namedef{#1hyphenmins}{#2}%
2671 \fi}

```

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

```

2672 \def\set@hyphenmins#1#2{%
2673 \lefthyphenmin#1\relax
2674 \righthyphenmin#2\relax}

```

`\ProvidesLanguage` The identification code for each file is something that was introduced in L^AT_EX 2_ε. 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.

```

2675 \ifx\ProvidesFile\@undefined
2676 \def\ProvidesLanguage#1[#2 #3 #4]{%
2677 \wlog{Language: #1 #4 #3 <#2>}%
2678 }
2679 \else
2680 \def\ProvidesLanguage#1{%
2681 \begingroup
2682 \catcode`\ 10 %
2683 \@makeother\/%
2684 \ifnextchar[%]
2685 {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}
2686 \def\@provideslanguage#1[#2]{%
2687 \wlog{Language: #1 #2}%
2688 \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
2689 \endgroup}
2690 \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`

```

2691 \def\LdfInit{%
2692 \chardef\atcatcode=\catcode`\@
2693 \catcode`\@=11\relax
2694 \input babel.def\relax
2695 \catcode`\@=\atcatcode \let\atcatcode\relax
2696 \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`.

```
2697 \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.

```
2698 \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’:

```
2699 \providecommand\setlocale{%
2700   \bbl@error
2701   {Not yet available}%
2702   {Find an armchair, sit down and wait}}
2703 \let\uselocale\setlocale
2704 \let\locale\setlocale
2705 \let\selectlocale\setlocale
2706 \let\textlocale\setlocale
2707 \let\textlanguage\setlocale
2708 \let\languagetext\setlocale
```

11.2 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 $\LaTeX 2_{\epsilon}$, so we can safely use its error handling interface. Otherwise we’ll have to ‘keep it simple’.

```
2709 \edef\bbl@nulllanguage{\string\language=0}
2710 \ifx\PackageError\@undefined
2711   \def\bbl@error#1#2{%
2712     \begingroup
2713     \newlinechar=`^^J
2714     \def\{^^J(babel) }%
2715     \errhelp{#2}\errmessage{\{#1}%
2716     \endgroup}
2717   \def\bbl@warning#1{%
2718     \begingroup
2719     \newlinechar=`^^J
2720     \def\{^^J(babel) }%
2721     \message{\{#1}%
2722     \endgroup}
2723   \def\bbl@info#1{%
2724     \begingroup
2725     \newlinechar=`^^J
2726     \def\{^^J}%
2727     \wlog{#1}%
2728     \endgroup}
2729 \else
2730   \def\bbl@error#1#2{%
2731     \begingroup
```

```

2732     \def\{\MessageBreak}%
2733     \PackageError{babel}{#1}{#2}%
2734   \endgroup}
2735   \def\bbl@warning#1{%
2736     \begingroup
2737       \def\{\MessageBreak}%
2738       \PackageWarning{babel}{#1}%
2739     \endgroup}
2740   \def\bbl@info#1{%
2741     \begingroup
2742       \def\{\MessageBreak}%
2743       \PackageInfo{babel}{#1}%
2744     \endgroup}
2745 \fi
2746 \@ifpackagewith{babel}{silent}
2747   {\let\bbl@info@gobble
2748    \let\bbl@warning@gobble}
2749   {}
2750 \def\bbl@nocaption{\protect\bbl@nocaption@i}
2751 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
2752   \global\@namedef{#2}{\textbf{?#1?}}%
2753   \@nameuse{#2}%
2754   \bbl@warning{%
2755     \@backslashchar#2 not set. Please, define\\%
2756     it in the preamble with something like:\\%
2757     \string\renewcommand\@backslashchar#2{..}\\%
2758     Reported}}
2759 \def\@nolanerr#1{%
2760   \bbl@error
2761   {You haven't defined the language #1\space yet}%
2762   {Your command will be ignored, type <return> to proceed}}
2763 \def\@nopatterns#1{%
2764   \bbl@warning
2765   {No hyphenation patterns were preloaded for\\%
2766    the language `#1' into the format.\\%
2767    Please, configure your TeX system to add them and\\%
2768    rebuild the format. Now I will use the patterns\\%
2769    preloaded for \bbl@nulllanguage\space instead}}
2770 \let\bbl@usehooks\@gobbletwo
2771 </kernel>
2772 <*patterns>

```

12 Loading hyphenation patterns

The following code is meant to be read by \LaTeX 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 \TeX fills the register. There are some problems with this approach though.

- When someone wants to use several hyphenation patterns with \TeX the above scheme won't work. The reason is that \TeX overwrites the contents of the `\everyjob` register with its own message.
- Plain \TeX does not use the `\everyjob` register so the message would not be displayed.

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

To make sure that \TeX 2.09 executes the `\@begindocumenthook` we would want to alter `\begin{document}`, but as this done too often already, we add the new code at the front of `\@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.

```

2773 <<Make sure ProvidesFile is defined>>
2774 \ProvidesFile{hyphen.cfg}[\<date>] [\<version>] Babel hyphens]
2775 \xdef\bb1@format{\jobname}
2776 \ifx\AtBeginDocument\@undefined
2777   \def\@empty{}
2778   \let\orig@dump\dump
2779   \def\dump{%
2780     \ifx\@ztryfc\@undefined
2781       \else
2782         \toks0=\expandafter{\@preamblecmds}%
2783         \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
2784         \def\@begindocumenthook{}%
2785       \fi
2786       \let\dump\orig@dump\let\orig@dump\@undefined\dump}
2787 \fi
2788 <<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.

```

2789 \def\process@line#1#2 #3 #4 {%
2790   \ifx=#1%
2791     \process@synonym{#2}%
2792   \else
2793     \process@language{#1#2}{#3}{#4}%
2794   \fi
2795   \ignorespaces}

```

`\process@synonym` This macro takes care of the lines which start with an `=`. It needs an empty token register to begin with. `\bb1@languages` is also set to empty.

```

2796 \toks@{}
2797 \def\bb1@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.

```

2798 \def\process@synonym#1{%
2799   \ifnum\last@language=\m@ne
2800     \toks\expandafter{\the\toks@relax\process@synonym{#1}}%
2801   \else
2802     \expandafter\chardef\csname l@#1\endcsname\last@language
2803     \wlog{\string\l@#1=\string\language\the\last@language}%
2804     \expandafter\let\csname #1hyphenmins\expandafter\endcsname
2805       \csname\language\hyphenmins\endcsname
2806     \let\bbl@elt\relax
2807     \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}{}}%
2808   \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.

```

2809 \def\process@language#1#2#3{%
2810   \expandafter\addlanguage\csname l@#1\endcsname
2811   \expandafter\language\csname l@#1\endcsname
2812   \edef\language{#1}%
2813   \bbl@hook@everylanguage{#1}%
2814   \bbl@get@enc#1::@@@
2815   \begingroup
2816     \lefthyphenmin\m@ne
2817     \bbl@hook@loadpatterns{#2}%

```

```

2818 \ifnum\lefthyphenmin=\m@ne
2819 \else
2820 \expandafter\xdef\csname #1hyphenmins\endcsname{%
2821 \the\lefthyphenmin\the\righthyphenmin}%
2822 \fi
2823 \endgroup
2824 \def\bbl@tempa{#3}%
2825 \ifx\bbl@tempa\@empty\else
2826 \bbl@hook@loadexceptions{#3}%
2827 \fi
2828 \let\bbl@elt\relax
2829 \edef\bbl@languages{%
2830 \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
2831 \ifnum\the\language=\z@
2832 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2833 \set@hyphenmins\tw@\thr@@\relax
2834 \else
2835 \expandafter\expandafter\expandafter\set@hyphenmins
2836 \csname #1hyphenmins\endcsname
2837 \fi
2838 \the\toks@
2839 \toks@{}}%
2840 \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.

```

2841 \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.

```

2842 \def\bbl@hook@everylanguage#1{}
2843 \def\bbl@hook@loadpatterns#1{\input #1\relax}
2844 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
2845 \let\bbl@hook@loadkernel\bbl@hook@loadpatterns
2846 \begingroup
2847 \def\AddBabelHook#1#2{%
2848 \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
2849 \def\next{\toks1}%
2850 \else
2851 \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname####1}%
2852 \fi
2853 \next}
2854 \ifx\directlua\@undefined
2855 \ifx\XeTeXinputencoding\@undefined\else
2856 \input xebabel.def
2857 \fi
2858 \else
2859 \input luababel.def
2860 \fi
2861 \openin1 = babel-\bbl@format.cfg
2862 \ifeof1
2863 \else
2864 \input babel-\bbl@format.cfg\relax
2865 \fi
2866 \closein1
2867 \endgroup
2868 \bbl@hook@loadkernel{switch.def}

```

`\readconfigfile` The configuration file can now be opened for reading.

```
2869 \openin1 = language.dat
```

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

```
2870 \def\languagename{english}%
2871 \ifeof1
2872   \message{I couldn't find the file language.dat,\space
2873           I will try the file hyphen.tex}
2874   \input hyphen.tex\relax
2875   \chardef\l@english\z@
2876 \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`.

```
2877   \last@language\m@ne
```

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

```
2878   \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.

```
2879     \endlinechar\m@ne
2880     \read1 to \bbl@line
2881     \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.

```
2882     \if T\ifeof1F\fi T\relax
2883     \ifx\bbl@line\@empty\else
2884       \edef\bbl@line{\bbl@line\space\space\space}%
2885       \expandafter\process@line\bbl@line\relax
2886     \fi
2887   \repeat
```

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

```
2888   \begingroup
2889     \def\bbl@elt#1#2#3#4{%
2890       \global\language=#2\relax
2891       \gdef\languagename{#1}%
2892       \def\bbl@elt##1##2##3##4{}}%
2893   \bbl@languages
2894   \endgroup
2895 \fi
```

and close the configuration file.

```
2896 \closein1
```

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

```
2897 \if/\the\toks@/\else
2898   \errhelp{language.dat loads no language, only synonyms}
2899   \errmessage{Orphan language synonym}
2900 \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.

```

2901 \let\bbl@line\@undefined
2902 \let\process@line\@undefined
2903 \let\process@synonym\@undefined
2904 \let\process@language\@undefined
2905 \let\bbl@get@enc\@undefined
2906 \let\bbl@hyph@enc\@undefined
2907 \let\bbl@tempa\@undefined
2908 \let\bbl@hook@loadkernel\@undefined
2909 \let\bbl@hook@everylanguage\@undefined
2910 \let\bbl@hook@loadpatterns\@undefined
2911 \let\bbl@hook@loadexceptions\@undefined
2912 \</patterns>

```

Here the code for `iniTEX` ends.

13 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].

```

2913 <<(*More package options)>> ≡
2914 \ifodd\bbl@engine
2915   \DeclareOption{bidi=basic-r}%
2916     {\ExecuteOptions{bidi=basic}}
2917   \DeclareOption{bidi=basic}%
2918     {\let\bbl@beforeforeign\leavevmode
2919       \newattribute\bbl@attr@dir
2920       \bbl@exp{\output{\bodydir\pagedir\the\output}}}%
2921       \AtEndOfPackage{\EnableBabelHook{babel-bidi}}}
2922 \else
2923   \DeclareOption{bidi=basic-r}%
2924     {\ExecuteOptions{bidi=basic}}
2925   \DeclareOption{bidi=basic}%
2926     {\bbl@error
2927       {The bidi method 'basic' is available only in\%
2928        luatex. I'll continue with 'bidi=default', so\%
2929        expect wrong results}%
2930       {See the manual for further details.}%
2931       \let\bbl@beforeforeign\leavevmode
2932       \AtEndOfPackage{%
2933         \EnableBabelHook{babel-bidi}%
2934         \bbl@xebidipar}}
2935 \fi
2936 \DeclareOption{bidi=default}%
2937   {\let\bbl@beforeforeign\leavevmode
2938     \ifodd\bbl@engine
2939       \newattribute\bbl@attr@dir
2940       \bbl@exp{\output{\bodydir\pagedir\the\output}}}%
2941   \fi
2942   \AtEndOfPackage{%
2943     \EnableBabelHook{babel-bidi}%
2944     \ifodd\bbl@engine\else
2945       \bbl@xebidipar
2946     \fi}}

```

2947 <</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.

```

2948 <<*Font selection>> ≡
2949 \bbl@trace{Font handling with fontspec}
2950 \@onlypreamble\babelfont
2951 \newcommand\babelfont[2][{}]{% 1=langs/scripts 2=fam
2952   \edef\bbl@tempa{#1}%
2953   \def\bbl@tempb{#2}%
2954   \ifx\fontspec\undefined
2955     \usepackage{fontspec}%
2956   \fi
2957   \EnableBabelHook{babel-fontspec}%
2958   \bbl@bblfont}
2959 \newcommand\bbl@bblfont[2][{}]{% 1=features 2=fontname
2960   \bbl@ifunset{\bbl@tempb family}{\bbl@providfam{\bbl@tempb}}{}}%
2961   \bbl@ifunset{\bbl@lsys\language}{\bbl@provide@lsys{\language}}{}}%
2962   \expandafter\bbl@ifblank\expandafter{\bbl@tempa}%
2963   {\bbl@csarg\edef{\bbl@tempb dflt@}{<{#1}{#2}}% save bbl@rmdflt@
2964     \bbl@exp{%
2965       \let\<bbl@\bbl@tempb dflt@\language>\<bbl@\bbl@tempb dflt@>%
2966       \\\bbl@font@set\<bbl@\bbl@tempb dflt@\language>%
2967         \<bbl@tempb default>\<bbl@tempb family>}}%
2968     {\bbl@foreach\bbl@tempa{% ie bbl@rmdflt@lang / *scrt
2969       \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:

```

2970 \def\bbl@providfam#1{%
2971   \bbl@exp{%
2972     \\\newcommand\<#1default>{}% Just define it
2973     \\\bbl@add@list\\bbl@font@fams{#1}%
2974     \\\DeclareRobustCommand\<#1family>{%
2975       \\\not@math@alphabet\<#1family>\relax
2976       \\\fontfamily\<#1default>\selectfont}%
2977     \\\DeclareTextFontCommand{\<text#1>}{\<#1family>}}}

```

The following macro is activated when the hook babel-fontspec is enabled.

```

2978 \def\bbl@switchfont{%
2979   \bbl@ifunset{\bbl@lsys\language}{\bbl@provide@lsys{\language}}{}}%
2980   \bbl@exp{% eg Arabic -> arabic
2981     \lowercase{\edef\\bbl@tempa{\bbl@cs{sname\language}}}}%
2982   \bbl@foreach\bbl@font@fams{%
2983     \bbl@ifunset{\bbl@##1dflt\language}% (1) language?
2984     {\bbl@ifunset{\bbl@##1dflt*\bbl@tempa}% (2) from script?
2985       {\bbl@ifunset{\bbl@##1dflt@}% 2=F - (3) from generic?
2986         {}% 123=F - nothing!
2987         {\bbl@exp{% 3=T - from generic
2988           \global\let\<bbl@##1dflt\language>%
2989             \<bbl@##1dflt@>}}}%
2990         {\bbl@exp{% 2=T - from script
2991           \global\let\<bbl@##1dflt\language>%
2992             \<bbl@##1dflt*\bbl@tempa>}}}%
2993       {}% 1=T - language, already defined
2994   \def\bbl@tempa{%
2995     \bbl@warning{The current font is not a standard family:\\%
2996       \fontname\font\\%
2997       Script and Language are not applied. Consider defining a\\%
2998       new family with \string\babelfont. Reported}}}%

```

```

2999 \bbl@foreach\bbl@font@fams{%      don't gather with prev for
3000   \bbl@ifunset\bbl@##1dflt@\languagename}%
3001   {\bbl@cs{famrst@##1}%
3002    \global\bbl@csarg\let{famrst@##1}\relax}%
3003   {\bbl@exp{% order is relevant
3004    \\bbl@add\\originalTeX{%
3005     \\bbl@font@rst{\bbl@cs{##1dflt@\languagename}}%
3006     \<##1default>\<##1family>{##1}}%
3007     \\bbl@font@set\<bbl@##1dflt@\languagename>% the main part!
3008     \<##1default>\<##1family>}}}%
3009 \bbl@ifrestoring{}\bbl@tempa}%

```

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.

```

3010 \def\bbl@font@set#1#2#3{% eg \bbl@rmdflt@lang \rmdefault \rmfamily
3011   \bbl@xin@{<>}{#1}%
3012   \ifin@
3013     \bbl@exp{\\bbl@fontspec@set\\#1\expandafter\@gobbletwo#1}%
3014   \fi
3015   \bbl@exp{%
3016     \def\\#2{#1}%      eg, \rmdefault{\bbl@rmdflt@lang}
3017     \\bbl@ifsamestring{#2}{\f@family}{\\#3\let\\bbl@tempa\relax}{}}%
3018 \def\bbl@fontspec@set#1#2#3{% eg \bbl@rmdflt@lang fnt-opt fnt-nme
3019   \let\bbl@tempe\bbl@mapselect
3020   \let\bbl@mapselect\relax
3021   \bbl@exp{\<fontspec_set_family:Nnn>\\#1%
3022     {\bbl@cs{lsys@\languagename},#2}}{#3}%
3023   \let\bbl@mapselect\bbl@tempe
3024   \bbl@tglobal#1}%

```

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.

```

3025 \def\bbl@font@rst#1#2#3#4{%
3026   \bbl@csarg\def{famrst@#4}{\bbl@font@set{#1}#2#3}}

```

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

```

3027 \def\bbl@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 :-).

```

3028 \newcommand\babelFSstore[2][{}]{%
3029   \bbl@ifblank{#1}%
3030   {\bbl@csarg\def{sname@#2}{Latin}}%
3031   {\bbl@csarg\def{sname@#2}{#1}}%
3032   \bbl@provide@dirs{#2}%
3033   \bbl@csarg\ifnum{wdir@#2}>\z@
3034     \let\bbl@beforeforeign\leavevmode
3035     \EnableBabelHook{babel-bidi}%
3036   \fi
3037   \bbl@foreach{#2}{%
3038     \bbl@FSstore{##1}{rm}\rmdefault\bbl@save@rmdefault
3039     \bbl@FSstore{##1}{sf}\sfdefault\bbl@save@sfdefault
3040     \bbl@FSstore{##1}{tt}\ttdefault\bbl@save@ttdefault}}

```

```

3041 \def\bbl@FSstore#1#2#3#4{%
3042   \bbl@csarg\edef{#2default#1}{#3}%
3043   \expandafter\addto\csname extras#1\endcsname{%
3044     \let#4#3%
3045     \ifx#3\f@family
3046       \edef#3{\csname bbl@#2default#1\endcsname}%
3047       \fontfamily{#3}\selectfont
3048     \else
3049       \edef#3{\csname bbl@#2default#1\endcsname}%
3050       \fi}%
3051   \expandafter\addto\csname noextras#1\endcsname{%
3052     \ifx#3\f@family
3053       \fontfamily{#4}\selectfont
3054       \fi
3055     \let#3#4}}
3056 \let\bbl@langfeatures\@empty
3057 \def\babelFSfeatures{% make sure \fontspec is redefined once
3058   \let\bbl@ori@fontspec\fontspec
3059   \renewcommand\fontspec[1][{}%
3060     \bbl@ori@fontspec[\bbl@langfeatures##1]}
3061   \let\babelFSfeatures\bbl@FSfeatures
3062   \babelFSfeatures}
3063 \def\bbl@FSfeatures#1#2{%
3064   \expandafter\addto\csname extras#1\endcsname{%
3065     \babel@save\bbl@langfeatures
3066     \edef\bbl@langfeatures{#2,}}
3067 <</Font selection>>

```

14 Hooks for XeTeX and LuaTeX

14.1 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.

```

3068 <<{*Restore Unicode catcodes before loading patterns}>> ≡
3069 \begingroup
3070   % Reset chars "80-"C0 to category "other", no case mapping:
3071   \catcode\@=11 \count@=128
3072   \loop\ifnum\count@<192
3073     \global\uccode\count@=0 \global\lccode\count@=0
3074     \global\catcode\count@=12 \global\sffcode\count@=1000
3075     \advance\count@ by 1 \repeat
3076   % Other:
3077   \def\O ##1 {%
3078     \global\uccode"##1=0 \global\lccode"##1=0
3079     \global\catcode"##1=12 \global\sffcode"##1=1000 }%
3080   % Letter:
3081   \def\L ##1 ##2 ##3 {\global\catcode"##1=11
3082     \global\uccode"##1="##2
3083     \global\lccode"##1="##3
3084     % Uppercase letters have sffcode=999:
3085     \ifnum"##1="##3 \else \global\sffcode"##1=999 \fi }%
3086   % Letter without case mappings:

```

```

3087 \def\l ##1 {\L ##1 ##1 ##1 }%
3088 \l 00AA
3089 \L 00B5 039C 00B5
3090 \l 00BA
3091 \O 00D7
3092 \l 00DF
3093 \O 00F7
3094 \L 00FF 0178 00FF
3095 \endgroup
3096 \input #1\relax
3097 <</Restore Unicode catcodes before loading patterns>>

```

Some more common code.

```

3098 <<{*Footnote changes}>> ≡
3099 \bbl@trace{Bidi footnotes}
3100 \ifx\bbl@beforeforeign\leavevmode
3101 \def\bbl@footnote#1#2#3{%
3102 \@ifnextchar[%
3103 {\bbl@footnote@o{#1}{#2}{#3}}%
3104 {\bbl@footnote@x{#1}{#2}{#3}}}
3105 \def\bbl@footnote@x#1#2#3#4{%
3106 \bgroup
3107 \select@language@x{\bbl@main@language}%
3108 \bbl@fn@footnote{#2#1{\ignorespaces#4}#3}%
3109 \egroup}
3110 \def\bbl@footnote@o#1#2#3[#4]#5{%
3111 \bgroup
3112 \select@language@x{\bbl@main@language}%
3113 \bbl@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}%
3114 \egroup}
3115 \def\bbl@footnotetext#1#2#3{%
3116 \@ifnextchar[%
3117 {\bbl@footnotetext@o{#1}{#2}{#3}}%
3118 {\bbl@footnotetext@x{#1}{#2}{#3}}}
3119 \def\bbl@footnotetext@x#1#2#3#4{%
3120 \bgroup
3121 \select@language@x{\bbl@main@language}%
3122 \bbl@fn@footnotetext{#2#1{\ignorespaces#4}#3}%
3123 \egroup}
3124 \def\bbl@footnotetext@o#1#2#3[#4]#5{%
3125 \bgroup
3126 \select@language@x{\bbl@main@language}%
3127 \bbl@fn@footnotetext[#4]{#2#1{\ignorespaces#5}#3}%
3128 \egroup}
3129 \def\BabelFootnote#1#2#3#4{%
3130 \ifx\bbl@fn@footnote\undefined
3131 \let\bbl@fn@footnote\footnote
3132 \fi
3133 \ifx\bbl@fn@footnotetext\undefined
3134 \let\bbl@fn@footnotetext\footnotetext
3135 \fi
3136 \bbl@ifblank{#2}%
3137 {\def#1{\bbl@footnote{\@firstofone}{#3}{#4}}
3138 \@namedef{\bbl@stripslash#1text}%
3139 {\bbl@footnotetext{\@firstofone}{#3}{#4}}}%
3140 {\def#1{\bbl@exp{\bbl@footnote{\foreignlanguage{#2}}}{#3}{#4}}%
3141 \@namedef{\bbl@stripslash#1text}%
3142 {\bbl@exp{\bbl@footnotetext{\foreignlanguage{#2}}}{#3}{#4}}}%
3143 \fi

```

```
3144 <</Footnote changes>>
```

Now, the code.

```
3145 (*xetex)
3146 \def\BabelStringsDefault{unicode}
3147 \let\xebbl@stop\relax
3148 \AddBabelHook{xetex}{encodedcommands}{%
3149   \def\bbl@tempa{#1}%
3150   \ifx\bbl@tempa\@empty
3151     \XeTeXinputencoding"bytes"%
3152   \else
3153     \XeTeXinputencoding"#1"%
3154   \fi
3155   \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
3156 \AddBabelHook{xetex}{stopcommands}{%
3157   \xebbl@stop
3158   \let\xebbl@stop\relax}
3159 \AddBabelHook{xetex}{loadkernel}{%
3160 <<Restore Unicode catcodes before loading patterns>>}
3161 \ifx\DisableBabelHook\@undefined\endinput\fi
3162 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
3163 \DisableBabelHook{babel-fontspec}
3164 <<Font selection>>
3165 \input txtbabel.def
3166 </xetex>
```

14.2 Layout

In progress.

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

\bbl@startskip and \bbl@endskip are available to package authors. Thanks to the T_EX expansion mechanism the following constructs are valid: \adim\bbl@startskip, \advance\bbl@startskip\adim, \bbl@startskip\adim.

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

```
3167 (*texxet)
3168 \bbl@trace{Redefinitions for bidi layout}
3169 \def\bbl@sspre@caption{%
3170   \bbl@exp{\everyhbox{\bbl@textdir\bbl@cs{wdir}\bbl@main@language}}}}
3171 \ifx\bbl@opt@layout\@nnil\endinput\fi % No layout
3172 \def\bbl@startskip{\ifcase\bbl@thepardir\leftskip\else\rightskip\fi}
3173 \def\bbl@endskip{\ifcase\bbl@thepardir\rightskip\else\leftskip\fi}
3174 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
3175   \def\@hangfrom#1{%
3176     \setbox\@tempboxa\hbox{#1}%
3177     \hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
3178     \noindent\box\@tempboxa}
3179 \def\raggedright{%
3180   \let\@centercr
3181   \bbl@startskip\z@skip
3182   \@rightskip\@flushglue
```

```

3183 \bbl@endskip\@rightskip
3184 \parindent\z@
3185 \parfillskip\bbl@startskip}
3186 \def\raggedleft{%
3187 \let\\@centercr
3188 \bbl@startskip\@flushglue
3189 \bbl@endskip\z@skip
3190 \parindent\z@
3191 \parfillskip\bbl@endskip}
3192 \fi
3193 \IfBabelLayout{lists}
3194 {\def\list#1#2{%
3195 \ifnum \@listdepth >5\relax
3196 \@toodeep
3197 \else
3198 \global\advance\@listdepth\@ne
3199 \fi
3200 \rightmargin\z@
3201 \listparindent\z@
3202 \itemindent\z@
3203 \csname @list\romannumeral\the\@listdepth\endcsname
3204 \def\@itemlabel{#1}%
3205 \let\makelabel\@mklab
3206 \@nmbrrlistfalse
3207 #2\relax
3208 \@trivlist
3209 \parskip\parsep
3210 \parindent\listparindent
3211 \advance\linewidth-\rightmargin
3212 \advance\linewidth-\leftmargin
3213 \advance\@totalleftmargin
3214 \ifcase\bbl@thepardir\leftmargin\else\rightmargin\fi
3215 \parshape\@ne\@totalleftmargin\linewidth
3216 \ignorespaces}%
3217 \ifcase\bbl@engine
3218 \def\labelenumii{}\theenumii{%
3219 \def\p@enumiii{\p@enumii}\theenumii}%
3220 \fi
3221 \def@verbatim{%
3222 \trivlist \item\relax
3223 \if@minipage\else\vskip\parskip\fi
3224 \bbl@startskip\textwidth
3225 \advance\bbl@startskip-\linewidth
3226 \bbl@endskip\z@skip
3227 \parindent\z@
3228 \parfillskip\@flushglue
3229 \parskip\z@skip
3230 \@@par
3231 \language\l@nohyphenation
3232 \@tempswafalse
3233 \def\par{%
3234 \if@tempswa
3235 \leavevmode\null
3236 \@@par\penalty\interlinepenalty
3237 \else
3238 \@tempwattrue
3239 \ifhmode\@@par\penalty\interlinepenalty\fi
3240 \fi}%
3241 \let\do\@makeother \dospecials

```

```

3242 \obeylines \verbatim@font \@noligs
3243 \everypar\expandafter{\the\everypar\unpenalty}}
3244 {}
3245 \IfBabelLayout{contents}
3246 {\def\@dottedtocline#1#2#3#4#5{%
3247 \ifnum#1>\c@tocdepth\else
3248 \vskip \z@ \@plus.2\p@
3249 {\bbl@startskip#2\relax
3250 \bbl@endskip\@tocrmarg
3251 \parfillskip-\bbl@endskip
3252 \parindent#2\relax
3253 \@afterindenttrue
3254 \interlinepenalty\@M
3255 \leavevmode
3256 \@tempdima#3\relax
3257 \advance\bbl@startskip\@tempdima
3258 \null\nobreak\hskip-\bbl@startskip
3259 {#4}\nobreak
3260 \leaders\hbox{%
3261 $m@th\mkern\@dotsep mu\hbox{.}\mkern\@dotsep mu$}%
3262 \hfill\nobreak
3263 \hb@xt@\@pnumwidth{\hfil\normalfont\normalcolor#5}%
3264 \par}%
3265 \fi}}
3266 {}
3267 \IfBabelLayout{columns}
3268 {\def\@outputdblcol{%
3269 \if@firstcolumn
3270 \global\@firstcolumnfalse
3271 \global\setbox\@leftcolumn\copy\@outputbox
3272 \splitmaxdepth\maxdimen
3273 \vbadness\maxdimen
3274 \setbox\@outputbox\vbox{\unvbox\@outputbox\unskip}%
3275 \setbox\@outputbox\vsplit\@outputbox to\maxdimen
3276 \toks\@expandafter{\topmark}%
3277 \xdef\@firstcoltopmark{\the\toks@}%
3278 \toks\@expandafter{\splitfirstmark}%
3279 \xdef\@firstcolfirstmark{\the\toks@}%
3280 \ifx\@firstcolfirstmark\@empty
3281 \global\let\@setmarks\relax
3282 \else
3283 \gdef\@setmarks{%
3284 \let\firstmark\@firstcolfirstmark
3285 \let\topmark\@firstcoltopmark}%
3286 \fi
3287 \else
3288 \global\@firstcolumntrue
3289 \setbox\@outputbox\vbox{%
3290 \hb@xt@\textwidth{%
3291 \hskip\columnwidth
3292 \hfil
3293 {\normalcolor\vrule \@width\columnseprule}%
3294 \hfil
3295 \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
3296 \hskip-\textwidth
3297 \hb@xt@\columnwidth{\box\@outputbox \hss}%
3298 \hskip\columnsep
3299 \hskip\columnwidth}}%
3300 \@combinedblfloats

```



```

3301 \setmarks
3302 \@outputpage
3303 \begingroup
3304 \dblfloatplacement
3305 \@startdblcolumn
3306 \@whiles\if@fcolmade \fi{\@outputpage
3307 \startdblcolumn}%
3308 \endgroup
3309 \fi}}%
3310 {}
3311 <<Footnote changes>>
3312 \IfBabelLayout{footnotes}%
3313 {\BabelFootnote\footnote\language\{}}%
3314 \BabelFootnote\localfootnote\language\{}}%
3315 \BabelFootnote\mainfootnote\{}}{}
3316 {}

```

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.

```

3317 \IfBabelLayout{counters}%
3318 {\let\bbl@latinarabic=\@arabic
3319 \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
3320 \let\bbl@asciroman=\@roman
3321 \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
3322 \let\bbl@asciiRoman=\@Roman
3323 \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}}{}
3324 </texet>

```

14.3 LuaTeX

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

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

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

Other preloaded languages could be read twice, if they has 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.

```

3325 (*luatex)
3326 \ifx\AddBabelHook\@undefined
3327 \bbl@trace{Read language.dat}
3328 \begingroup
3329 \toks@{}
3330 \count@ \z@ % 0=start, 1=0th, 2=normal
3331 \def\bbl@process@line#1#2 #3 #4 {%
3332   \ifx=#1%
3333     \bbl@process@synonym{#2}%
3334   \else
3335     \bbl@process@language{#1#2}{#3}{#4}%
3336   \fi
3337   \ignorespaces}
3338 \def\bbl@manylang{%
3339   \ifnum\bbl@last>\@ne
3340     \bbl@info{Non-standard hyphenation setup}%
3341   \fi
3342   \let\bbl@manylang\relax}
3343 \def\bbl@process@language#1#2#3{%
3344   \ifcase\count@
3345     \@ifundefined{zth@#1}{\count@\tw@}{\count@\@ne}%
3346   \or
3347     \count@\tw@
3348   \fi
3349   \ifnum\count@=\tw@
3350     \expandafter\addlanguage\csname l@#1\endcsname
3351     \language\allocationnumber
3352     \chardef\bbl@last\allocationnumber
3353     \bbl@manylang
3354     \let\bbl@elt\relax
3355     \xdef\bbl@languages{%
3356       \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
3357   \fi
3358   \the\toks@
3359   \toks@{}}
3360 \def\bbl@process@synonym@aux#1#2{%
3361   \global\expandafter\chardef\csname l@#1\endcsname#2\relax
3362   \let\bbl@elt\relax
3363   \xdef\bbl@languages{%
3364     \bbl@languages\bbl@elt{#1}{#2}{}}}%
3365 \def\bbl@process@synonym#1{%
3366   \ifcase\count@
3367     \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
3368   \or
3369     \@ifundefined{zth@#1}{\bbl@process@synonym@aux{#1}{}}}%
3370   \else
3371     \bbl@process@synonym@aux{#1}{\the\bbl@last}%
3372   \fi}
3373 \ifx\bbl@languages\@undefined % Just a (sensible?) guess
3374   \chardef\l@english\z@
3375   \chardef\l@USenglish\z@
3376   \chardef\bbl@last\z@
3377   \global\@namedef{bbl@hyphendata@0}{\hyphen.tex}}
3378 \gdef\bbl@languages{%

```

```

3379 \bbl@elt{english}{0}{hyphen.tex}{}%
3380 \bbl@elt{USenglish}{0}{}{}
3381 \else
3382 \global\let\bbl@languages@format\bbl@languages
3383 \def\bbl@elt#1#2#3#4{% Remove all except language 0
3384 \ifnum#2>\z@ \else
3385 \noexpand\bbl@elt{#1}{#2}{#3}{#4}%
3386 \fi}%
3387 \xdef\bbl@languages{\bbl@languages}%
3388 \fi
3389 \def\bbl@elt#1#2#3#4{\@namedef{zth#1}{} } % Define flags
3390 \bbl@languages
3391 \openin1=language.dat
3392 \ifeof1
3393 \bbl@warning{I couldn't find language.dat. No additional\\%
3394 patterns loaded. Reported}%
3395 \else
3396 \loop
3397 \endlinechar\m@ne
3398 \read1 to \bbl@line
3399 \endlinechar\^^M
3400 \if T\ifeof1F\fi T\relax
3401 \ifx\bbl@line\@empty\else
3402 \edef\bbl@line{\bbl@line\space\space\space}%
3403 \expandafter\bbl@process@line\bbl@line\relax
3404 \fi
3405 \repeat
3406 \fi
3407 \endgroup
3408 \bbl@trace{Macros for reading patterns files}
3409 \def\bbl@get@enc#1:#2:#3\@@{\def\bbl@hyph@enc{#2}}
3410 \ifx\babelcatcodetablenum\@undefined
3411 \def\babelcatcodetablenum{5211}
3412 \fi
3413 \def\bbl@luapatterns#1#2{%
3414 \bbl@get@enc#1::\@@@
3415 \setbox\z@\hbox\bgroup
3416 \begingroup
3417 \ifx\catcodetable\@undefined
3418 \let\savecatcodetable\luatexsavecatcodetable
3419 \let\initcatcodetable\luatexinitcatcodetable
3420 \let\catcodetable\luatexcatcodetable
3421 \fi
3422 \savecatcodetable\babelcatcodetablenum\relax
3423 \initcatcodetable\numexpr\babelcatcodetablenum+1\relax
3424 \catcodetable\numexpr\babelcatcodetablenum+1\relax
3425 \catcode`\#=6 \catcode`\$=3 \catcode`\&=4 \catcode`\^=7
3426 \catcode`\_ =8 \catcode`\{=1 \catcode`\}=2 \catcode`\-=13
3427 \catcode`\@=11 \catcode`\^^I=10 \catcode`\^^J=12
3428 \catcode`\<=12 \catcode`\>=12 \catcode`\*=12 \catcode`\.=12
3429 \catcode`\-=12 \catcode`\/=12 \catcode`\[=12 \catcode`\]=12
3430 \catcode`\`=12 \catcode`\'=12 \catcode`\`=12
3431 \input #1\relax
3432 \catcodetable\babelcatcodetablenum\relax
3433 \endgroup
3434 \def\bbl@tempa{#2}%
3435 \ifx\bbl@tempa\@empty\else
3436 \input #2\relax
3437 \fi

```

```

3438 \egroup}%
3439 \def\bbl@patterns@lua#1{%
3440 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
3441 \csname l@#1\endcsname
3442 \edef\bbl@tempa{#1}%
3443 \else
3444 \csname l@#1:\f@encoding\endcsname
3445 \edef\bbl@tempa{#1:\f@encoding}%
3446 \fi\relax
3447 \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
3448 \@ifundefined{bbl@hyphendata@the\language}%
3449 {\def\bbl@elt##1##2##3##4{%
3450 \ifnum##2=\csname l@\bbl@tempa\endcsname % #2=spanish, dutch:OT1...
3451 \def\bbl@tempb{##3}%
3452 \ifx\bbl@tempb\empty\else % if not a synonymous
3453 \def\bbl@tempc{##3}{##4}%
3454 \fi
3455 \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
3456 \fi}%
3457 \bbl@languages
3458 \@ifundefined{bbl@hyphendata@the\language}%
3459 {\bbl@info{No hyphenation patterns were set for\%
3460 language '\bbl@tempa'. Reported}}%
3461 {\expandafter\expandafter\expandafter\bbl@luapatterns
3462 \csname bbl@hyphendata@the\language\endcsname}}}%
3463 \endinput\fi
3464 \begingroup
3465 \catcode`\%=12
3466 \catcode`\'=12
3467 \catcode`\%=12
3468 \catcode`\:=12
3469 \directlua{
3470 Babel = Babel or {}
3471 function Babel.bytes(line)
3472 return line:gsub(".",
3473 function (chr) return unicode.utf8.char(string.byte(chr)) end)
3474 end
3475 function Babel.begin_process_input()
3476 if luatexbase and luatexbase.add_to_callback then
3477 luatexbase.add_to_callback('process_input_buffer',
3478 Babel.bytes,'Babel.bytes')
3479 else
3480 Babel.callback = callback.find('process_input_buffer')
3481 callback.register('process_input_buffer',Babel.bytes)
3482 end
3483 end
3484 function Babel.end_process_input ()
3485 if luatexbase and luatexbase.remove_from_callback then
3486 luatexbase.remove_from_callback('process_input_buffer','Babel.bytes')
3487 else
3488 callback.unregister('process_input_buffer',Babel.callback)
3489 end
3490 end
3491 function Babel.addpatterns(pp, lg)
3492 local lg = lang.new(lg)
3493 local pats = lang.patterns(lg) or ''
3494 lang.clear_patterns(lg)
3495 for p in pp:gmatch('[^%s]+') do
3496 ss = ''

```

```

3497     for i in string.utfcharacters(p:gsub('%d', '')) do
3498         ss = ss .. '%d?' .. i
3499     end
3500     ss = ss:gsub('^%%d%?%', '%%.') .. '%d?'
3501     ss = ss:gsub('%.%%d%?$', '%%.')
3502     pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
3503     if n == 0 then
3504         tex.sprint(
3505             [[\string\csname\space bbl@info\endcsname{New pattern: }]]
3506             .. p .. [[]])
3507         pats = pats .. ' ' .. p
3508     else
3509         tex.sprint(
3510             [[\string\csname\space bbl@info\endcsname{Renew pattern: }]]
3511             .. p .. [[]])
3512     end
3513 end
3514 lang.patterns(lg, pats)
3515 end
3516 }
3517 \endgroup
3518 \def\BabelStringsDefault{unicode}
3519 \let\luabbl@stop\relax
3520 \AddBabelHook{luatex}{encodedcommands}{%
3521     \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
3522     \ifx\bbl@tempa\bbl@tempb\else
3523         \directlua{Babel.begin_process_input()}%
3524     \def\luabbl@stop{%
3525         \directlua{Babel.end_process_input()}}%
3526     \fi}%
3527 \AddBabelHook{luatex}{stopcommands}{%
3528     \luabbl@stop
3529     \let\luabbl@stop\relax}
3530 \AddBabelHook{luatex}{patterns}{%
3531     \@ifundefined{bbl@hyphendata@the\language}%
3532     {\def\bbl@elt##1##2##3##4{%
3533         \ifnum##2=\csname l@##2\endcsname % #2=spanish, dutch:OT1...
3534         \def\bbl@tempb{##3}%
3535         \ifx\bbl@tempb@empty\else % if not a synonymous
3536             \def\bbl@tempc{##3}{##4}%
3537         \fi
3538         \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
3539         \fi}%
3540     \bbl@languages
3541     \@ifundefined{bbl@hyphendata@the\language}%
3542     {\bbl@info{No hyphenation patterns were set for\%
3543         language '#2'. Reported}}%
3544     {\expandafter\expandafter\expandafter\bbl@luapatterns
3545         \csname bbl@hyphendata@the\language\endcsname}}}%
3546 \@ifundefined{bbl@patterns@}{}%
3547 \begingroup
3548     \bbl@xin@{,\number\language,}{,\bbl@pttnlist}%
3549     \ifin@else
3550         \ifx\bbl@patterns@empty\else
3551             \directlua{ Babel.addpatterns(
3552                 [[\bbl@patterns@]], \number\language) }%
3553         \fi
3554         \@ifundefined{bbl@patterns@#1}%
3555         \@empty

```

```

3556         {\directlua{ Babel.addpatterns(
3557             [[\space\csname bbl@patterns@#1\endcsname]],
3558             \number\language) }}}%
3559         \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
3560         \fi
3561     \endgroup}}
3562 \AddBabelHook{luatex}{everylanguage}{%
3563     \def\process@language##1##2##3{%
3564         \def\process@line####1####2 ####3 ####4 {}}}
3565 \AddBabelHook{luatex}{loadpatterns}{%
3566     \input #1\relax
3567     \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
3568         {{#1}}}}
3569 \AddBabelHook{luatex}{loadexceptions}{%
3570     \input #1\relax
3571     \def\bbl@tempb##1##2{{##1}{##2}}}%
3572     \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
3573         {\expandafter\expandafter\expandafter\bbl@tempb
3574             \csname bbl@hyphendata@the\language\endcsname}}

```

`\babelpatterns` This macro adds patterns. Two macros are used to store them: `\bbl@patterns@` for the global ones and `\bbl@patterns@<lang>` for language ones. We make sure there is a space between words when multiple commands are used.

```

3575 \@onlypreamble\babelpatterns
3576 \AtEndOfPackage{%
3577     \newcommand\babelpatterns[2][\@empty]{%
3578         \ifx\bbl@patterns@\relax
3579             \let\bbl@patterns@\@empty
3580         \fi
3581         \ifx\bbl@pttnlist\@empty\else
3582             \bbl@warning{%
3583                 You must not intermingle \string\selectlanguage\space and\\%
3584                 \string\babelpatterns\space or some patterns will not\\%
3585                 be taken into account. Reported}%
3586             \fi
3587             \ifx\@empty#1%
3588                 \protected@edef\bbl@patterns@{\bbl@patterns@\space#2}%
3589             \else
3590                 \edef\bbl@tempb{\zap@space#1 \@empty}%
3591                 \bbl@for\bbl@tempa\bbl@tempb{%
3592                     \bbl@fixname\bbl@tempa
3593                     \bbl@iflanguage\bbl@tempa{%
3594                         \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
3595                             \@ifundefined{bbl@patterns@\bbl@tempa}%
3596                             \@empty
3597                             {\csname bbl@patterns@\bbl@tempa\endcsname\space}%
3598                             #2}}}%
3599                 \fi}}

```

15 Southeast Asian scripts

```

3600 \def\bbl@seainterword{
3601     \let\bbl@seainterword\relax
3602     \directlua{
3603         Babel = Babel or {}
3604         function Babel.sea_disc_to_space (head)
3605             local useinterword = false

```

```

3606     local size = 10
3607     for item in node.traverse(head) do
3608         local i = item.id
3609         if i == node.id'glyph' then
3610             f = font.getfont(item.font)
3611             size = f.size    %% 10 pt = 655360 = 10 * 65536
3612             if (item.char > 0x0E00 and item.char < 0x0E7F) or    % Thai
3613                 (item.char > 0x0E80 and item.char < 0x0EFF) or    % Lao
3614                 (item.char > 0x1780 and item.char < 0x17FF) then % Khemer
3615                 useinterword = true
3616             else
3617                 useinterword = false
3618             end
3619         end
3620         if i == 7 and item.subtype == 3 and useinterword then
3621             local n = node.new(12, 13)
3622             node.setglue(n, \BabelInterword * size,
3623                         \BabelInterwordStretch * size,
3624                         \BabelInterwordShrink * size)
3625             node.insert_before(head, item, n)
3626             node.remove(head, item)
3627         end
3628     end
3629 end
3630 luatexbase.add_to_callback('hyphenate',
3631     function (head, tail)
3632         lang.hyphenate(head)
3633         Babel.sea_disc_to_space(head)
3634     end,
3635     'Babel.sea_disc_to_space')
3636 }}

```

Common stuff.

```

3637 \AddBabelHook{luatex}{loadkernel}{%
3638 <<Restore Unicode catcodes before loading patterns>>}}
3639 \ifx\DisableBabelHook\undefined\endinginput\fi
3640 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
3641 \DisableBabelHook{babel-fontspec}
3642 <<Font selection>>

```

15.1 Layout

Work in progress.

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

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

There are, however, a number of issues when the text direction is not the same as the box direction (as set by `\bodydir`), and when `\parbox` and `\hangindent` are involved.

Fortunately, latest releases of luatex simplify a lot the solution with `\shapemode`.

```

3643 \bbl@trace{Redefinitions for bidi layout}
3644 \ifx\@eqnnum\undefined\else
3645   \ifx\bbl@attr@dir\undefined\else
3646     \edef\@eqnnum{%
3647       \unexpanded{\ifcase\bbl@attr@dir\else\bbl@textdir\@ne\fi}%
3648       \unexpanded\expandafter{\@eqnnum}}
3649   \fi
3650 \fi

```

```

3651 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
3652 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
3653 \def\bbl@nextfake#1{%
3654   \mathdir\bodydir % non-local, use always inside a group!
3655   \bbl@exp{%
3656     #1%           Once entered in math, set boxes to restore values
3657     \everyvbox{%
3658       \the\everyvbox
3659       \bodydir\the\bodydir
3660       \mathdir\the\mathdir
3661       \everyhbox{\the\everyhbox}%
3662       \everyvbox{\the\everyvbox}}%
3663     \everyhbox{%
3664       \the\everyhbox
3665       \bodydir\the\bodydir
3666       \mathdir\the\mathdir
3667       \everyhbox{\the\everyhbox}%
3668       \everyvbox{\the\everyvbox}}}%
3669 \def\@hangfrom#1{%
3670   \setbox\@tempboxa\hbox{{#1}}%
3671   \hangindent\wd\@tempboxa
3672   \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
3673     \shapemode\@ne
3674   \fi
3675   \noindent\box\@tempboxa}
3676 \fi
3677 \IfBabelLayout{tabular}
3678 {\def\@tabular{%
3679   \leavevmode\hbox\bgroup\bbl@nextfake$%   %$
3680   \let\@acol\@tabacol   \let\@classz\@tabclassz
3681   \let\@classiv\@tabclassiv \let\@tabularcr\@tabarray}}
3682 {}
3683 \IfBabelLayout{lists}
3684 {\def\list#1#2{%
3685   \ifnum \@listdepth >5\relax
3686     \@toodeep
3687   \else
3688     \global\advance\@listdepth\@ne
3689   \fi
3690   \rightmargin\z@
3691   \listparindent\z@
3692   \itemindent\z@
3693   \csname @list\romannumeral\the\@listdepth\endcsname
3694   \def\@itemlabel{#1}%
3695   \let\makelabel\@mklab
3696   \@nmbrlistfalse
3697   #2\relax
3698   \@trivlist
3699   \parskip\parsep
3700   \parindent\listparindent
3701   \advance\linewidth -\rightmargin
3702   \advance\linewidth -\leftmargin
3703   \advance\totalleftmargin \leftmargin
3704   \parshape \@ne
3705   \@totalleftmargin \linewidth
3706   \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
3707     \shapemode\tw@
3708   \fi
3709   \ignorespaces}}

```



```
3710 {}
```

Implicitly reverses sectioning labels in bidi=basic-r, because the full stop is not in contact with L numbers any more. I think there must be a better way. Assumes bidi=basic-r, but there are some additional readjustments for bidi=default.

```
3711 \IfBabelLayout{counters}%
3712 {\def\@textsuperscript#1{% lua has separate settings for math
3713   \m@th
3714   \mathdir\pagedir % required with basic-r; ok with default, too
3715   \ensuremath{^{\mbox{\fontsize\sf@size\z@ #1}}}%
3716   \let\bbl@latinarabic=\@arabic
3717   \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
3718   \@ifpackagewith{babel}{bidi=default}%
3719   {\let\bbl@asciroman=\@roman
3720    \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
3721    \let\bbl@asciiRoman=\@Roman
3722    \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}%
3723    \def\labelenumii{}\theenumii}%
3724    \def\p@enumii{\p@enumii}\theenumii}}}%
3725 <<Footnote changes>>
3726 \IfBabelLayout{footnotes}%
3727 {\BabelFootnote\footnote\languagename{}{}}%
3728 \BabelFootnote\localfootnote\languagename{}{}}%
3729 \BabelFootnote\mainfootnote{}{}}%
3730 {}
```

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

```
3731 \IfBabelLayout{extras}%
3732 {\def\@underline#1{%
3733   \relax
3734   \ifmode\@underline{#1}%
3735   \else\bbl@nextfake$@@underline{\hbox{#1}}\m@th$\relax\fi}%
3736 \DeclareRobustCommand{\LaTeXe}{\mbox{\m@th
3737   \if b\expandafter\@car\@series\@nil\boldmath\fi
3738   \babelsublr{%
3739     \LaTeX\kern.15em2\bbl@nextfake$_{\textstyle\varepsilon}$}}}%
3740 {}
3741 </luatex>
```

15.2 Auto bidi with basic-r

The file babel-bidi.lua currently only contains data. It is a large and boring file and it's not shown here. See the generated file.

Now the basic-r bidi mode. One of the aims is to implement a fast and simple bidi algorithm, with a single loop. I managed to do it for R texts, with a second smaller loop for a special case. The code is still somewhat chaotic, but its behavior is essentially correct. I cannot resist copying the following text from Emacs bidi.c (which also attempts to implement the bidi algorithm with a single loop):

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.

TODO: math mode (as weak L?)

```

3742 (*basic-r)
3743 Babel = Babel or {}
3744
3745 require('babel-bidi.lua')
3746
3747 local characters = Babel.characters
3748 local ranges = Babel.ranges
3749
3750 local DIR = node.id("dir")
3751
3752 local function dir_mark(head, from, to, outer)
3753   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
3754   local d = node.new(DIR)
3755   d.dir = '+' .. dir
3756   node.insert_before(head, from, d)
3757   d = node.new(DIR)
3758   d.dir = '-' .. dir
3759   node.insert_after(head, to, d)
3760 end
3761
3762 function Babel.pre_otfload_v(head)
3763   -- head = Babel.numbers(head)
3764   head = Babel.bidi(head, true)
3765   return head
3766 end
3767
3768 function Babel.pre_otfload_h(head)
3769   -- head = Babel.numbers(head)
3770   head = Babel.bidi(head, false)
3771   return head
3772 end
3773
3774 function Babel.bidi(head, ispar)
3775   local first_n, last_n          -- first and last char with nums
3776   local last_es                  -- an auxiliary 'last' used with nums
3777   local first_d, last_d          -- first and last char in L/R block
3778   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):

```

3779   local strong = ('TRT' == tex.pardir) and 'r' or 'l'
3780   local strong_lr = (strong == 'l') and 'l' or 'r'

```

```

3781 local outer = strong
3782
3783 local new_dir = false
3784 local first_dir = false
3785
3786 local last_lr
3787
3788 local type_n = ''
3789
3790 for item in node.traverse(head) do
3791
3792   -- three cases: glyph, dir, otherwise
3793   if item.id == node.id'glyph'
3794     or (item.id == 7 and item.subtype == 2) then
3795
3796     local itemchar
3797     if item.id == 7 and item.subtype == 2 then
3798       itemchar = item.replace.char
3799     else
3800       itemchar = item.char
3801     end
3802     local chardata = characters[itemchar]
3803     dir = chardata and chardata.d or nil
3804     if not dir then
3805       for nn, et in ipairs(ranges) do
3806         if itemchar < et[1] then
3807           break
3808         elseif itemchar <= et[2] then
3809           dir = et[3]
3810           break
3811         end
3812       end
3813     end
3814     dir = dir or 'l'

```

Next is based on the assumption babel sets the language AND switches the script with its dir. We treat a language block as a separate Unicode sequence. The following piece of code is executed at the first glyph after a 'dir' node. We don't know the current language until then.

```

3815   if new_dir then
3816     attr_dir = 0
3817     for at in node.traverse(item.attr) do
3818       if at.number == luatexbase.registernumber'bbl@attr@dir' then
3819         attr_dir = at.value % 3
3820       end
3821     end
3822     if attr_dir == 1 then
3823       strong = 'r'
3824     elseif attr_dir == 2 then
3825       strong = 'al'
3826     else
3827       strong = 'l'
3828     end
3829     strong_lr = (strong == 'l') and 'l' or 'r'
3830     outer = strong_lr
3831     new_dir = false
3832   end
3833
3834   if dir == 'nsm' then dir = strong end

```

-- W1

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

```
3835     dir_real = dir          -- We need dir_real to set strong below
3836     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:

```
3837     if strong == 'al' then
3838         if dir == 'en' then dir = 'an' end          -- W2
3839         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
3840         strong_lr = 'r'                             -- W3
3841     end
```

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

```
3842     elseif item.id == node.id'dir' then
3843         new_dir = true
3844         dir = nil
3845     else
3846         dir = nil          -- Not a char
3847     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>.

```
3848     if dir == 'en' or dir == 'an' or dir == 'et' then
3849         if dir ~= 'et' then
3850             type_n = dir
3851         end
3852         first_n = first_n or item
3853         last_n = last_es or item
3854         last_es = nil
3855     elseif dir == 'es' and last_n then -- W3+W6
3856         last_es = item
3857     elseif dir == 'cs' then          -- it's right - do nothing
3858     elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
3859         if strong_lr == 'r' and type_n ~= '' then
3860             dir_mark(head, first_n, last_n, 'r')
3861         elseif strong_lr == 'l' and first_d and type_n == 'an' then
3862             dir_mark(head, first_n, last_n, 'r')
3863             dir_mark(head, first_d, last_d, outer)
3864             first_d, last_d = nil, nil
3865         elseif strong_lr == 'l' and type_n ~= '' then
3866             last_d = last_n
3867         end
3868         type_n = ''
3869         first_n, last_n = nil, nil
3870     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:

```
3871     if dir == 'l' or dir == 'r' then
3872         if dir ~= outer then
3873             first_d = first_d or item
```

```

3874     last_d = item
3875     elseif first_d and dir ~= strong_lr then
3876         dir_mark(head, first_d, last_d, outer)
3877         first_d, last_d = nil, nil
3878     end
3879 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.

```

3880     if dir and not last_lr and dir ~= 'l' and outer == 'r' then
3881         item.char = characters[item.char] and
3882             characters[item.char].m or item.char
3883     elseif (dir or new_dir) and last_lr ~= item then
3884         local mir = outer .. strong_lr .. (dir or outer)
3885         if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
3886             for ch in node.traverse(node.next(last_lr)) do
3887                 if ch == item then break end
3888                 if ch.id == node.id'glyph' then
3889                     ch.char = characters[ch.char].m or ch.char
3890                 end
3891             end
3892         end
3893     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).

```

3894     if dir == 'l' or dir == 'r' then
3895         last_lr = item
3896         strong = dir_real          -- Don't search back - best save now
3897         strong_lr = (strong == 'l') and 'l' or 'r'
3898     elseif new_dir then
3899         last_lr = nil
3900     end
3901 end

```

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

```

3902     if last_lr and outer == 'r' then
3903         for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
3904             ch.char = characters[ch.char].m or ch.char
3905         end
3906     end
3907     if first_n then
3908         dir_mark(head, first_n, last_n, outer)
3909     end
3910     if first_d then
3911         dir_mark(head, first_d, last_d, outer)
3912     end

```

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

```

3913     return node.prev(head) or head
3914 end
3915 </basic-r>

```

And here the Lua code for bidi=basic:

```

3916 <*basic>

```

```

3917 Babel = Babel or {}
3918
3919 Babel.fontmap = Babel.fontmap or {}
3920 Babel.fontmap[0] = {}      -- l
3921 Babel.fontmap[1] = {}      -- r
3922 Babel.fontmap[2] = {}      -- al/an
3923
3924 function Babel.pre_otfload_v(head)
3925     -- head = Babel.numbers(head)
3926     head = Babel.bidi(head, true)
3927     return head
3928 end
3929
3930 function Babel.pre_otfload_h(head, gc, sz, pt, dir)
3931     -- head = Babel.numbers(head)
3932     head = Babel.bidi(head, false, dir)
3933     return head
3934 end
3935
3936 require('babel-bidi.lua')
3937
3938 local characters = Babel.characters
3939 local ranges = Babel.ranges
3940
3941 local DIR = node.id('dir')
3942 local GLYPH = node.id('glyph')
3943
3944 local function insert_implicit(head, state, outer)
3945     local new_state = state
3946     if state.sim and state.eim and state.sim ~= state.eim then
3947         dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
3948         local d = node.new(DIR)
3949         d.dir = '+' .. dir
3950         node.insert_before(head, state.sim, d)
3951         local d = node.new(DIR)
3952         d.dir = '-' .. dir
3953         node.insert_after(head, state.eim, d)
3954     end
3955     new_state.sim, new_state.eim = nil, nil
3956     return head, new_state
3957 end
3958
3959 local function insert_numeric(head, state)
3960     local new
3961     local new_state = state
3962     if state.san and state.ean and state.san ~= state.ean then
3963         local d = node.new(DIR)
3964         d.dir = '+TLT'
3965         _, new = node.insert_before(head, state.san, d)
3966         if state.san == state.sim then state.sim = new end
3967         local d = node.new(DIR)
3968         d.dir = '-TLT'
3969         _, new = node.insert_after(head, state.ean, d)
3970         if state.ean == state.eim then state.eim = new end
3971     end
3972     new_state.san, new_state.ean = nil, nil
3973     return head, new_state
3974 end
3975

```

```

3976 -- \hbox with an explicit dir can lead to wrong results
3977 -- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>
3978
3979 function Babel.bidi(head, ispar, hdir)
3980   local d -- d is used mainly for computations in a loop
3981   local prev_d = ''
3982   local new_d = false
3983
3984   local nodes = {}
3985   local outer_first = nil
3986
3987   local glue_d = nil
3988   local glue_i = nil
3989
3990   local has_en = false
3991   local first_et = nil
3992
3993   local ATDIR = luatexbase.registernumber'bbl@attr@dir'
3994
3995   local save_outer
3996   local temp = node.get_attribute(head, ATDIR)
3997   if temp then
3998     temp = temp % 3
3999     save_outer = (temp == 0 and 'l') or
4000                  (temp == 1 and 'r') or
4001                  (temp == 2 and 'al')
4002   elseif ispar then -- Or error? Shouldn't happen
4003     save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
4004   else
4005     save_outer = ('TRT' == hdir) and 'r' or 'l'
4006   end
4007   local outer = save_outer
4008   local last = outer
4009   -- 'al' is only taken into account in the first, current loop
4010   if save_outer == 'al' then save_outer = 'r' end
4011
4012   local fontmap = Babel.fontmap
4013
4014   for item in node.traverse(head) do
4015
4016     -- In what follows, #node is the last (previous) node, because the
4017     -- current one is not added until we start processing the neutrals.
4018
4019     -- three cases: glyph, dir, otherwise
4020     if item.id == GLYPH
4021        or (item.id == 7 and item.subtype == 2) then
4022
4023       local d_font = nil
4024       local item_r
4025       if item.id == 7 and item.subtype == 2 then
4026         item_r = item.replace -- automatic discs have just 1 glyph
4027       else
4028         item_r = item
4029       end
4030       local chardata = characters[item_r.char]
4031       d = chardata and chardata.d or nil
4032       if not d or d == 'nsm' then
4033         for nn, et in ipairs(ranges) do
4034           if item_r.char < et[1] then

```

```

4035         break
4036     elseif item_r.char <= et[2] then
4037         if not d then d = et[3]
4038         elseif d == 'nsm' then d_font = et[3]
4039         end
4040         break
4041     end
4042 end
4043 end
4044 d = d or 'l'
4045 d_font = d_font or d
4046
4047 d_font = (d_font == 'l' and 0) or
4048         (d_font == 'nsm' and 0) or
4049         (d_font == 'r' and 1) or
4050         (d_font == 'al' and 2) or
4051         (d_font == 'an' and 2) or nil
4052 if d_font and fontmap and fontmap[d_font][item_r.font] then
4053     item_r.font = fontmap[d_font][item_r.font]
4054 end
4055
4056 if new_d then
4057     table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
4058     attr_d = node.get_attribute(item, ATDIR)
4059     attr_d = attr_d % 3
4060     if attr_d == 1 then
4061         outer_first = 'r'
4062         last = 'r'
4063     elseif attr_d == 2 then
4064         outer_first = 'r'
4065         last = 'al'
4066     else
4067         outer_first = 'l'
4068         last = 'l'
4069     end
4070     outer = last
4071     has_en = false
4072     first_et = nil
4073     new_d = false
4074 end
4075
4076 if glue_d then
4077     if (d == 'l' and 'l' or 'r') ~= glue_d then
4078         table.insert(nodes, {glue_i, 'on', nil})
4079     end
4080     glue_d = nil
4081     glue_i = nil
4082 end
4083
4084 elseif item.id == DIR then
4085     d = nil
4086     new_d = true
4087
4088 elseif item.id == node.id'glue' and item.subtype == 13 then
4089     glue_d = d
4090     glue_i = item
4091     d = nil
4092
4093 else

```



```

4094     d = nil
4095 end
4096
4097 -- AL <= EN/ET/ES      -- W2 + W3 + W6
4098 if last == 'al' and d == 'en' then
4099     d = 'an'           -- W3
4100 elseif last == 'al' and (d == 'et' or d == 'es') then
4101     d = 'on'           -- W6
4102 end
4103
4104 -- EN + CS/ES + EN      -- W4
4105 if d == 'en' and #nodes >= 2 then
4106     if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
4107         and nodes[#nodes-1][2] == 'en' then
4108         nodes[#nodes][2] = 'en'
4109     end
4110 end
4111
4112 -- AN + CS + AN         -- W4 too, because uax9 mixes both cases
4113 if d == 'an' and #nodes >= 2 then
4114     if (nodes[#nodes][2] == 'cs')
4115         and nodes[#nodes-1][2] == 'an' then
4116         nodes[#nodes][2] = 'an'
4117     end
4118 end
4119
4120 -- ET/EN                -- W5 + W7->1 / W6->on
4121 if d == 'et' then
4122     first_et = first_et or (#nodes + 1)
4123 elseif d == 'en' then
4124     has_en = true
4125     first_et = first_et or (#nodes + 1)
4126 elseif first_et then    -- d may be nil here !
4127     if has_en then
4128         if last == 'l' then
4129             temp = 'l'    -- W7
4130         else
4131             temp = 'en'   -- W5
4132         end
4133     else
4134         temp = 'on'       -- W6
4135     end
4136     for e = first_et, #nodes do
4137         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
4138     end
4139     first_et = nil
4140     has_en = false
4141 end
4142
4143 if d then
4144     if d == 'al' then
4145         d = 'r'
4146         last = 'al'
4147     elseif d == 'l' or d == 'r' then
4148         last = d
4149     end
4150     prev_d = d
4151     table.insert(nodes, {item, d, outer_first})
4152 end

```

```

4153
4154     outer_first = nil
4155
4156 end
4157
4158 -- TODO -- repeated here in case EN/ET is the last node. Find a
4159 -- better way of doing things:
4160 if first_et then      -- dir may be nil here !
4161     if has_en then
4162         if last == 'l' then
4163             temp = 'l'    -- W7
4164         else
4165             temp = 'en'    -- W5
4166         end
4167     else
4168         temp = 'on'        -- W6
4169     end
4170     for e = first_et, #nodes do
4171         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
4172     end
4173 end
4174
4175 -- dummy node, to close things
4176 table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
4177
4178 ----- NEUTRAL -----
4179
4180 outer = save_outer
4181 last = outer
4182
4183 local first_on = nil
4184
4185 for q = 1, #nodes do
4186     local item
4187
4188     local outer_first = nodes[q][3]
4189     outer = outer_first or outer
4190     last = outer_first or last
4191
4192     local d = nodes[q][2]
4193     if d == 'an' or d == 'en' then d = 'r' end
4194     if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
4195
4196     if d == 'on' then
4197         first_on = first_on or q
4198     elseif first_on then
4199         if last == d then
4200             temp = d
4201         else
4202             temp = outer
4203         end
4204         for r = first_on, q - 1 do
4205             nodes[r][2] = temp
4206             item = nodes[r][1]    -- MIRRORING
4207             if item.id == GLYPH and temp == 'r' then
4208                 item.char = characters[item.char].m or item.char
4209             end
4210         end
4211         first_on = nil

```

```

4212     end
4213
4214     if d == 'r' or d == 'l' then last = d end
4215 end
4216
4217 ----- IMPLICIT, REORDER -----
4218
4219 outer = save_outer
4220 last = outer
4221
4222 local state = {}
4223 state.has_r = false
4224
4225 for q = 1, #nodes do
4226
4227     local item = nodes[q][1]
4228
4229     outer = nodes[q][3] or outer
4230
4231     local d = nodes[q][2]
4232
4233     if d == 'nsm' then d = last end          -- W1
4234     if d == 'en' then d = 'an' end
4235     local isdir = (d == 'r' or d == 'l')
4236
4237     if outer == 'l' and d == 'an' then
4238         state.san = state.san or item
4239         state.ean = item
4240     elseif state.san then
4241         head, state = insert_numeric(head, state)
4242     end
4243
4244     if outer == 'l' then
4245         if d == 'an' or d == 'r' then      -- im -> implicit
4246             if d == 'r' then state.has_r = true end
4247             state.sim = state.sim or item
4248             state.eim = item
4249         elseif d == 'l' and state.sim and state.has_r then
4250             head, state = insert_implicit(head, state, outer)
4251         elseif d == 'l' then
4252             state.sim, state.eim, state.has_r = nil, nil, false
4253         end
4254     else
4255         if d == 'an' or d == 'l' then
4256             state.sim = state.sim or item
4257             state.eim = item
4258         elseif d == 'r' and state.sim then
4259             head, state = insert_implicit(head, state, outer)
4260         elseif d == 'r' then
4261             state.sim, state.eim = nil, nil
4262         end
4263     end
4264
4265     if isdir then
4266         last = d          -- Don't search back - best save now
4267     elseif d == 'on' and state.san then
4268         state.san = state.san or item
4269         state.ean = item
4270     end

```

```

4271
4272   end
4273
4274   return node.prev(head) or head
4275 end
4276 </basic>

```

16 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.

```

4277 <*nil>
4278 \ProvidesLanguage{nil}[<<date>> <<version>> Nil language]
4279 \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.

```

4280 \ifx\l@nohyphenation\@undefined
4281   \@nopatterns{nil}
4282   \adddialect\l@nil0
4283 \else
4284   \let\l@nil\l@nohyphenation
4285 \fi

```

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

```

4286 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}

```

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

```

\captionnil
\datenil
4287 \let\captionnil\@empty
4288 \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.

```

4289 \ldf@finish{nil}
4290 </nil>

```

17 Support for Plain \TeX (plain.def)

17.1 Not renaming hyphen.tex

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

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

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

The files `bplain.tex` and `lplain.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 \LaTeX , you will get a file called either `bplain.fmt` or `lplain.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 \LaTeX sees, we need to set some category codes just to be able to change the definition of `\input`

```
4291 (*bplain | bplain)
4292 \catcode`\{=1 % left brace is begin-group character
4293 \catcode`\}=2 % right brace is end-group character
4294 \catcode`\#=6 % hash mark is macro parameter character
```

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

```
4295 \openin 0 hyphen.cfg
```

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

```
4296 \ifeof0
4297 \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).

```
4298 \let\input
```

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

```
4299 \def\input #1 {%
4300   \let\input\input
4301   \input hyphen.cfg
```

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

```
4302 \let\input\undefined
4303 }
4304 \fi
4305 (/bplain | lplain)
```

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

```
4306 (bplain)\input plain.tex
4307 (lplain)\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.

```
4308 (bplain)\def\fmtname{babel-plain}
4309 (lplain)\def\fmtname{babel-lplain}
```

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

17.2 Emulating some \LaTeX features

The following code duplicates or emulates parts of $\text{\LaTeX} 2_{\epsilon}$ that are needed for `babel`.

```
4310 (*plain)
4311 \def\@empty{}
4312 \def\loadlocalcfg#1{%
```

```

4313 \openin0#1.cfg
4314 \ifeof0
4315 \closein0
4316 \else
4317 \closein0
4318 {\immediate\write16{*****}%
4319 \immediate\write16{* Local config file #1.cfg used}%
4320 \immediate\write16{*}%
4321 }
4322 \input #1.cfg\relax
4323 \fi
4324 \@endoflfd}

```

17.3 General tools

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

```

4325 \long\def\@firstofone#1{#1}
4326 \long\def\@firstoftwo#1#2{#1}
4327 \long\def\@secondoftwo#1#2{#2}
4328 \def\@nnil{\@nil}
4329 \def\@gobbletwo#1#2{}
4330 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
4331 \def\@star@or@long#1{%
4332 \@ifstar
4333 {\let\l@ngrel@x\relax#1}%
4334 {\let\l@ngrel@x\long#1}}
4335 \let\l@ngrel@x\relax
4336 \def\@car#1#2\@nil{#1}
4337 \def\@cdr#1#2\@nil{#2}
4338 \let\@typeset@protect\relax
4339 \let\protected@edef\edef
4340 \long\def\@gobble#1{}
4341 \edef\@backslashchar{\expandafter\@gobble\string\}
4342 \def\strip@prefix#1>{}
4343 \def\g@addto@macro#1#2{%
4344 \toks@{\expandafter{#1#2}}%
4345 \xdef#1{\the\toks@}}
4346 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
4347 \def\@nameuse#1{\csname #1\endcsname}
4348 \def\@ifundefined#1{%
4349 \expandafter\ifx\csname#1\endcsname\relax
4350 \expandafter\@firstoftwo
4351 \else
4352 \expandafter\@secondoftwo
4353 \fi}
4354 \def\@expandtwoargs#1#2#3{%
4355 \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
4356 \def\zap@space#1 #2{%
4357 #1%
4358 \ifx#2\@empty\else\expandafter\zap@space\fi
4359 #2}

```

\LaTeX 2_ε has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

4360 \ifx\@preamblecmds\@undefined
4361 \def\@preamblecmds{}
4362 \fi
4363 \def\@onlypreamble#1{%

```

```

4364 \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
4365   \@preamblecmds\do#1}}
4366 \@onlypreamble\@onlypreamble

```

Mimick \LaTeX 's `\AtBeginDocument`; for this to work the user needs to add `\begindocument` to his file.

```

4367 \def\begindocument{%
4368   \@begindocumenthook
4369   \global\let\@begindocumenthook\@undefined
4370   \def\do#1{\global\let##1\@undefined}%
4371   \@preamblecmds
4372   \global\let\do\noexpand}

4373 \ifx\@begindocumenthook\@undefined
4374   \def\@begindocumenthook{}
4375 \fi
4376 \@onlypreamble\@begindocumenthook
4377 \def\AtBeginDocument{\g@addto@macro\@begindocumenthook}

```

We also have to mimick \LaTeX 's `\AtEndOfPackage`. Our replacement macro is much simpler; it stores its argument in `\@endofldf`.

```

4378 \def\AtEndOfPackage#1{\g@addto@macro\@endofldf{#1}}
4379 \@onlypreamble\AtEndOfPackage
4380 \def\@endofldf{}
4381 \@onlypreamble\@endofldf
4382 \let\bbl@afterlang\@empty
4383 \chardef\bbl@opt@hyphenmap\z@

```

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

```

4384 \ifx\if@files\@undefined
4385   \expandafter\let\csname if@files\expandafter\endcsname
4386   \csname iffalse\endcsname
4387 \fi

```

Mimick \LaTeX 's commands to define control sequences.

```

4388 \def\newcommand{\@star@or@long\new@command}
4389 \def\new@command#1{%
4390   \@testopt{\@newcommand#1}0}
4391 \def\@newcommand#1[#2]{%
4392   \@ifnextchar [{\@xargdef#1[#2]}%
4393     {\@argdef#1[#2]}}
4394 \long\def\@argdef#1[#2]#3{%
4395   \@yargdef#1\@ne{#2}{#3}}
4396 \long\def\@xargdef#1[#2][#3]#4{%
4397   \expandafter\def\expandafter#1\expandafter{%
4398     \expandafter\@protected@testopt\expandafter #1%
4399     \csname\string#1\expandafter\endcsname{#3}}%
4400   \expandafter\@yargdef \csname\string#1\endcsname
4401   \tw@{#2}{#4}}
4402 \long\def\@yargdef#1#2#3{%
4403   \@tempcnta#3\relax
4404   \advance \@tempcnta \@ne
4405   \let\@hash@\relax
4406   \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
4407   \@tempcntb #2%
4408   \@whilenum\@tempcntb <\@tempcnta
4409   \do{%
4410     \edef\reserved@a{\reserved@a\@hash@\the\@tempcntb}%

```

```

4411 \advance\@tempcntb \@ne}%
4412 \let\@hash@##%
4413 \l@ngrel\x\expandafter\def\expandafter#1\reserved@a}
4414 \def\providecommand{\@star@or@long\provide@command}
4415 \def\provide@command#1{%
4416 \begingroup
4417 \escapechar\m@ne\xdef\@gtempa{\string#1}%
4418 \endgroup
4419 \expandafter\@ifundefined\@gtempa
4420 {\def\reserved@a{\new@command#1}}%
4421 {\let\reserved@a\relax
4422 \def\reserved@a{\new@command\reserved@a}}%
4423 \reserved@a}%

4424 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
4425 \def\declare@robustcommand#1{%
4426 \edef\reserved@a{\string#1}%
4427 \def\reserved@b{#1}%
4428 \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
4429 \edef#1{%
4430 \ifx\reserved@a\reserved@b
4431 \noexpand\x@protect
4432 \noexpand#1%
4433 \fi
4434 \noexpand\protect
4435 \expandafter\noexpand\csname
4436 \expandafter\@gobble\string#1 \endcsname
4437 }%
4438 \expandafter\new@command\csname
4439 \expandafter\@gobble\string#1 \endcsname
4440 }
4441 \def\x@protect#1{%
4442 \ifx\protect\@typeset@protect\else
4443 \@x@protect#1%
4444 \fi
4445 }
4446 \def\@x@protect#1\fi#2#3{%
4447 \fi\protect#1%
4448 }

```

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`.

```

4449 \def\bbl@tempa{\csname newif\endcsname\in@}
4450 \ifx\in@\@undefined
4451 \def\in@#1#2{%
4452 \def\in@##1#1##2##3\in@{%
4453 \ifx\in@##2\in@false\else\in@true\fi}%
4454 \in@#2#1\in@\in@}
4455 \else
4456 \let\bbl@tempa\@empty
4457 \fi
4458 \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).

```
4459 \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.

```
4460 \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.

```
4461 \ifx\@tempcnta\@undefined
4462   \csname newcount\endcsname\@tempcnta\relax
4463 \fi
4464 \ifx\@tempcntb\@undefined
4465   \csname newcount\endcsname\@tempcntb\relax
4466 \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`).

```
4467 \ifx\bye\@undefined
4468   \advance\count10 by -2\relax
4469 \fi
4470 \ifx\@ifnextchar\@undefined
4471   \def\@ifnextchar#1#2#3{%
4472     \let\reserved@d=#1%
4473     \def\reserved@a{#2}\def\reserved@b{#3}%
4474     \futurelet\@let@token\@ifnch}
4475   \def\@ifnch{%
4476     \ifx\@let@token\@sptoken
4477       \let\reserved@c\@xifnch
4478     \else
4479       \ifx\@let@token\reserved@d
4480         \let\reserved@c\reserved@a
4481       \else
4482         \let\reserved@c\reserved@b
4483       \fi
4484     \fi
4485     \reserved@c}
4486   \def\:{\let\@sptoken= } \: % this makes \@sptoken a space token
4487   \def\:{\@xifnch} \expandafter\def\:{\futurelet\@let@token\@ifnch}
4488 \fi
4489 \def\@testopt#1#2{%
4490   \@ifnextchar[#{1}{#1[#2]}}
4491 \def\@protected@testopt#1{%
4492   \ifx\protect\@typeset@protect
4493     \expandafter\@testopt
4494   \else
4495     \@x@protect#1%
4496   \fi}
4497 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
4498   #2\relax}\fi}
4499 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
4500   \else\expandafter\@gobble\fi{#1}}
```

17.4 Encoding related macros

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

```

4501 \def\DeclareTextCommand{%
4502   \@dec@text@cmd\providecommand
4503 }
4504 \def\ProvideTextCommand{%
4505   \@dec@text@cmd\providecommand
4506 }
4507 \def\DeclareTextSymbol#1#2#3{%
4508   \@dec@text@cmd\chardef#1{#2}#3\relax
4509 }
4510 \def\@dec@text@cmd#1#2#3{%
4511   \expandafter\def\expandafter#2%
4512     \expandafter{%
4513       \csname#3-cmd\expandafter\endcsname
4514       \expandafter#2%
4515       \csname#3\string#2\endcsname
4516     }%
4517 %   \let\@ifdefinable\rc@ifdefinable
4518   \expandafter#1\csname#3\string#2\endcsname
4519 }
4520 \def\@current@cmd#1{%
4521   \ifx\protect\@typeset@protect\else
4522     \noexpand#1\expandafter\@gobble
4523   \fi
4524 }
4525 \def\@changed@cmd#1#2{%
4526   \ifx\protect\@typeset@protect
4527     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
4528       \expandafter\ifx\csname ?\string#1\endcsname\relax
4529         \expandafter\def\csname ?\string#1\endcsname{%
4530           \@changed@x@err{#1}%
4531         }%
4532       \fi
4533       \global\expandafter\let
4534         \csname\cf@encoding \string#1\expandafter\endcsname
4535         \csname ?\string#1\endcsname
4536       \fi
4537       \csname\cf@encoding\string#1%
4538       \expandafter\endcsname
4539     \else
4540       \noexpand#1%
4541     \fi
4542 }
4543 \def\@changed@x@err#1{%
4544   \errhelp{Your command will be ignored, type <return> to proceed}%
4545   \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
4546 \def\DeclareTextCommandDefault#1{%
4547   \DeclareTextCommand#1?%
4548 }
4549 \def\ProvideTextCommandDefault#1{%
4550   \ProvideTextCommand#1?%
4551 }
4552 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
4553 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
4554 \def\DeclareTextAccent#1#2#3{%
4555   \DeclareTextCommand#1{#2}[1]{\accent#3 #1}
4556 }
4557 \def\DeclareTextCompositeCommand#1#2#3#4{%
4558   \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
4559   \edef\reserved@b{\string#1}%

```

```

4560 \edef\reserved@c{%
4561   \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
4562 \ifx\reserved@b\reserved@c
4563   \expandafter\expandafter\expandafter\ifx
4564     \expandafter\@car\reserved@a\relax\relax\@nil
4565     \@text@composite
4566   \else
4567     \edef\reserved@b##1{%
4568       \def\expandafter\noexpand
4569         \csname#2\string#1\endcsname###1{%
4570         \noexpand\@text@composite
4571         \expandafter\noexpand\csname#2\string#1\endcsname
4572         ###1\noexpand\@empty\noexpand\@text@composite
4573         {##1}%
4574       }%
4575     }%
4576     \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
4577   \fi
4578   \expandafter\def\csname\expandafter\string\csname
4579     #2\endcsname\string#1-\string#3\endcsname{#4}
4580 \else
4581   \errhelp{Your command will be ignored, type <return> to proceed}%
4582   \errmessage{\string\DeclareTextCompositeCommand\space used on
4583     inappropriate command \protect#1}
4584 \fi
4585 }
4586 \def\@text@composite#1#2#3\@text@composite{%
4587   \expandafter\@text@composite@x
4588     \csname\string#1-\string#2\endcsname
4589 }
4590 \def\@text@composite@x#1#2{%
4591   \ifx#1\relax
4592     #2%
4593   \else
4594     #1%
4595   \fi
4596 }
4597 %
4598 \def\@strip@args#1:#2-#3\@strip@args{#2}
4599 \def\DeclareTextComposite#1#2#3#4{%
4600   \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
4601   \bgroup
4602     \lccode\@=#4%
4603     \lowercase{%
4604   \egroup
4605     \reserved@a @%
4606   }%
4607 }
4608 %
4609 \def\UseTextSymbol#1#2{%
4610   \let\@curr@enc\cf@encoding
4611   \@use@text@encoding{#1}%
4612   #2%
4613   \@use@text@encoding\@curr@enc
4614 }
4615 \def\UseTextAccent#1#2#3{%
4616   \let\@curr@enc\cf@encoding
4617   \@use@text@encoding{#1}%
4618   #2{\@use@text@encoding\@curr@enc\selectfont#3}%

```

```

4619 % \use@text@encoding\@curr@enc
4620 }
4621 \def\use@text@encoding#1{%
4622 % \edef\font@encoding{#1}%
4623 % \xdef\font@name{%
4624 % \csname\curr@fontshape/\font@size\endcsname
4625 % }%
4626 % \pickup@font
4627 % \font@name
4628 % \@@enc@update
4629 }
4630 \def\DeclareTextSymbolDefault#1#2{%
4631 \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
4632 }
4633 \def\DeclareTextAccentDefault#1#2{%
4634 \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
4635 }
4636 \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.

```

4637 \DeclareTextAccent{"}{OT1}{127}
4638 \DeclareTextAccent{'}{OT1}{19}
4639 \DeclareTextAccent{^}{OT1}{94}
4640 \DeclareTextAccent{\`}{OT1}{18}
4641 \DeclareTextAccent{\~}{OT1}{126}

```

The following control sequences are used in `babel.def` but are not defined for `PLAIN TEX`.

```

4642 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
4643 \DeclareTextSymbol{\textquotedblright}{OT1}{\`"}
4644 \DeclareTextSymbol{\textquoteleft}{OT1}{\`'}
4645 \DeclareTextSymbol{\textquoteright}{OT1}{\`'}
4646 \DeclareTextSymbol{\i}{OT1}{16}
4647 \DeclareTextSymbol{\ss}{OT1}{25}

```

For a couple of languages we need the \LaTeX -control sequence `\scriptsize` to be available. Because plain $T_{\text{E}}X$ doesn't have such a sophisticated font mechanism as \LaTeX has, we just `\let` it to `\sevenrm`.

```

4648 \ifx\scriptsize\@undefined
4649 \let\scriptsize\sevenrm
4650 \fi
4651 \</plain>

```

18 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] Donald E. Knuth, *The T_EXbook*, Addison-Wesley, 1986.
- [3] Leslie Lamport, *L^AT_EX, A document preparation System*, Addison-Wesley, 1986.
- [4] K.F. Treebus. *Tekstwijzer, een gids voor het grafisch verwerken van tekst*. SDU Uitgeverij ('s-Gravenhage, 1988).
- [5] Hubert Partl, *German T_EX*, *TUGboat* 9 (1988) #1, p. 70–72.
- [6] Leslie Lamport, in: T_EXhax Digest, Volume 89, #13, 17 February 1989.
- [7] Johannes Braams, Victor Eijkhout and Nico Poppelier, *The development of national L^AT_EX styles*, *TUGboat* 10 (1989) #3, p. 401–406.
- [8] Yannis Haralambous, *Fonts & Encodings*, O'Reilly, 2007.
- [9] Joachim Schrod, *International L^AT_EX is ready to use*, *TUGboat* 11 (1990) #1, p. 87–90.
- [10] Apostolos Syropoulos, Antonis Tsolomitis and Nick Sofroniu, *Digital typography using L^AT_EX*, Springer, 2002, p. 301–373.