

Babel

Version 3.15

2017/11/03

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	7
1.7	Basic language selectors	8
1.8	Auxiliary language selectors	8
1.9	More on selection	9
1.10	Shorthands	10
1.11	Package options	14
1.12	The base option	16
1.13	ini files	16
1.14	Selecting fonts	22
1.15	Modifying a language	24
1.16	Creating a language	24
1.17	Getting the current language name	26
1.18	Hooks	26
1.19	Hyphenation tools	27
1.20	Selecting scripts	29
1.21	Selecting directions	29
1.22	Language attributes	31
1.23	Languages supported by babel	31
1.24	Tips, workarounds, know issues and notes	32
1.25	Current and future work	33
1.26	Tentative and experimental code	35
2	Loading languages with language.dat	35
2.1	Format	35
3	The interface between the core of babel and the language definition files	36
3.1	Basic macros	37
3.2	Skeleton	39
3.3	Support for active characters	39
3.4	Support for saving macro definitions	40
3.5	Support for extending macros	40
3.6	Macros common to a number of languages	40
3.7	Encoding-dependent strings	41
4	Changes	44
4.1	Changes in babel version 3.9	44
4.2	Changes in babel version 3.7	45
II	The code	46
5	Identification and loading of required files	46
6	Tools	46
6.1	Multiple languages	50

7	The Package File (\LaTeX, babel.sty)	51
7.1	base	51
7.2	key=value options and other general option	52
7.3	Conditional loading of shorthands	53
7.4	Language options	55
8	The kernel of Babel (babel.def, common)	57
8.1	Tools	58
8.2	Hooks	60
8.3	Setting up language files	61
8.4	Shorthands	63
8.5	Language attributes	72
8.6	Support for saving macro definitions	75
8.7	Short tags	75
8.8	Hyphens	76
8.9	Multiencoding strings	77
8.10	Macros common to a number of languages	83
8.11	Making glyphs available	83
8.11.1	Quotation marks	84
8.11.2	Letters	85
8.11.3	Shorthands for quotation marks	86
8.11.4	Umlauts and tremas	87
8.12	Layout	88
9	The kernel of Babel (babel.def, only \LaTeX)	89
9.1	The redefinition of the style commands	89
9.2	Creating languages	89
9.3	Cross referencing macros	95
9.4	Marks	98
9.5	Preventing clashes with other packages	99
9.5.1	ifthen	99
9.5.2	varioref	100
9.5.3	hhline	100
9.5.4	hyperref	101
9.5.5	fancyhdr	101
9.6	Encoding and fonts	102
9.7	Basic bidi support	103
9.8	Local Language Configuration	106
10	Multiple languages (switch.def)	106
10.1	Selecting the language	107
10.2	Errors	115
11	Loading hyphenation patterns	116
12	Font handling with fontspec	121
13	Hooks for XeTeX and LuaTeX	124
13.1	XeTeX	124
13.2	Layout	125
13.3	LuaTeX	128
13.4	Layout	133
13.5	Auto bidi with basic-r	134
14	The ‘nil’ language	138

15	Support for Plain T_EX (plain.def)	139
15.1	Not renaming hyphen.tex	139
15.2	Emulating some L ^A T _E X features	140
15.3	General tools	140
15.4	Encoding related macros	144
15.5	Babel options	147
16	Acknowledgements	147

Part I

User guide

This user guide focuses on \LaTeX . There are also some notes on its use with Plain \TeX . 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>. 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.

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:

```
\documentclass{article}

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

\usepackage[french]{babel}

\begin{document}

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

\end{document}
```

WARNING A common source of trouble is a wrong setting of the input encoding. 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.

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 \LaTeX , the preamble of the document:

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

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

You can also set the main language explicitly:

```
\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.

```
\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 behaviour 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 luatex. With them you can use babel to localize the documents.

The Latin script is covered by default in current L^AT_EX (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`.

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 `\babelfont` is used, described below).

```
\documentclass{article}

\usepackage[russian]{babel}

\babelfont{rm}{DejaVu Serif}

\begin{document}

Россия, находящаяся на пересечении множества культур, а также
```

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

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

```
! Package babel Error: Unknown language `LANG'. Either you have misspelled
(babel)                its name, it has not been installed, or you requested
(babel)                it in a previous run. Fix its name, install it or just
(babel)                rerun the file, respectively
```

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.

- 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 (MacT_EX, MikT_EX, T_EXLive, etc.) for further info about how to configure it.

1.6 Plain

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

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

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

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

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

```
{\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` $\langle language \rangle$ $\langle text \rangle$

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).

1.8 Auxiliary language selectors

`\begin{otherlanguage}` $\langle language \rangle$... `\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:

⁴Even in the babel kernel there were some macros not compatible with plain. Hopefully these issues will be fixed soon.

```

\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 behaviour and it is just a version as environment of `\foreignlanguage`.

`\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` {<tag1> = <language1>, <tag2> = <language2>, ...}

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

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

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

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

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

```
\babelensure{polish}
```

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

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

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

1.10 Shorthands

A *shorthand* is a sequence of one or two characters that expands to arbitrary \TeX code. Shorthands can be used for different kinds of things, 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 discretionaries and breaks can be inserted easily with "-", "=", etc. The package `inputenc` as well as `xetex` and `luatex` have alleviated entering non-ASCII characters, but minority languages and some kinds of text can still require characters not directly available on the keyboards (and sometimes not even as separated or precomposed Unicode characters). As to the point 2, now `pdfTeX` provides `\knbccode`, and `luatex` can manipulate the glyph list. Tools for point 3 can be still very useful in general.

⁵With it encoded string may not work as expected.

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

It is sometimes necessary to switch a shorthand character off temporarily, because it must be used in an entirely different way. For this purpose, the user commands `\shorthandoff` and `\shorthandon` are provided. They each take a list of characters as their arguments. The command `\shorthandoff` sets the `\catcode` for each of the characters in its argument to other (12); the command `\shorthandon` sets the `\catcode` to active (13). Both commands only work on ‘known’ shorthand characters. If a character is not known to be a shorthand character its category code will be left unchanged.

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.

`\usesshorthands` `*{\char}`

The command `\usesshorthands` 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 `\usesshorthands*{\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 `\usesshorthands`. This restriction will be lifted in a future release.

\defineshorthand [*<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 discretionaries (languages do not define shorthands consistently, and “-”, “\-”, “=” have different meanings). You could start with, say:

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

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

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

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

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

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

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

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

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.

1.11 Package options

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

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

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

activegrave Same for `.

shorthands= $\langle char \rangle \langle char \rangle \dots$ | 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 spoilt by an unexpected .cfg file. However, if the key config is set, this file is loaded.

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

nocase **New 3.9l** Language settings for uppercase and lowercase mapping (as set by `\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 T_EX, 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 L^AT_EX tools, so use it only as a last resort).

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

New 3.9g Sets the behaviour 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=

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

layout=

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

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 behaviour it might be added. On the other hand, `other` is provided even if I [JBL] think it isn't really useful, but who knows.

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

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

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

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

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

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

1.13 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.

EXAMPLE Although Georgian has its own `ldf` file, here is how to declare this language with an `ini` file in Unicode engines. The `nil` language is required, because currently `babel` raises an error if there is no language.

```
\documentclass{book}

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

\babelfont{rm}{DejaVu Sans}

\begin{document}

\tableofcontents
```

\chapter{სამმართველო და სუფრის ტრადიციები}

ქართული ტრადიციული სამმართველო ერთ-ერთი უმდიდრესია მთელ მსოფლიოში.

\end{document}

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

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

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

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

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

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

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

luxembourgish	punjabi-gurmukhi
luyia	punjabi-guru
macedonian	punjabi
machame	quechua
makhuwameetto	romanian
makonde	romansh
malagasy	rombo
malay-bn	rundi
malay-brunei	russian
malay-sg	rwa
malay-singapore	sakha
malay	samburu
malayalam	samin
maltese	sango
manx	sangu
marathi	scottishgaelic
masai	sena
mazanderani	serbian-cyrillic-bosniaherzegovina
meru	serbian-cyrillic-kosovo
meta	serbian-cyrillic-montenegro
mexican	serbian-cyrillic
mongolian	serbian-cyrl-ba
morisyen	serbian-cyrl-me
mundang	serbian-cyrl-xk
nama	serbian-cyrl
nepali	serbian-latin-bosniaherzegovina
newzealand	serbian-latin-kosovo
ngiemboon	serbian-latin-montenegro
ngomba	serbian-latin
norsk	serbian-latn-ba
northernluri	serbian-latn-me
northernsami	serbian-latn-xk
northndebele	serbian-latn
norwegianbokmal	serbian
norwegiannynorsk	shambala
nswissgerman	shona
nuer	sichuanyi
nyankole	sinhala
nynorsk	slovak
occitan	slovene
oriya	slovenian
oromo	soga
ossetic	somali
pashto	spanish-mexico
persian	spanish-mx
piedmontese	spanish
polish	standardmoroccantamazight
portuguese-br	swahili
portuguese-brazil	swedish
portuguese-portugal	swissgerman
portuguese-pt	tachelhit-latin
portuguese	tachelhit-latn
punjabi-arab	tachelhit-tfng
punjabi-arabic	tachelhit-tifinagh

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

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`.

`\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}

\babelprovide[import=he]{hebrew}

\babelfont{rm}{FreeSerif}

\begin{document}

Svenska \foreignlanguage{hebrew}{עִבְרִית} 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`.

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 behaviour.

WARNING Do not use `\setxxxxfont` and `\babelfont` at the same time. `\babelfont` follows the standard \LaTeX 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 behaviour of a language (say, the chapter “caption”), is sometimes necessary, but not always trivial.

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

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

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

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

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

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

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

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

NOTE 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\maylangchaptername{..}  
(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.

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`, there is a `\<language>date` macro with three arguments: year, month and day numbers. In fact, `\today` calls `\<language>today`, which in turn calls `\<language>date{\the\year}{\the\month}{\the\day}`.

captions= *<language-tag>*

Loads only the strings. For example:

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

hyphenrules= *<language-list>*

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

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

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

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

```

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

```

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

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

script= $\langle script-name \rangle$

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.

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 `\useshortands` and `\defineshortand` as described above. (2) Captions and `\today` are “ensured” with `\babelensure` (this is the default in ini-based languages).

1.17 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 $\{\langle language \rangle\}\{\langle true \rangle\}\{\langle false \rangle\}$

If more than one language is used, it might be necessary to know which language is active at a specific time. This can be checked by a call to `\iflanguage`, but note here “language” is used in the T_EX 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.

1.18 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:

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

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

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

defaultcommands Used (locally) in `\StartBabelCommands`.

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

stopcommands Used to reset the the above, if necessary.

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

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

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

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

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

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

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

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

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

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

loadkernel (file) By default loads `switch.def`. It can be used to load a different version of this 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).

1.19 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 behaviour 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, it 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 oportunity 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` [`<language>`, `<language>`, ...]{`<exceptions>`}

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

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

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

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

New 3.9m *In luatex only*,¹³ 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.

¹³With *luatex* exceptions and patterns can be modified almost freely. However, this is very likely a task for a separate package and `babel` only provides the most basic tools.

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

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

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.¹⁴

Some languages sharing the same script define macros to switch it (eg, `\textcyrillic`), but be aware they may also set the language to a certain default. Even the babel core defined `\textlatin`, but it was somewhat buggy because in some cases it messed up encodings and fonts (for example, if the main latin encoding was LY1), and therefore it has been deprecated.¹⁵

`\ensureascii` `{⟨text⟩}`

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

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

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 Setting bidi text has many subtleties (see for example

<https://www.w3.org/TR/html-bidi/>). This means the babel bidi code may take some time before it is truly stable. 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).

¹⁴The 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. As to directionality, it poses special challenges because it also affects individual characters and layout elements.

¹⁵But still defined for backwards compatibility.

There are some package options controlling bidi writing.

bidi= default | 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.

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, which will be improved in the future. Remember basic-r is available in luatex only.

```
\documentclass{article}

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

\babelprovide[import=ar, main]{arabic}

\babelfont{rm}{FreeSerif}

\begin{document}

    وقد عرفت شبه جزيرة العرب طيلة العصر الهيليني (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}
```

layout= sectioning | counters | lists | contents | columns

New 3.16 Selects with layout elements are adapted in bidi documents. This list will be expanded in future releases. *To be expanded.*

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

counters required in all engines to reorder correctly section numbers and the like (eg, $\langle subsection \rangle . \langle section \rangle$); required in xetex and pdftex for counters in general, as well as in luatex with bidi=default;

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

contents required in xetex and pdftex, but only in multilingual documents in luatex;

columns required in xetex and pdftex to reverse the column order (standard two column mode).

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

However, digits in pdftex must be marked up explicitly (unlike luatex with bidi=basic-r and, usually, xetex). Mainly for it (although available in all engines, because it can be useful), this command is provided to set $\{\langle lr-text \rangle\}$ in L mode. It's mainly intended for what Unicode calls weak characters, because words are best set with the corresponding language. For this reason, there is not a r1 counterpart.

\BabelPatchSection {<section-name>}

At work – 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, while the section text is still the current language. The latter is passed to tocs and marks, too. It can be set individually or to all standard sectioning commands. TODO: In the latter case, `\markboth` and `\markright` are not redefined, because the text passed includes already the language selector. TODO: `layout=nomarks` ?

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 Languages supported by babel

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

Afrikaans	afrikaans
Azerbaijani	azerbaijani
Basque	basque
Breton	breton
Bulgarian	bulgarian
Catalan	catalan
Croatian	croatian
Czech	czech
Danish	danish
Dutch	dutch
English	english, USenglish, american, UKenglish, british, canadian, australian, newzealand
Esperanto	esperanto
Estonian	estonian
Finnish	finnish
French	french, francais, canadien, acadian
Galician	galician
German	austrian, german, germanb, ngerman, naustrian
Greek	greek, polutonikogreek
Hebrew	hebrew
Icelandic	icelandic
Indonesian	bahasa, indonesian, indon, bahasai
Interlingua	interlingua
Irish Gaelic	irish
Italian	italian
Latin	latin

Lower Sorbian lowersorbian
Malay bahasam, malay, melayu
North Sami samin
Norwegian norsk, nynorsk
Polish polish
Portuguese portuges, portuguese, brazilian, brazil
Romanian romanian
Russian russian
Scottish Gaelic scottish
Spanish spanish
Slovakian slovak
Slovenian slovene
Swedish swedish
Serbian serbian
Turkish turkish
Ukrainian ukrainian
Upper Sorbian uppersorbian
Welsh welsh

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

Most of them work out of the box, but some may require extra fonts, encoding files, a preprocessor or even a complete framework (like CJK). 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 .

1.24 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\extrasrussian{\inputencoding{koi8-r}}
```

(A recent version of inputenc is required.)

- For the hyphenation to work correctly, lccodes cannot change, because T_EX only takes into account the values when the paragraph is hyphenated, i.e., when it has been finished.¹⁶ So, if you write a chunk of French text with `\foreignlanguage`, the apostrophes might not be taken into account. This is a limitation of T_EX, not of babel. Alternatively, you may use `\useshortands` 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 T_EX 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.

mkipattern Generates hyphenation patterns.

tracklang Tracks which languages have been requested.

1.25 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

¹⁶This explains why L^AT_EX assumes the lowercase mapping of T1 and does not provide a tool for multiple mappings. Unfortunately, `\savingshyphcodes` is not a solution either, because lccodes for hyphenation are frozen in the format and cannot be changed.

the roadmap are R layouts (lists, footnotes, tables, column order), page and section numbering, and maybe kashida justification.

As to Thai line breaking, here is the basic idea of what luatex can do for us, with the Thai patterns and a little script (the final version will not be so little, of course). It replaces each discretionary by the equivalent to ZWJ.

```
\documentclass{article}

\usepackage[nil]{babel}

\babelprovide[import=th, main]{thai}

\babelfont{rm}{FreeSerif}

\directlua{
local GLYPH = node.id'glyph'
function insertsp (head)
  local size = 0
  for item in node.traverse(head) do
    local i = item.id
    if i == GLYPH then
      f = font.getfont(item.font)
      size = f.size
    elseif i == 7 then
      local n = node.new(12, 0)
      node.setglue(n, 0, size * 1) % 1 is a factor
      node.insert_before(head, item, n)
      node.remove(head, item)
    end
  end
end
end

luatexbase.add_to_callback('hyphenate',
  function (head, tail)
    lang.hyphenate(head)
    insertsp(head)
  end, 'insertsp')
}

\begin{document}

(Thai text.)

\end{document}
```

Useful additions would be, for example, time, currency, addresses and personal names.¹⁷. But that is the easy part, because they don't require modifying the L^AT_EX 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)-ból”, in Spanish an item labelled “3.^o” may be referred to as either “ítem 3.^o” or “3.^{er} ítem”, and so on.

¹⁷See for example POSIX, ISO 14652 and the Unicode Common Locale Data Repository (CLDR). Those system, however, have limited application to T_EX because their aim is just to display information and not fine typesetting.

1.26 Tentative and experimental code

Handling of “**Unicode**” fonts is problematic. There is fontspec, but special macros are required (not only the NFSS ones) and it doesn’t provide “orthogonal axis” for features, including those related to the language (mainly language and script). A couple of tentative macros, were provided by babel ($\geq 3.9g$) with a partial solution. These macros are now deprecated – use `\babelfont`.

- `\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 is taking its *first steps*. *First steps* means exactly that. For example, in luatex any Arabic text must be marked up explicitly in L mode. On the other hand, xetex poses quite different challenges. Document layout (lists, footnotes, etc.) is not touched at all. 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 T_EX.

2 Loading languages with language.dat

T_EX and most engines based on it (pdfT_EX, xetex, ϵ -T_EX, the main exception being luatex) require hyphenation patterns to be preloaded when a format is created (eg, L^AT_EX, XeL^AT_EX, pdfL^AT_EX). 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).¹⁸ 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).¹⁹

2.1 Format

In that file the person who maintains a T_EX environment has to record for which languages he has hyphenation patterns *and* in which files these are stored²⁰. When hyphenation exceptions are stored in a separate file this can be indicated by naming that file *after* the file with the hyphenation patterns.

¹⁸This feature was added to 3.9o, but it was buggy. Both 3.9o and 3.9p are deprecated.

¹⁹The 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`.

²⁰This is because different operating systems sometimes use very different file-naming conventions.

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.²¹ 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
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.

²¹This is not a new feature, but in former versions it didn't work correctly.

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

Some recommendations:

- The preferred shorthand is `"`, which is not used in \LaTeX (quotes are entered as ``` and `'`). Other good choices are characters which are not used in a certain context (eg, `=` in an ancient language). Note however `=`, `<`, `>`, `:` and the like can be dangerous, because they may be used as part of the syntax of some elements (numeric expressions, key/value pairs, etc.).
- Captions should not contain shorthands or encoding dependent commands (the latter is not always possible, but should be clearly documented). They should be defined using the LICR. You may also use the new tools for encoded strings, described below.
- Avoid adding things to `\noextras⟨lang⟩` except for `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.²²
- 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 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

²²But not removed, for backward compatibility.

	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 T _E X sense of set of hyphenation patterns.
<code>\adddialect</code>	The macro <code>\adddialect</code> 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 behaviour of the babel system is to define this language as a ‘dialect’ of the language for which the patterns were loaded as <code>\language0</code> . Here “language” is used in the T _E X sense of set of hyphenation patterns.
<code>\<lang>hyphenmins</code>	The macro <code>\<lang>hyphenmins</code> is used to store the values of the <code>\lefthyphenmin</code> and <code>\righthyphenmin</code> . Redefine this macro to set your own values, with two numbers corresponding to these two parameters. For example:
	<pre>\renewcommand\spanishhyphenmins{34}</pre>
	(Assigning <code>\lefthyphenmin</code> and <code>\righthyphenmin</code> directly in <code>\extras<lang></code> has no effect.)
<code>\providehyphenmins</code>	The macro <code>\providehyphenmins</code> should be used in the language definition files to set <code>\lefthyphenmin</code> and <code>\righthyphenmin</code> . 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 <i>not</i> set them).
<code>\captions<lang></code>	The macro <code>\captions<lang></code> defines the macros that hold the texts to replace the original hard-wired texts.
<code>\date<lang></code>	The macro <code>\date<lang></code> defines <code>\today</code> .
<code>\extras<lang></code>	The macro <code>\extras<lang></code> 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.
<code>\noextras<lang></code>	Because we want to let the user switch between languages, but we do not know what state T _E X might be in after the execution of <code>\extras<lang></code> , a macro that brings T _E X into a predefined state is needed. It will be no surprise that the name of this macro is <code>\noextras<lang></code> .
<code>\bbl@declare@ttribute</code>	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.
<code>\main@language</code>	To postpone the activation of the definitions needed for a language until the beginning of a document, all language definition files should use <code>\main@language</code> instead of <code>\selectlanguage</code> . This will just store the name of the language, and the proper language will be activated at the start of the document.
<code>\ProvidesLanguage</code>	The macro <code>\ProvidesLanguage</code> should be used to identify the language definition files. Its syntax is similar to the syntax of the L ^A T _E X command <code>\ProvidesPackage</code> .
<code>\LdfInit</code>	The macro <code>\LdfInit</code> 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.
<code>\ldf@quit</code>	The macro <code>\ldf@quit</code> 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 <code>\begin{document}</code> time, and ending the input stream.
<code>\ldf@finish</code>	The macro <code>\ldf@finish</code> 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 <code>\begin{document}</code> time.
<code>\loadlocalcfg</code>	After processing a language definition file, L ^A T _E X can be instructed to load a local configuration file. This file can, for instance, be used to add strings to <code>\captions<lang></code> to support local document classes. The user will be informed that this configuration file has been loaded. This macro is called by <code>\ldf@finish</code> .
<code>\substitutefontfamily</code>	(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.

3.2 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. 3.7 (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>

\bbld@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\monthinname{<name of first month>}
% More strings

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

\StartBabelCommands*{<dialect>}{date}
\SetString\monthinname{<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>}
```

3.3 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

L^AT_EX to give a character the category code ‘active’. When a character has been made active it will remain that way until the end of the document. Its definition may vary.

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

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

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

`\bbl@add@special` The T_EXbook states: “Plain T_EX 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]

`\bbl@remove@special` 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`. L^AT_EX 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.

3.4 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 this²³.

`\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.

3.5 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 behaviour is preserved for backward compatibility. If you are using `etoolbox`, by Philipp Lehman, consider using the tools provided by this package instead of `\addto`.

3.6 Macros common to a number of languages

`\bbl@allowhyphens` In several languages compound words are used. This means that when T_EX 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.

²³This mechanism was introduced by Bernd Raichle.

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 behaviour of `\bbl@allowhyphens`.

`\set@low@box` For some languages, quotes need to be lowered to the baseline. For this purpose the macro `\set@low@box` is available. It takes one argument and puts that argument in an `\hbox`, at the baseline. The result is available in `\box0` for further processing.

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

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

3.7 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 `\category` is either captions, date or extras. You must stick to these three categories, even if no error is raised when using other name.²⁴ It may be empty, too, but in such a case using `\SetString` is an error (but not `\SetCase`).

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

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

\EndBabelCommands
```

A real example is:

```
\StartBabelCommands{austrian}{date}
  [unicode, fontenc=TU EU1 EU2, charset=utf8]
  \SetString\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\monthxiname{Oktober}
  \SetString\monthxiiname{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 `\category``\language` are overridden, which means the old way to define strings still works and used by default (to be precise, is first set

²⁴In future releases further categories may be added.

to undefined and then strings are added). However, when used in the preamble or in a package, new settings are added to the previous ones, if the language exists (in the babel sense, ie, if `\date<language>` exists).

\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.²⁵

\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 `\@uc1clist` (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=`İ\relax
```

²⁵This replaces in 3.9g a short-lived `\UseStrings` which has been removed because it did not work.

```

\uccode`ı=`İ\relax}
{\lccode`İ=`ı\relax
 \lccode`I=`ı\relax}

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

\EndBabelCommands

```

(Note the mapping for 0T1 is not complete.)

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

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{\langle uccode \rangle}{\langle lccode \rangle}` 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{\langle uccode-from \rangle}{\langle uccode-to \rangle}{\langle step \rangle}{\langle lccode-from \rangle}` 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{\langle uccode-from \rangle}{\langle uccode-to \rangle}{\langle step \rangle}{\langle lccode \rangle}` 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}{\text{11F}}{2}{\text{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 are 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 behaviour 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`name. 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.

4.2 Changes in babel version 3.7

In `babel` version 3.7 a number of bugs that were found in version 3.6 are fixed. Also a number of changes and additions have occurred:

- Shorthands are expandable again. The disadvantage is that one has to type `'{ }a` when the acute accent is used as a shorthand character. The advantage is that a number of other problems (such as the breaking of ligatures, etc.) have vanished.
- Two new commands, `\shorthandon` and `\shorthandoff` have been introduced to enable to temporarily switch off one or more shorthands.
- Support for typesetting Hebrew (and potential support for typesetting other right-to-left written languages) is now available thanks to Rama Porrat and Boris Lavva.
- A language attribute has been added to the `\mark . . .` commands in order to make sure that a Greek header line comes out right on the last page before a language switch.
- Hyphenation pattern files are now read *inside a group*; therefore any changes a pattern file needs to make to lowercase codes, uppercase codes, and category codes are kept local to that group. If they are needed for the language, these changes will need to be repeated and stored in `\extras . . .`
- The concept of language attributes is introduced. It is intended to give the user some control over the features a language-definition file provides. Its first use is for the Greek language, where the user can choose the `πολυτονικό` (“polytonikó” or multi-accented) Greek way of typesetting texts.
- The environment `hyphenrules` is introduced.
- The syntax of the file `language.dat` has been extended to allow (optionally) specifying the font encoding to be used while processing the patterns file.
- The command `\providehyphenmins` should now be used in language definition files in order to be able to keep any settings provided by the pattern file.

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.

```
1 <<version=3.15>>
2 <<date=2017/11/03>>
```

6 Tools

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

We define some basic macros which just make the code cleaner. `\bbl@add` is now used internally instead of `\addto` because of the unpredictable behaviour 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 loaded until options have been defined. This does not hurt, but should be fixed somehow.

```
3 <<*Basic macros>> ≡
4 \def\bbl@stripslash{\expandafter\@gobble\string}
5 \def\bbl@add#1#2{%
6   \bbl@ifunset{\bbl@stripslash#1}%
7     {\def#1{#2}}%
8     {\expandafter\def\expandafter#1\expandafter{#1#2}}
9 \def\bbl@xin@{\@expandtwoargs\in@}
10 \def\bbl@csarg#1#2{\expandafter#1\csname bbl@#2\endcsname}%
11 \def\bbl@cs#1{\csname bbl@#1\endcsname}
12 \def\bbl@loop#1#2#3{\bbl@loop#1{#3}#2,\@nnil,}
13 \def\bbl@loopx#1#2{\expandafter\bbl@loop\expandafter#1\expandafter{#2}}
14 \def\bbl@loop@loop#1#2#3,{%
15   \ifx\@nnil#3\relax\else
```

```

16   \def#1{#3}#2\bbl@afterfi\bbl@loop#1{#2}%
17   \fi}
18 \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.

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

\bbl@afterelse Because the code that is used in the handling of active characters may need to look ahead,
\bbl@afterfi we take extra care to ‘throw’ it over the \else and \fi parts of an \if-statement26. These
               macros will break if another \if... \fi statement appears in one of the arguments and it
               is not enclosed in braces.

25 \long\def\bbl@afterelse#1\else#2\fi{\fi#1}
26 \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.

27 \def\bbl@tempa#1{%
28   \long\def\bbl@trim##1##2{%
29     \futurelet\bbl@trim@a\bbl@trim@c##2\@nil\@nil#1\@nil\relax{##1}}%
30   \def\bbl@trim@c{%
31     \ifx\bbl@trim@a\@sptoken
32       \expandafter\bbl@trim@b
33     \else
34       \expandafter\bbl@trim@b\expandafter#1%
35     \fi}%
36   \long\def\bbl@trim@b#1##1 \@nil{\bbl@trim@i##1}}
37 \bbl@tempa{ }
38 \long\def\bbl@trim@i#1\@nil#2\relax#3{#3{#1}}
39 \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  $\epsilon$ -tex engine, it is based on \ifcsname, which is more
              efficient, and do not waste memory.

40 \def\bbl@ifunset#1{%
41   \expandafter\ifx\csname#1\endcsname\relax
42     \expandafter\@firstoftwo
43   \else
44     \expandafter\@secondoftwo
45   \fi}
46 \bbl@ifunset{ifcsname}%
47 {}%
48 {\def\bbl@ifunset#1{%
49   \ifcsname#1\endcsname
50     \expandafter\ifx\csname#1\endcsname\relax

```

²⁶This code is based on code presented in TUGboat vol. 12, no2, June 1991 in “An expansion Power Lemma” by Sonja Maus.


```

51      \bbl@afterelse\expandafter\@firstoftwo
52      \else
53      \bbl@afterfi\expandafter\@secondoftwo
54      \fi
55      \else
56      \expandafter\@firstoftwo
57      \fi}}

```

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

```

58 \def\bbl@ifblank#1{%
59   \bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
60 \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).

```

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

```

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

```

\bbl@replace

```

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

```

93 \def\bbl@exp#1{%
94   \begingroup
95     \let\\noexpand
96     \def\<##1>{\expandafter\noexpand\csname##1\endcsname}%
97     \edef\bbl@exp@aux{\endgroup#1}%
98     \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 pdfTeX, 1 is luatex, and 2 is xetex. You may use the latter it in your language style if you want.

```

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

```

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

```

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

```

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

```

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

```

136 <<*Load macros for plain if not LaTeX>> ≡
137 \ifx\AtBeginDocument\@undefined

```

```

138 \input plain.def\relax
139 \fi
140 <</Load macros for plain if not LaTeX>>

```

6.1 Multiple languages

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

```

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

```

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

7 The Package File (\LaTeX , babel.sty)

In order to make use of the features of $\text{\LaTeX} 2_{\epsilon}$, 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 behaviour 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.

7.1 base

The first option to be processed is base, which sets 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.

```
162 (*package)
163 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
164 \ProvidesPackage{babel}[\langle\langle date \rangle\rangle \langle\langle version \rangle\rangle The Babel package]
165 \@ifpackagewith{babel}{debug}
166   {\let\bbl@debug\@firstofone}
167   {\let\bbl@debug\@gobble}
168 \input switch.def\relax
169 \langle\langle Load patterns in luatex \rangle\rangle
170 \langle\langle Basic macros \rangle\rangle
171 \def\AfterBabelLanguage#1{%
172   \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.

```
173 \ifx\bbl@languages\undefined\else
174   \begingroup
175     \catcode\^^I=12
176     \@ifpackagewith{babel}{showlanguages}{%
177       \begingroup
178         \def\bbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}%
179         \wlog{<*languages>}%
180         \bbl@languages
181         \wlog{</languages>}%
182       \endgroup}{%
183     \endgroup
184     \def\bbl@elt#1#2#3#4{%
185       \ifnum#2=\z@
186         \gdef\bbl@nulllanguage{#1}%
187         \def\bbl@elt##1##2##3##4{%
188           \fi}%
189       \bbl@languages
190     \fi
191     \ifodd\bbl@engine
192       \@ifpackagewith{babel}{bidi=basic-r}{% must go before any \DeclareOption
193         \let\bbl@beforeforeign\leavevmode
194         \AtEndOfPackage{\EnableBabelHook{babel-bidi}}%
195         \RequirePackage{luatexbase}%
196         \directlua{
197           require('babel-bidi.lua')}

```

```

198     require('babel-bidi-basic-r.lua')
199     luatexbase.add_to_callback('pre_linebreak_filter',
200         Babel.pre_otfload,
201         'Babel.pre_otfload',
202         luatexbase.priority_in_callback('pre_linebreak_filter',
203             'luaotfload.node_processor') or nil)
204     luatexbase.add_to_callback('hpack_filter',
205         Babel.pre_otfload,
206         'Babel.pre_otfload',
207         luatexbase.priority_in_callback('hpack_filter',
208             'luaotfload.node_processor') or nil)}}}
209 \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.

```

210 \@ifpackagewith{babel}{base}{%
211   \ifx\directlua\undefined
212     \DeclareOption*{\bbl@patterns{\CurrentOption}}%
213   \else
214     \DeclareOption*{\bbl@patterns@lua{\CurrentOption}}%
215   \fi
216   \DeclareOption{base}{}%
217   \DeclareOption{showlanguages}{}%
218   \ProcessOptions
219   \global\expandafter\let\csname opt@babel.sty\endcsname\relax
220   \global\expandafter\let\csname ver@babel.sty\endcsname\relax
221   \global\let@ifl@ter@@\ifl@ter
222   \def@ifl@ter#1#2#3#4#5{\global\let@ifl@ter\ifl@ter@@}%
223   \endinput}}}%

```

7.2 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 `\load keyval`, for example.

```

224 \bbl@csarg\let{tempa\expandafter}\csname opt@babel.sty\endcsname
225 \def\bbl@tempb#1.#2{%
226   #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%
227 \def\bbl@tempd#1.#2@nnil{%
228   \ifx\@empty#2%
229     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
230   \else
231     \in@{=}{#1}\ifin@
232     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
233   \else
234     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
235   \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
236   \fi
237 \fi}
238 \let\bbl@tempc\@empty
239 \bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty@nnil}
240 \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.

```

241 \DeclareOption{KeepShorthandsActive}{}
242 \DeclareOption{activeacute}{}
243 \DeclareOption{activegrave}{}
244 \DeclareOption{debug}{}
245 \DeclareOption{noconfigs}{}
246 \DeclareOption{showlanguages}{}
247 \DeclareOption{silent}{}
248 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
249 <<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.

```

250 \let\bbl@opt@shorthands\@nnil
251 \let\bbl@opt@config\@nnil
252 \let\bbl@opt@main\@nnil
253 \let\bbl@opt@headfoot\@nnil
254 \let\bbl@opt@layout\@nnil

```

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

```

255 \def\bbl@tempa#1=#2\bbl@tempa{%
256   \bbl@csarg\ifx{opt@#1}\@nnil
257     \bbl@csarg\edef{opt@#1}{#2}%
258   \else
259     \bbl@error{%
260       Bad option `#1=#2'. Either you have misspelled the\\%
261       key or there is a previous setting of `#1'}{%
262       Valid keys are `shorthands', `config', `strings', `main',\\%
263       `headfoot', `safe', `math'}
264   \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.

```

265 \let\bbl@language@opts\@empty
266 \DeclareOption*{%
267   \bbl@xin@{\string=}{\CurrentOption}%
268   \ifin@
269     \expandafter\bbl@tempa\CurrentOption\bbl@tempa
270   \else
271     \bbl@add@list\bbl@language@opts{\CurrentOption}%
272   \fi}

```

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

```

273 \ProcessOptions*

```

7.3 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@ifshorthands` is always true, and it is always false if `shorthands` is empty. Also, some code makes sense only with `shorthands=...`

```

274 \def\bbl@sh@string#1{%
275   \ifx#1\@empty\else
276     \ifx#1t\string~%
277     \else\ifx#1c\string,%
278     \else\string#1%
279     \fi\fi
280   \expandafter\bbl@sh@string
281 \fi}
282 \ifx\bbl@opt@shorthands\@nnil
283   \def\bbl@ifshorthand#1#2#3{#2}%
284 \else\ifx\bbl@opt@shorthands\@empty
285   \def\bbl@ifshorthand#1#2#3{#3}%
286 \else

```

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

```

287 \def\bbl@ifshorthand#1{%
288   \bbl@xin@{\string#1}{\bbl@opt@shorthands}%
289   \ifin@
290     \expandafter\@firstoftwo
291   \else
292     \expandafter\@secondoftwo
293   \fi}

```

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

```

294 \edef\bbl@opt@shorthands{%
295   \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.

```

296 \bbl@ifshorthand{'}%
297   {\PassOptionsToPackage{activeacute}{babel}}{}
298 \bbl@ifshorthand{`}%
299   {\PassOptionsToPackage{activegrave}{babel}}{}
300 \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.

```

301 \ifx\bbl@opt@headfoot\@nnil\else
302   \g@addto@macro\@resetactivechars{%
303     \set@typeset@protect
304     \expandafter\select@language@x\expandafter{\bbl@opt@headfoot}%
305     \let\protect\noexpand}
306 \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.

```

307 \ifx\bbl@opt@safe\@undefined
308   \def\bbl@opt@safe{BR}
309 \fi
310 \ifx\bbl@opt@main\@nnil\else
311   \edef\bbl@language@opts{%
312     \ifx\bbl@language@opts\@empty\else\bbl@language@opts,\fi
313     \bbl@opt@main}
314 \fi

```

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

```

315 \ifx\bbl@opt@layout\@nnil

```

```

316 \newcommand\IfBabelLayout[3]{#3}%
317 \else
318 \newcommand\IfBabelLayout[1]{%
319 \expandafter\in@{.#1.}{.\bbl@opt@layout.}%
320 \ifin@
321 \expandafter\@firstoftwo
322 \else
323 \expandafter\@secondoftwo
324 \fi}
325 \fi

```

7.4 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).

```

326 \let\bbl@afterlang\relax
327 \let\BabelModifiers\relax
328 \let\bbl@loaded\@empty
329 \def\bbl@load@language#1{%
330 \InputIfFileExists{#1.ldf}%
331 {\edef\bbl@loaded{\CurrentOption
332 \ifx\bbl@loaded\@empty\else,\bbl@loaded\fi}%
333 \expandafter\let\expandafter\bbl@afterlang
334 \csname\CurrentOption.ldf-h@@k\endcsname
335 \expandafter\let\expandafter\BabelModifiers
336 \csname bbl@mod@\CurrentOption\endcsname}%
337 {\bbl@error{%
338 Unknown option '\CurrentOption'. Either you misspelled it\\
339 or the language definition file \CurrentOption.ldf was not found}}%
340 Valid options are: shorthands=, KeepShorthandsActive,\\
341 activeacute, activegrave, noconfigs, safe=, main=, math=\\
342 headfoot=, strings=, config=, hyphenmap=, or a language name.}}}

```

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

```

343 \def\bbl@try@load@lang#1#2#3{%
344 \IfFileExists{\CurrentOption.ldf}%
345 {\bbl@load@language{\CurrentOption}}%
346 {#1\bbl@load@language{#2}#3}}
347 \DeclareOption{afrikaans}{\bbl@try@load@lang{}{dutch}}
348 \DeclareOption{brazil}{\bbl@try@load@lang{}{portuges}}
349 \DeclareOption{brazilian}{\bbl@try@load@lang{}{portuges}}
350 \DeclareOption{hebrew}{%
351 \input{rlbabel.def}%
352 \bbl@load@language{hebrew}}
353 \DeclareOption{hungarian}{\bbl@try@load@lang{}{magyar}}
354 \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}}
355 \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}}
356 \DeclareOption{polutonikogreek}{%
357 \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}}
358 \DeclareOption{portuguese}{\bbl@try@load@lang{}{portuges}}
359 \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}}
360 \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}}
361 \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.

```

362 \ifx\bbbl@opt@config\@nnil
363   \@ifpackagewith{babel}{noconfigs}{}%
364     {\InputIfFileExists{bblopts.cfg}%
365       {\typeout{*****^^J%
366                 * Local config file bblopts.cfg used^^J%
367                 *}}}%
368     }{}%
369 \else
370   \InputIfFileExists{\bbbl@opt@config.cfg}%
371     {\typeout{*****^^J%
372               * Local config file \bbbl@opt@config.cfg used^^J%
373               *}}}%
374     {\bbbl@error{%
375       Local config file '\bbbl@opt@config.cfg' not found}{%
376       Perhaps you misspelled it.}}}%
377 \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 `bbbl@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.

```

378 \bbbl@for\bbbl@tempa\bbbl@language@opts{%
379   \bbbl@ifunset{ds@\bbbl@tempa}%
380     {\edef\bbbl@tempb{%
381       \noexpand\DeclareOption
382       {\bbbl@tempa}%
383       {\noexpand\bbbl@load@language{\bbbl@tempa}}}%
384     \bbbl@tempb}%
385   \@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.

```

386 \bbbl@foreach\@classoptionslist{%
387   \bbbl@ifunset{ds@#1}%
388     {\IfFileExists{#1.ldf}%
389       {\DeclareOption{#1}{\bbbl@load@language{#1}}}%
390     }{}%
391   {}}

```

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

```

392 \ifx\bbbl@opt@main\@nnil\else
393   \expandafter
394   \let\expandafter\bbbl@loadmain\csname ds@\bbbl@opt@main\endcsname
395   \DeclareOption{\bbbl@opt@main}{}
396 \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):

```

397 \def\AfterBabelLanguage#1{%
398   \bbbl@ifsamestring\CurrentOption{#1}{\global\bbbl@add\bbbl@afterlang}{}}
399 \DeclareOption*{}
400 \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`.

```

401 \ifx\bbl@opt@main\@nnil
402 \edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
403 \let\bbl@tempc\@empty
404 \bbl@for\bbl@tempb\bbl@tempa{%
405   \bbl@xin@{,\bbl@tempb,}{,\bbl@loaded,}%
406   \ifin@edef\bbl@tempc{\bbl@tempb}\fi}
407 \def\bbl@tempa#1,#2\@nnil{\def\bbl@tempb{#1}}
408 \expandafter\bbl@tempa\bbl@loaded,\@nnil
409 \ifx\bbl@tempb\bbl@tempc\else
410   \bbl@warning{%
411     Last declared language option is '\bbl@tempc',\%
412     but the last processed one was '\bbl@tempb'.\%
413     The main language cannot be set as both a global\%
414     and a package option. Use 'main=\bbl@tempc' as\%
415     option. Reported}%
416 \fi
417 \else
418 \DeclareOption{\bbl@opt@main}{\bbl@loadmain}
419 \ExecuteOptions{\bbl@opt@main}
420 \DeclareOption*{}
421 \ProcessOptions*
422 \fi
423 \def\AfterBabelLanguage{%
424   \bbl@error
425   {Too late for \string\AfterBabelLanguage}%
426   {Languages have been loaded, so I can do nothing}}
427 \ifx\bbl@main@language\@undefined
428   \bbl@error{%
429     You haven't specified a language option}{%
430     You need to specify a language, either as a global option\%
431     or as an optional argument to the \string\usepackage\space
432     command;\%
433     You shouldn't try to proceed from here, type x to quit.}
434 \fi
435 \endpackage

```

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

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

will have to be checked in a number of places. Some of the code below is common to plain 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.

8.1 Tools

```

436 < *core >
437 \ifx\ldf@quit\undefined
438 \else
439   \expandafter\endinput
440 \fi
441 <<Make sure ProvidesFile is defined>>
442 \ProvidesFile{babel.def}[\<date>] \<version> Babel common definitions]
443 <<Load macros for plain if not LaTeX>>
444 \ifx\bbl@ifshorthand\undefined
445   \def\bbl@ifshorthand#1#2#3{#2}%
446   \def\bbl@opt@safe{BR}
447   \def\AfterBabelLanguage#1#2{}
448   \let\bbl@afterlang\relax
449   \let\bbl@language@opts\empty
450 \fi
451 \input switch.def\relax
452 \ifx\bbl@languages\undefined
453   \ifx\directlua\undefined
454     \openin1 = language.def
455     \ifeof1
456       \closein1
457       \message{I couldn't find the file language.def}
458     \else
459       \closein1
460       \begingroup
461         \def\addlanguage#1#2#3#4#5{%
462           \expandafter\ifx\csname lang@#1\endcsname\relax\else
463             \global\expandafter\let\csname l@#1\endcsname
464               \csname lang@#1\endcsname
465           \fi}%
466         \def\uselanguage#1{}%
467         \input language.def
468       \endgroup
469     \fi
470   \fi
471   \chardef\l@english\z@
472 \fi
473 <<Load patterns in luatex>>
474 <<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.

```

475 \def\addto#1#2{%
476   \ifx#1\@undefined
477     \def#1{#2}%
478   \else
479     \ifx#1\relax
480       \def#1{#2}%
481     \else
482       {\toks@\expandafter{#1#2}%
483        \xdef#1{\the\toks@}}%
484   \fi
485 \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.

```

486 \def\bbl@withactive#1#2{%
487   \begingroup
488   \lccode`~=#2\relax
489   \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@. . .`

```

490 \def\bbl@redefine#1{%
491   \edef\bbl@tempa{\bbl@stripslash#1}%
492   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
493   \expandafter\def\csname\bbl@tempa\endcsname}

```

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

```

494 \@onlypreamble\bbl@redefine

```

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

```

495 \def\bbl@redefine@long#1{%
496   \edef\bbl@tempa{\bbl@stripslash#1}%
497   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
498   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
499 \@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_`.

```

500 \def\bbl@redefineroobust#1{%
501   \edef\bbl@tempa{\bbl@stripslash#1}%
502   \bbl@ifunset{\bbl@tempa\space}%
503   {\expandafter\let\csname org@\bbl@tempa\endcsname#1%
504    \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
505   {\bbl@exp{\let\<org@\bbl@tempa>\<\bbl@tempa\space>}}}%
506   \@namedef{\bbl@tempa\space}}

```

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

```

507 \@onlypreamble\bbl@redefineroobust

```

8.2 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.

```

508 \def\AddBabelHook#1#2{%
509   \bbl@ifunset{bbl@hk@#1}{\EnableBabelHook{#1}}{}%
510   \def\bbl@tempa##1,#2=##2,##3\@empty{\def\bbl@tempb{##2}}%
511   \expandafter\bbl@tempa\bbl@evargs,#2=,\@empty
512   \bbl@ifunset{bbl@ev@#1@#2}%
513     {\bbl@csarg\bbl@add{ev@#2}{\bbl@elt{#1}}}%
514     \bbl@csarg\newcommand}%
515     {\bbl@csarg\let{ev@#1@#2}\relax
516     \bbl@csarg\newcommand}%
517     {ev@#1@#2}[\bbl@tempb]}
518 \def\EnableBabelHook#1{\bbl@csarg\let{hk@#1}\@firstofone}
519 \def\DisableBabelHook#1{\bbl@csarg\let{hk@#1}\@gobble}
520 \def\bbl@usehooks#1#2{%
521   \def\bbl@elt##1{%
522     \@nameuse{bbl@hk@##1}{\@nameuse{bbl@ev@##1@#1}#2}}%
523   \@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).

```

524 \def\bbl@evargs{% don't delete the comma
525   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
526   adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
527   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
528   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.

```

529 \newcommand\babelensure[2][{}% TODO - revise test files
530   \AddBabelHook{babel-ensure}{afterextras}%
531   \ifcase\bbl@select@type
532     \@nameuse{bbl@e@\language name}%
533   \fi}%
534 \begin{group}
535   \let\bbl@ens@include\@empty
536   \let\bbl@ens@exclude\@empty
537   \def\bbl@ens@fontenc{\relax}%
538   \def\bbl@tempb##1{%
539     \ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}%
540   \edef\bbl@tempa{\bbl@tempb#1\@empty}%
541   \def\bbl@tempb##1=##2\@{\@namedef{bbl@ens@##1}{##2}}%

```

```

542 \bbl@foreach\bbl@tempa{\bbl@tempb##1\@@}%
543 \def\bbl@tempc{\bbl@ensure}%
544 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
545 \expandafter{\bbl@ens@include}}%
546 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
547 \expandafter{\bbl@ens@exclude}}%
548 \toks@\expandafter{\bbl@tempc}%
549 \bbl@exp{%
550 \endgroup
551 \def\<bbl@e@#2>{\the\toks@{\bbl@ens@fontenc}}}%
552 \def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
553 \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
554 \ifx##1\@empty\else
555 \in@{##1}{#2}%
556 \ifin\else
557 \bbl@ifunset{\bbl@ensure@\language}%
558 {\bbl@exp{%
559 \\\DeclareRobustCommand\<bbl@ensure@\language>[1]{%
560 \\\foreignlanguage{\language}%
561 {\ifx\relax#3\else
562 \\\fontencoding{#3}\selectfont
563 \fi
564 #####1}}}%
565 }%
566 \toks@\expandafter{##1}%
567 \edef##1{%
568 \bbl@csarg\noexpand{ensure@\language}%
569 {\the\toks@}}%
570 \fi
571 \expandafter\bbl@tempb
572 \fi}%
573 \expandafter\bbl@tempb\bbl@captionslist\today\@empty
574 \def\bbl@tempa##1{% elt for include list
575 \ifx##1\@empty\else
576 \bbl@csarg\in@{ensure@\language\expandafter}\expandafter{##1}%
577 \ifin\else
578 \bbl@tempb##1\@empty
579 \fi
580 \expandafter\bbl@tempa
581 \fi}%
582 \bbl@tempa#1\@empty}
583 \def\bbl@captionslist{%
584 \prefacename\refname\abstractname\bibname\chaptername\appendixname
585 \contentsname\listfigurename\listtablename\indexname\figurename
586 \tablename\partname\enclname\ccname\headtoname\pagename\seename
587 \alsoname\proofname\glossaryname}

```

8.3 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`.

```

588 \def\bbl@ldfinit{%
589   \let\bbl@screset\@empty
590   \let\BabelStrings\bbl@opt@string
591   \let\BabelOptions\@empty
592   \let\BabelLanguages\relax
593   \ifx\originalTeX\@undefined
594     \let\originalTeX\@empty
595   \else
596     \originalTeX
597   \fi}
598 \def\LdfInit#1#2{%
599   \chardef\atcatcode=\catcode`\@
600   \catcode`\@=11\relax
601   \chardef\eqcatcode=\catcode`\=
602   \catcode`\==12\relax
603   \expandafter\if\expandafter\@backslashchar
604     \expandafter\@car\string#2\@nil
605     \ifx#2\@undefined\else
606       \ldf@quit{#1}%
607     \fi
608   \else
609     \expandafter\ifx\csname#2\endcsname\relax\else
610       \ldf@quit{#1}%
611     \fi
612   \fi
613   \bbl@ldfinit}

```

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

```

614 \def\ldf@quit#1{%
615   \expandafter\main@language\expandafter{#1}%
616   \catcode`\@=\atcatcode \let\atcatcode\relax
617   \catcode`\==\eqcatcode \let\eqcatcode\relax
618   \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.

```

619 \def\bbl@afterldf#1{%
620   \bbl@afterlang
621   \let\bbl@afterlang\relax
622   \let\BabelModifiers\relax
623   \let\bbl@screset\relax}%
624 \def\ldf@finish#1{%
625   \loadlocalcfg{#1}%

```

```

626 \bbl@afterldf{#1}%
627 \expandafter\main@language\expandafter{#1}%
628 \catcode`\@=\atcatcode \let\atcatcode\relax
629 \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 .

```

630 \@onlypreamble\LdfInit
631 \@onlypreamble\ldf@quit
632 \@onlypreamble\ldf@finish

```

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

```

633 \def\main@language#1{%
634   \def\bbl@main@language{#1}%
635   \let\language\name\bbl@main@language
636   \bbl@patterns{\language\name}}

```

We also have to make sure that some code gets executed at the beginning of the document.

```

637 \AtBeginDocument{%
638   \expandafter\selectlanguage\expandafter{\bbl@main@language}%
639   \ifcase\bbl@engine\or\pagedir\bodydir\fi} % TODO - a better place

```

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

```

640 \def\select@language@x#1{%
641   \ifcase\bbl@select@type
642     \bbl@ifsamestring\language\name{#1}{\select@language{#1}}%
643   \else
644     \select@language{#1}%
645   \fi}

```

8.4 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.

```

646 \def\bbl@add@special#1{% 1:a macro like "\", "\?", etc.
647   \bbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
648   \bbl@ifunset{@sanitize}{\bbl@add@sanitize{\makeother#1}}%
649   \ifx\nfss@catcodes\undefined\else % TODO - same for above
650     \begingroup
651       \catcode`#1\active
652       \nfss@catcodes
653       \ifnum\catcode`#1=\active
654         \endgroup
655       \bbl@add\nfss@catcodes{\makeother#1}%
656     \else
657       \endgroup
658     \fi
659   \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.

```

660 \def\bbl@remove@special#1{%
661   \begingroup
662     \def\x##1##2{\ifnum`#1=`##2\noexpand\@empty
663       \else\noexpand##1\noexpand##2\fi}%
664     \def\do{\x\do}%
665     \def\@makeother{\x\@makeother}%
666   \edef\x{\endgroup
667     \def\noexpand\dospecials{\dospecials}%
668     \expandafter\ifx\csname @sanitize\endcsname\relax\else
669       \def\noexpand\@sanitize{\@sanitize}%
670     \fi}%
671   \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).

```

672 \def\bbl@active@def#1#2#3#4{%
673   \@namedef{#3#1}{%
674     \expandafter\ifx\csname#2@sh@#1\endcsname\relax
675       \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
676     \else
677       \bbl@afterfi\csname#2@sh@#1\endcsname
678     \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.

```

679   \long\@namedef{#3@arg#1}##1{%
680     \expandafter\ifx\csname#2@sh@#1\string##1\endcsname\relax
681       \bbl@afterelse\csname#4#1\endcsname##1%
682     \else
683       \bbl@afterfi\csname#2@sh@#1\string##1\endcsname
684     \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.

```

685 \def\initiate@active@char#1{%

```

```

686 \bbl@ifunset{active@char\string#1}%
687   {\bbl@withactive
688     {\expandafter\@initiate@active@char\expandafter}\#1\string#1}%
689   {}}

```

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`).

```

690 \def\@initiate@active@char#1#2#3{%
691   \bbl@csarg\edef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%
692   \ifx#1\@undefined
693     \bbl@csarg\edef{oridef@#2}{\let\noexpand#1\noexpand\@undefined}%
694   \else
695     \bbl@csarg\let{oridef@#2}\#1%
696     \bbl@csarg\edef{oridef@#2}{%
697       \let\noexpand#1%
698       \expandafter\noexpand\csname bbl@oridef@#2\endcsname}%
699   \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`).

```

700   \ifx#1#3\relax
701     \expandafter\let\csname normal@char#2\endcsname#3%
702   \else
703     \bbl@info{Making #2 an active character}%
704     \ifnum\mathcode`#2="8000
705       \@namedef{normal@char#2}{%
706         \textormath{#3}{\csname bbl@oridef@#2\endcsname}}%
707     \else
708       \@namedef{normal@char#2}{#3}%
709   \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).

```

710   \bbl@restoreactive{#2}%
711   \AtBeginDocument{%
712     \catcode`#2\active
713     \if@filesw
714       \immediate\write\@mainaux{\catcode`\string#2\active}%
715     \fi}%
716   \expandafter\bbl@add@special\csname#2\endcsname
717   \catcode`#2\active
718 \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⟩`).

```

719 \let\bbl@tempa\@firstoftwo
720 \if\string^#2%
721 \def\bbl@tempa{\noexpand\textormath}%
722 \else
723 \ifx\bbl@mathnormal\@undefined\else
724 \let\bbl@tempa\bbl@mathnormal
725 \fi
726 \fi
727 \expandafter\edef\csname active@char#2\endcsname{%
728 \bbl@tempa
729 {\noexpand\if@safe@actives
730 \noexpand\expandafter
731 \expandafter\noexpand\csname normal@char#2\endcsname
732 \noexpand\else
733 \noexpand\expandafter
734 \expandafter\noexpand\csname bbl@doactive#2\endcsname
735 \noexpand\fi}%
736 {\expandafter\noexpand\csname normal@char#2\endcsname}}%
737 \bbl@csarg\edef{doactive#2}{%
738 \expandafter\noexpand\csname user@active#2\endcsname}%

```

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

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

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

```

739 \bbl@csarg\edef{active@#2}{%
740 \noexpand\active@prefix\noexpand#1%
741 \expandafter\noexpand\csname active@char#2\endcsname}%
742 \bbl@csarg\edef{normal@#2}{%
743 \noexpand\active@prefix\noexpand#1%
744 \expandafter\noexpand\csname normal@char#2\endcsname}%
745 \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.

```

746 \bbl@active@def#2\user@group{user@active}{language@active}%
747 \bbl@active@def#2\language@group{language@active}{system@active}%
748 \bbl@active@def#2\system@group{system@active}{normal@char}%

```

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

```

749 \expandafter\edef\csname\user@group @sh#2@@\endcsname
750 {\expandafter\noexpand\csname normal@char#2\endcsname}%
751 \expandafter\edef\csname\user@group @sh#2@\string\protect@\endcsname
752 {\expandafter\noexpand\csname user@active#2\endcsname}%

```

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

```

753 \if\string'#2%
754 \let\prim@s\bbl@prim@s

```

```

755 \let\active@math@prime#1%
756 \fi
757 \bbl@usehooks{initiateactive}{\#1}{\#2}{\#3}}

```

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

```

758 <(*More package options)> ≡
759 \DeclareOption{math=active}{}
760 \DeclareOption{math=normal}{\def\bbl@mathnormal{\noexpand\textormath}}
761 <(/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.

```

762 \ifpackagewith{babel}{KeepShorthandsActive}%
763 {\let\bbl@restoreactive\@gobble}%
764 {\def\bbl@restoreactive#1{%
765   \bbl@exp{%
766     \\\AfterBabelLanguage\\\CurrentOption
767     {\catcode`#1=\the\catcode`#1\relax}%
768     \\\AtEndOfPackage
769     {\catcode`#1=\the\catcode`#1\relax}}}%
770 \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.

```

771 \def\bbl@sh@select#1#2{%
772   \expandafter\ifx\csname#1@sh@#2@sel\endcsname\relax
773     \bbl@afterelse\bbl@scndcs
774   \else
775     \bbl@afterfi\csname#1@sh@#2@sel\endcsname
776   \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`.

```

777 \def\active@prefix#1{%
778   \ifx\protect\@typeset@protect
779     \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).

```

780   \ifx\protect\@unexpandable@protect
781     \noexpand#1%
782   \else
783     \protect#1%
784   \fi
785   \expandafter\@gobble
786   \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}`.

```

787 \newif\if@safe@actives
788 \@safe@activesfalse

\bb1@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.

789 \def\bb1@restore@actives{\if@safe@actives\@safe@activesfalse\fi}

\bb1@activate Both macros take one argument, like \initiate@active@char. The macro is used to
\bb1@deactivate change the definition of an active character to expand to \active@char⟨char⟩ in the case
of \bb1@activate, or \normal@char⟨char⟩ in the case of \bb1@deactivate.

790 \def\bb1@activate#1{%
791   \bb1@withactive{\expandafter\let\expandafter}#1%
792   \csname bbl@active@\string#1\endcsname}
793 \def\bb1@deactivate#1{%
794   \bb1@withactive{\expandafter\let\expandafter}#1%
795   \csname bbl@normal@\string#1\endcsname}

\bb1@firstcs These macros have two arguments. They use one of their arguments to build a control
\bb1@scndcs sequence from.

796 \def\bb1@firstcs#1#2{\csname#1\endcsname}
797 \def\bb1@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.

798 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil}
799 \def\@decl@short#1#2#3\@nil#4{%
800   \def\bb1@tempa{#3}%
801   \ifx\bb1@tempa\@empty
802     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bb1@scndcs
803     \bb1@ifunset{#1@sh@\string#2@}{}%
804     {\def\bb1@tempa{#4}%
805       \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bb1@tempa
806       \else
807         \bb1@info
808         {Redefining #1 shorthand \string#2\%
809          in language \CurrentOption}%
810       \fi}%
811     \@namedef{#1@sh@\string#2@}{#4}%
812   \else
813     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bb1@firstcs
814     \bb1@ifunset{#1@sh@\string#2@\string#3@}{}%
815     {\def\bb1@tempa{#4}%
816       \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bb1@tempa
817       \else
818         \bb1@info
819         {Redefining #1 shorthand \string#2\string#3\%
820          in language \CurrentOption}%
821       \fi}%
822     \@namedef{#1@sh@\string#2@\string#3@}{#4}%
823   \fi}

```

<code>\textormath</code>	<p>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 <code>\textormath</code> is provided.</p> <pre> 824 \def\textormath{% 825 \ifmmode 826 \expandafter\@secondoftwo 827 \else 828 \expandafter\@firstoftwo 829 \fi}</pre>
<code>\user@group</code> <code>\language@group</code> <code>\system@group</code>	<p>The current concept of ‘shorthands’ supports three levels or groups of shorthands. For each level the name of the level or group is stored in a macro. The default is to have a user group; use language group ‘english’ and have a system group called ‘system’.</p> <pre> 830 \def\user@group{user} 831 \def\language@group{english} 832 \def\system@group{system}</pre>
<code>\useshorthands</code>	<p>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.</p> <pre> 833 \def\useshorthands{% 834 \@ifstar\bbl@usesh@s{\bbl@usesh@x}} 835 \def\bbl@usesh@s#1{% 836 \bbl@usesh@x 837 {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bbl@activate{#1}}}% 838 {#1}} 839 \def\bbl@usesh@x#1#2{% 840 \bbl@ifshorthand{#2}% 841 {\def\user@group{user}% 842 \initiate@active@char{#2}% 843 #1% 844 \bbl@activate{#2}}% 845 {\bbl@error 846 {Cannot declare a shorthand turned off (\string#2)} 847 {Sorry, but you cannot use shorthands which have been\\ 848 turned off in the package options}}}</pre>
<code>\defineshorthand</code>	<p>Currently we only support two groups of user level shorthands, named internally <code>user</code> and <code>user@<lang></code> (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 <code>\defineshorthand</code>) a new level is inserted for it (<code>user@generic</code>, done by <code>\bbl@set@user@generic</code>); we make also sure <code>{}</code> and <code>\protect</code> are taken into account in this new top level.</p> <pre> 849 \def\user@language@group{user@\language@group} 850 \def\bbl@set@user@generic#1#2{% 851 \bbl@ifunset{user@generic@active#1}% 852 {\bbl@active@def#1\user@language@group{user@active}{user@generic@active}% 853 \bbl@active@def#1\user@group{user@generic@active}{\language@active}% 854 \expandafter\edef\csname#2@sh#1@@\endcsname{% 855 \expandafter\noexpand\csname normal@char#1\endcsname}% 856 \expandafter\edef\csname#2@sh#1@\string\protect@\endcsname{% 857 \expandafter\noexpand\csname user@active#1\endcsname}}% 858 \@empty} 859 \newcommand\defineshorthand[3][user]{% 860 \edef\bbl@tempa{\zap@space#1 \@empty}% </pre>

```

861 \bbl@for\bbl@tempb\bbl@tempa{%
862   \if*\expandafter\@car\bbl@tempb\@nil
863   \edef\bbl@tempb{user@\expandafter\@gobble\bbl@tempb}%
864   \expandtwoargs
865   \bbl@set@user@generic{\expandafter\string\@car#2\@nil}\bbl@tempb
866   \fi
867   \declare@shorthand{\bbl@tempb}{#2}{#3}}

```

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

```

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

```

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

```

869 \def\aliasshorthand#1#2{%
870   \bbl@ifshorthand{#2}%
871   {\expandafter\ifx\csname active@char\string#2\endcsname\relax
872     \ifx\document\@notprerr
873       \@notshorthand{#2}%
874     \else
875       \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`.

```

876     \expandafter\let\csname active@char\string#2\expandafter\endcsname
877     \csname active@char\string#1\endcsname
878     \expandafter\let\csname normal@char\string#2\expandafter\endcsname
879     \csname normal@char\string#1\endcsname
880     \bbl@activate{#2}%
881   \fi
882 \fi}%
883 {\bbl@error
884   {Cannot declare a shorthand turned off (\string#2)}
885   {Sorry, but you cannot use shorthands which have been\\%
886     turned off in the package options}}}

```

`\@notshorthand`

```

887 \def\@notshorthand#1{%
888   \bbl@error{%
889     The character '\string #1' should be made a shorthand character;\\%
890     add the command \string\usesshorthands\string{#1\string} to
891     the preamble.\\%
892     I will ignore your instruction}%
893   {You may proceed, but expect unexpected results}}

```

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

```

894 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
895 \DeclareRobustCommand*\shorthandoff{%
896   \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
897 \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.

```

898 \def\bbl@switch@sh#1#2{%
899   \ifx#2\@nnil\else
900     \bbl@ifunset{\bbl@active@\string#2}%
901     {\bbl@error
902       {I cannot switch '\string#2' on or off--not a shorthand}%
903       {This character is not a shorthand. Maybe you made\%
904         a typing mistake? I will ignore your instruction}}%
905     {\ifcase#1%
906       \catcode`#212\relax
907       \or
908       \catcode`#2\active
909       \or
910       \csname bbl@oricat@\string#2\endcsname
911       \csname bbl@oridef@\string#2\endcsname
912       \fi}%
913   \bbl@afterfi\bbl@switch@sh#1%
914 \fi}

```

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

```

915 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
916 \def\bbl@putsh#1{%
917   \bbl@ifunset{\bbl@active@\string#1}%
918   {\bbl@putsh@i#1\@empty\@nnil}%
919   {\csname bbl@active@\string#1\endcsname}}
920 \def\bbl@putsh@i#1#2\@nnil{%
921   \csname\language @sh@\string#1@%
922     \ifx\@empty#2\else\string#2\fi\endcsname}
923 \ifx\bbl@opt@shorthands\@nnil\else
924   \let\bbl@s@initiate@active@char\initiate@active@char
925   \def\initiate@active@char#1{%
926     \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
927   \let\bbl@s@switch@sh\bbl@switch@sh
928   \def\bbl@switch@sh#1#2{%
929     \ifx#2\@nnil\else
930       \bbl@afterfi
931       \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
932     \fi}
933   \let\bbl@s@activate\bbl@activate
934   \def\bbl@activate#1{%
935     \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
936   \let\bbl@s@deactivate\bbl@deactivate
937   \def\bbl@deactivate#1{%
938     \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
939 \fi

```

`\bbl@prim@s` One of the internal macros that are involved in substituting `\prime` for each right quote in 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.

```

940 \def\bbl@prim@s{%

```



```

941 \prime\futurelet\@let@token\bbl@pr@m@s}
942 \def\bbl@if@primes#1#2{%
943 \ifx#1\@let@token
944 \expandafter\@firstoftwo
945 \else\ifx#2\@let@token
946 \bbl@afterelse\expandafter\@firstoftwo
947 \else
948 \bbl@afterfi\expandafter\@secondoftwo
949 \fi\fi}
950 \begingroup
951 \catcode`\^=7 \catcode`\*=\active \lccode`\*=\^
952 \catcode`\'=12 \catcode`\"=\active \lccode`\"=\'\'
953 \lowercase{%
954 \gdef\bbl@pr@m@s{%
955 \bbl@if@primes"%
956 \pr@@@s
957 {\bbl@if@primes*\pr@@@t\egroup}}
958 \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).

```

959 \initiate@active@char{~}
960 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
961 \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.

```

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

```

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

```

964 \ifx\f@encoding\@undefined
965 \def\f@encoding{OT1}
966 \fi

```

8.5 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.

```

967 \newcommand\languageattribute[2]{%
968 \def\bbl@tempc{#1}%
969 \bbl@fixname\bbl@tempc
970 \bbl@iflanguage\bbl@tempc{%
971 \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@attrs`. When that control sequence is not yet defined this attribute is certainly not selected before.

```
972 \ifx\bbl@known@attrs\@undefined
973 \in@false
974 \else
```

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

```
975 \bbl@xin@{,\bbl@tempc-##1,}{,\bbl@known@attrs,}%
976 \fi
```

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

```
977 \ifin@
978 \bbl@warning{%
979 You have more than once selected the attribute '##1'\%
980 for language #1}%
981 \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.

```
982 \bbl@exp{%
983 \bbl@add@list\bbl@known@attrs{\bbl@tempc-##1}}%
984 \edef\bbl@tempa{\bbl@tempc-##1}%
985 \expandafter\bbl@ifknown@ttrib\expandafter{\bbl@tempa}\bbl@attributes%
986 {\csname\bbl@tempc @attr##1\endcsname}%
987 {\@attrerr{\bbl@tempc}{##1}}%
988 \fi}}
```

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

```
989 \@onlypreamble\languageattribute
```

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

```
990 \newcommand*{\@attrerr}[2]{%
991 \bbl@error
992 {The attribute #2 is unknown for language #1.}%
993 {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}`.

```
994 \def\bbl@declare@ttribute#1#2#3{%
995 \bbl@xin@{,#2,}{,\BabelModifiers,}%
996 \ifin@
997 \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
998 \fi
999 \bbl@add@list\bbl@attributes{#1-#2}%
1000 \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.

```
1001 \def\bbl@ifattributeset#1#2#3#4{%
```

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

```
1002 \ifx\bbbl@known@attribs\@undefined
1003   \in@false
1004 \else
```

The we need to check the list of known attributes.

```
1005   \bbbl@xin@{,#1-#2,}{,\bbbl@known@attribs,}%
1006 \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`'.

```
1007 \ifin@
1008   \bbbl@afterelse#3%
1009 \else
1010   \bbbl@afterfi#4%
1011 \fi
1012 }
```

`\bbbl@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.

```
1013 \def\bbbl@ifknown@ttrib#1#2{%
```

We first assume the attribute is unknown.

```
1014   \let\bbbl@tempa\@secondoftwo
```

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

```
1015   \bbbl@loopx\bbbl@tempb{#2}{%
1016     \expandafter\in@\expandafter{\expandafter,\bbbl@tempb,}{,#1,}%
1017     \ifin@
```

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

```
1018     \let\bbbl@tempa\@firstoftwo
1019   \else
1020   \fi}%
```

Finally we execute `\bbbl@tempa`.

```
1021   \bbbl@tempa
1022 }
```

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

```
1023 \def\bbbl@clear@ttribs{%
1024   \ifx\bbbl@attributes\@undefined\else
1025     \bbbl@loopx\bbbl@tempa{\bbbl@attributes}{%
1026       \expandafter\bbbl@clear@ttrib\bbbl@tempa.
1027     }%
1028   \let\bbbl@attributes\@undefined
1029 \fi}
1030 \def\bbbl@clear@ttrib#1-#2.{%
1031   \expandafter\let\csname#1@attr@#2\endcsname\@undefined}
1032 \AtBeginDocument{\bbbl@clear@ttribs}
```

8.6 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` 1033 `\def\babel@beginsave{\babel@savecnt\z@}`

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

1034 `\newcount\babel@savecnt`

1035 `\babel@beginsave`

`\babel@save` The macro `\babel@save<csname>` saves the current meaning of the control sequence `<csname>` to `\originalTeX`²⁷. To do this, we let the current meaning to a temporary control sequence, the restore commands are appended to `\originalTeX` and the counter is incremented.

1036 `\def\babel@save#1{%`

1037 `\expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax`

1038 `\toks@\expandafter{\originalTeX\let#1=}%`

1039 `\bbl@exp{%`

1040 `\def\\originalTeX{\the\toks@<babel@number\babel@savecnt>\relax}}%`

1041 `\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.

1042 `\def\babel@savevariable#1{%`

1043 `\toks@\expandafter{\originalTeX #1=}%`

1044 `\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
`\bbl@nonfrenchspacing` command `\bbl@frenchspacing` switches it on when it isn't already in effect and
`\bbl@nonfrenchspacing` switches it off if necessary.

1045 `\def\bbl@frenchspacing{%`

1046 `\ifnum\the\sfcodes\@m`

1047 `\let\bbl@nonfrenchspacing\relax`

1048 `\else`

1049 `\frenchspacing`

1050 `\let\bbl@nonfrenchspacing\nonfrenchspacing`

1051 `\fi}`

1052 `\let\bbl@nonfrenchspacing\nonfrenchspacing`

8.7 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.

1053 `\def\babeltags#1{%`

1054 `\edef\bbl@tempa{\zap@space#1 \@empty}%`

1055 `\def\bbl@tempb##1=##2\@{%`

²⁷`\originalTeX` has to be expandable, i. e. you shouldn't let it to `\relax`.

```

1056 \edef\bbl@tempc{%
1057 \noexpand\newcommand
1058 \expandafter\noexpand\csname ##1\endcsname{%
1059 \noexpand\protect
1060 \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
1061 \noexpand\newcommand
1062 \expandafter\noexpand\csname text##1\endcsname{%
1063 \noexpand\foreignlanguage{##2}}
1064 \bbl@tempc}%
1065 \bbl@for\bbl@tempa\bbl@tempa{%
1066 \expandafter\bbl@tempb\bbl@tempa\@@}}

```

8.8 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.

```

1067 \@onlypreamble\babelhyphenation
1068 \AtEndOfPackage{%
1069 \newcommand\babelhyphenation[2][\@empty]{%
1070 \ifx\bbl@hyphenation@\relax
1071 \let\bbl@hyphenation@\@empty
1072 \fi
1073 \ifx\bbl@hyphlist\@empty\else
1074 \bbl@warning{%
1075 You must not intermingle \string\selectlanguage\space and\\%
1076 \string\babelhyphenation\space or some exceptions will not\\%
1077 be taken into account. Reported}%
1078 \fi
1079 \ifx\@empty#1%
1080 \protected@edef\bbl@hyphenation@{\bbl@hyphenation@\space#2}%
1081 \else
1082 \bbl@vforeach{#1}{%
1083 \def\bbl@tempa{##1}%
1084 \bbl@fixname\bbl@tempa
1085 \bbl@iflanguage\bbl@tempa{%
1086 \bbl@csarg\protected@edef{hyphenation@\bbl@tempa}{%
1087 \bbl@ifunset{bbl@hyphenation@\bbl@tempa}%
1088 \@empty
1089 {\csname bbl@hyphenation@\bbl@tempa\endcsname\space}%
1090 #2}}}%
1091 \fi}}

```

`\bbl@allowhyphens` This macro makes hyphenation possible. Basically its definition is nothing more than `\nobreak \hskip 0pt plus 0pt28`.

```

1092 \def\bbl@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}
1093 \def\bbl@t@one{T1}
1094 \def\allowhyphens{\ifx\cf@encoding\bbl@t@one\else\bbl@allowhyphens\fi}

```

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

```

1095 \newcommand\babellnullhyphen{\char\hyphenchar\font}
1096 \def\babelhyphen{\active@prefix\babelhyphen\bbl@hyphen}

```

²⁸ \TeX begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.

```

1097 \def\bbl@hyphen{%
1098   \ifstar{\bbl@hyphen@i @}{\bbl@hyphen@i\@empty}}
1099 \def\bbl@hyphen@i#1#2{%
1100   \bbl@ifunset{bbl@hy@#1#2\@empty}%
1101   {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%
1102   {\csname bbl@hy@#1#2\@empty\endcsname}}

```

The following two commands are used to wrap the “hyphen” and set the behaviour 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.

```

1103 \def\bbl@usehyphen#1{%
1104   \leavevmode
1105   \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
1106   \nobreak\hskip\z@skip}
1107 \def\bbl@@usehyphen#1{%
1108   \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

The following macro inserts the hyphen char.

```

1109 \def\bbl@hyphenchar{%
1110   \ifnum\hyphenchar\font=\m@ne
1111     \babelnullhyphen
1112   \else
1113     \char\hyphenchar\font
1114   \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.

```

1115 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1116 \def\bbl@hy@@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
1117 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
1118 \def\bbl@hy@@hard{\bbl@usehyphen\bbl@hyphenchar}
1119 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}
1120 \def\bbl@hy@@nobreak{\mbox{\bbl@hyphenchar}}
1121 \def\bbl@hy@repeat{%
1122   \bbl@usehyphen{%
1123     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}
1124 \def\bbl@hy@@repeat{%
1125   \bbl@usehyphen{%
1126     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}
1127 \def\bbl@hy@empty{\hskip\z@skip}
1128 \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.

```

1129 \def\bbl@disc#1#2{\nobreak\discretionary{#2-}{}{#1}\bbl@allowhyphens}

```

8.9 Multiencoding strings

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

Tools But first, a couple of tools. The first one makes global a local variable. This is not the best solution, but it works.

```

1130 \def\bbl@tglobal#1{\global\let#1#1}
1131 \def\bbl@recatcode#1{%
1132   \@tempcnta="7F
1133   \def\bbl@tempa{%
1134     \ifnum\@tempcnta>"FF\else
1135       \catcode\@tempcnta=#1\relax
1136       \advance\@tempcnta\@ne
1137       \expandafter\bbl@tempa
1138     \fi}%
1139   \bbl@tempa}

```

The second one. We need to patch `\@uclclist`, but it is done once and only if `\SetCase` is used or if strings are encoded. The code is far from satisfactory for several reasons, including the fact `\@uclclist` is not a list any more. Therefore a package option is added to ignore it. Instead of gobbling the macro getting the next two elements (usually `\reserved@a`), we pass it as argument to `\bbl@uclc`. The parser is restarted inside `\(lang)\bbl@uclc` because we do not know how many expansions are necessary (depends on whether strings are encoded). The last part is tricky – when uppercasing, we have:

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

and starts over (and similarly when lowercasing).

```

1140 \@ifpackagewith{babel}{nocase}%
1141 {\let\bbl@patchuclc\relax}%
1142 {\def\bbl@patchuclc{%
1143   \global\let\bbl@patchuclc\relax
1144   \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}%
1145   \gdef\bbl@uclc##1{%
1146     \let\bbl@encoded\bbl@encoded@uclc
1147     \bbl@ifunset{\language @bbl@uclc}% and resumes it
1148     {##1}%
1149     {\let\bbl@tempa##1\relax % Used by LANG@bbl@uclc
1150      \csname\language @bbl@uclc\endcsname}%
1151     {\bbl@tolower\@empty}{\bbl@toupper\@empty}}%
1152   \gdef\bbl@tolower{\csname\language @bbl@lc\endcsname}%
1153   \gdef\bbl@toupper{\csname\language @bbl@uc\endcsname}}%
1154 <<More package options>> ≡
1155 \DeclareOption{nocase}{}
1156 <</More package options>>

```

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

```

1157 <<More package options>> ≡
1158 \let\bbl@opt@strings\@nnil % accept strings=value
1159 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1160 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1161 \def\BabelStringsDefault{generic}
1162 <</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.

```

1163 \@onlypreamble\StartBabelCommands
1164 \def\StartBabelCommands{%

```

```

1165 \begingroup
1166 \bbl@recatcode{11}%
1167 <⟨Macros local to BabelCommands⟩
1168 \def\bbl@provstring##1##2{%
1169   \providecommand##1{##2}%
1170   \bbl@tglobal##1}%
1171 \global\let\bbl@scafter\@empty
1172 \let\StartBabelCommands\bbl@startcmds
1173 \ifx\BabelLanguages\relax
1174   \let\BabelLanguages\CurrentOption
1175 \fi
1176 \begingroup
1177 \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
1178 \StartBabelCommands}
1179 \def\bbl@startcmds{%
1180   \ifx\bbl@screset\@nnil\else
1181     \bbl@usehooks{stopcommands}{}%
1182   \fi
1183 \endgroup
1184 \begingroup
1185 \@ifstar
1186   {\ifx\bbl@opt@strings\@nnil
1187     \let\bbl@opt@strings\BabelStringsDefault
1188   \fi
1189   \bbl@startcmds@i}%
1190 \bbl@startcmds@i}
1191 \def\bbl@startcmds@i#1#2{%
1192   \edef\bbl@L{\zap@space#1 \@empty}%
1193   \edef\bbl@G{\zap@space#2 \@empty}%
1194   \bbl@startcmds@ii}

```

Parse the encoding info to get the label, input, and font parts.

Select the behaviour 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.

```

1195 \newcommand\bbl@startcmds@ii[1][\@empty]{%
1196   \let\SetString\@gobbletwo
1197   \let\bbl@stringdef\@gobbletwo
1198   \let\AfterBabelCommands\@gobble
1199   \ifx\@empty#1%
1200     \def\bbl@sc@label{generic}%
1201     \def\bbl@encstring##1##2{%
1202       \ProvideTextCommandDefault##1{##2}%
1203       \bbl@tglobal##1%
1204       \expandafter\bbl@tglobal\csname\string?\string##1\endcsname}%
1205     \let\bbl@sctest\in@true
1206   \else
1207     \let\bbl@sc@charset\space % <- zapped below
1208     \let\bbl@sc@fontenc\space % <- " "
1209     \def\bbl@tempa##1=##2\@nil{%
1210       \bbl@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
1211     \bbl@vforeach{label=#1}{\bbl@tempa##1\@nil}%

```



```

1212 \def\bbl@tempa##1 ##2{% space -> comma
1213   ##1%
1214   \ifx\@empty##2\else\ifx,##1,\else,\fi\bbl@afterfi\bbl@tempa##2\fi}%
1215 \edef\bbl@sc@fontenc{\expandafter\bbl@tempa\bbl@sc@fontenc\@empty}%
1216 \edef\bbl@sc@label{\expandafter\zap@space\bbl@sc@label\@empty}%
1217 \edef\bbl@sc@charset{\expandafter\zap@space\bbl@sc@charset\@empty}%
1218 \def\bbl@encstring##1##2{%
1219   \bbl@foreach\bbl@sc@fontenc{%
1220     \bbl@ifunset{T####1}%
1221     {}%
1222     {\ProvideTextCommand##1{####1}{##2}%
1223     \bbl@tglobal##1%
1224     \expandafter
1225     \bbl@tglobal\csname####1\string##1\endcsname}}}%
1226 \def\bbl@sctest{%
1227   \bbl@xin@{\bbl@opt@strings,}{,\bbl@sc@label,\bbl@sc@fontenc,}}%
1228 \fi
1229 \ifx\bbl@opt@strings\@nnil % ie, no strings key -> defaults
1230 \else\ifx\bbl@opt@strings\relax % ie, strings=encoded
1231   \let\AfterBabelCommands\bbl@aftercmds
1232   \let\SetString\bbl@setstring
1233   \let\bbl@stringdef\bbl@encstring
1234 \else % ie, strings=value
1235   \bbl@sctest
1236 \fin@
1237   \let\AfterBabelCommands\bbl@aftercmds
1238   \let\SetString\bbl@setstring
1239   \let\bbl@stringdef\bbl@provstring
1240 \fi\fi\fi
1241 \bbl@scswitch
1242 \ifx\bbl@G\@empty
1243   \def\SetString##1##2{%
1244     \bbl@error{Missing group for string \string##1}%
1245     {You must assign strings to some category, typically\\%
1246     captions or extras, but you set none}}%
1247 \fi
1248 \ifx\@empty#1%
1249   \bbl@usehooks{defaultcommands}{}%
1250 \else
1251   \@expandtwoargs
1252   \bbl@usehooks{encodedcommands}{\bbl@sc@charset}\bbl@sc@fontenc}%
1253 \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).

```

1254 \def\bbl@forlang#1#2{%
1255   \bbl@for#1\bbl@L{%
1256     \bbl@xin@{,##1,}{,\BabelLanguages,}%
1257     \ifin@#2\relax\fi}}
1258 \def\bbl@scswitch{%
1259   \bbl@forlang\bbl@tempa{%
1260     \ifx\bbl@G\@empty\else

```

```

1261 \ifx\SetString\@gobbletwo\else
1262 \edef\bbl@GL{\bbl@G\bbl@tempa}%
1263 \bbl@xin@{,\bbl@GL,}{,\bbl@screset,}%
1264 \ifin@else
1265 \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
1266 \xdef\bbl@screset{\bbl@screset,\bbl@GL}%
1267 \fi
1268 \fi
1269 \fi}}
1270 \AtEndOfPackage{%
1271 \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{}{#2}}}%
1272 \let\bbl@scswitch\relax}
1273 \@onlypreamble\EndBabelCommands
1274 \def\EndBabelCommands{%
1275 \bbl@usehooks{stopcommands}{}%
1276 \endgroup
1277 \endgroup
1278 \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.

```

1279 \def\bbl@setstring#1#2{%
1280 \bbl@forlang\bbl@tempa{%
1281 \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
1282 \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
1283 {\global\expandafter % TODO - con \bbl@exp ?
1284 \bbl@add\csname\bbl@G\bbl@tempa\expandafter\endcsname\expandafter
1285 {\expandafter\bbl@scset\expandafter#1\csname\bbl@LC\endcsname}}}%
1286 }%
1287 \def\BabelString{#2}%
1288 \bbl@usehooks{stringprocess}{}%
1289 \expandafter\bbl@stringdef
1290 \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.

```

1291 \ifx\bbl@opt@strings\relax
1292 \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
1293 \bbl@patchuclc
1294 \let\bbl@encoded\relax
1295 \def\bbl@encoded@uclc#1{%
1296 \@inmathwarn#1%
1297 \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
1298 \expandafter\ifx\csname ?\string#1\endcsname\relax
1299 \TextSymbolUnavailable#1%
1300 \else
1301 \csname ?\string#1\endcsname
1302 \fi
1303 \else
1304 \csname\cf@encoding\string#1\endcsname
1305 \fi}

```

```

1306 \else
1307   \def\bbl@scset#1#2{\def#1{#2}}
1308 \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.

```

1309 <<*Macros local to BabelCommands>> ≡
1310 \def\SetStringLoop##1##2{%
1311   \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
1312   \count@\z@
1313   \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
1314     \advance\count@\@ne
1315     \toks@\expandafter{\bbl@tempa}%
1316     \bbl@exp{%
1317       \\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
1318       \count@=\the\count@\relax}}}%
1319 <</Macros local to BabelCommands>>

```

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

```

1320 \def\bbl@aftercmds#1{%
1321   \toks@\expandafter{\bbl@scafter#1}%
1322   \xdef\bbl@scafter{\the\toks@}}

```

Case mapping The command `\SetCase` provides a way to change the behaviour of `\MakeUppercase` and `\MakeLowercase`. `\bbl@tempa` is set by the patched `\@uclclist` to the parsing command.

```

1323 <<*Macros local to BabelCommands>> ≡
1324 \newcommand\SetCase[3][]{%
1325   \bbl@patchuclc
1326   \bbl@forlang\bbl@tempa{%
1327     \expandafter\bbl@encstring
1328     \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
1329     \expandafter\bbl@encstring
1330     \csname\bbl@tempa @bbl@uc\endcsname{##2}%
1331     \expandafter\bbl@encstring
1332     \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
1333 <</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.

```

1334 <<*Macros local to BabelCommands>> ≡
1335 \newcommand\SetHyphenMap[1]{%
1336   \bbl@forlang\bbl@tempa{%
1337     \expandafter\bbl@stringdef
1338     \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
1339 <</Macros local to BabelCommands>>

```

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

```

1340 \newcommand\BabelLower[2]{% one to one.
1341   \ifnum\lccode#1=#2\else
1342     \babel@savevariable{\lccode#1}%
1343     \lccode#1=#2\relax
1344   \fi}
1345 \newcommand\BabelLowerMM[4]{% many-to-many

```

```

1346 \@tempcnta=#1\relax
1347 \@tempcntb=#4\relax
1348 \def\bbl@tempa{%
1349   \ifnum\@tempcnta>#2\else
1350     \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
1351     \advance\@tempcnta#3\relax
1352     \advance\@tempcntb#3\relax
1353     \expandafter\bbl@tempa
1354   \fi}%
1355 \bbl@tempa}
1356 \newcommand\BabelLowerM0[4]{% many-to-one
1357 \@tempcnta=#1\relax
1358 \def\bbl@tempa{%
1359   \ifnum\@tempcnta>#2\else
1360     \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
1361     \advance\@tempcnta#3
1362     \expandafter\bbl@tempa
1363   \fi}%
1364 \bbl@tempa}

```

The following package options control the behaviour of hyphenation mapping.

```

1365 <<{*More package options}>> ≡
1366 \DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
1367 \DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\@ne}
1368 \DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
1369 \DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@@}
1370 \DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}
1371 <</More package options>>

```

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

```

1372 \AtEndOfPackage{%
1373   \ifx\bbl@opt@hyphenmap\undefined
1374     \bbl@xin@{,}{\bbl@language@opts}%
1375     \chardef\bbl@opt@hyphenmap\ifin@4\else\@ne\fi
1376   \fi}

```

8.10 Macros common to a number of languages

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

```

1377 \def\set@low@box#1{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}%
1378   \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@%
1379   \setbox\z@\hbox{\lower\dimen\z@ \box\z@}\ht\z@\ht\tw@ \dp\z@\dp\tw@}

```

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

```

1380 \def\save@sf@q#1{\leavevmode
1381   \begingroup
1382     \edef\@SF{\spacefactor\the\spacefactor}#1\@SF
1383   \endgroup}

```

8.11 Making glyphs available

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

8.11.1 Quotation marks

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

```
1384 \ProvideTextCommand{\quotedblbase}{OT1}{%
1385   \save@sf@q{\set@low@box{\textquotedblright\}%
1386     \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.

```
1387 \ProvideTextCommandDefault{\quotedblbase}{%
1388   \UseTextSymbol{OT1}{\quotedblbase}}
```

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

```
1389 \ProvideTextCommand{\quotesinglbase}{OT1}{%
1390   \save@sf@q{\set@low@box{\textquoteright\}%
1391     \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.

```
1392 \ProvideTextCommandDefault{\quotesinglbase}{%
1393   \UseTextSymbol{OT1}{\quotesinglbase}}
```

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

```
\guillemotright 1394 \ProvideTextCommand{\guillemotleft}{OT1}{%
1395   \ifmmode
1396     \ll
1397   \else
1398     \save@sf@q{\nobreak
1399       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bb1@allowhyphens}%
1400   \fi}
1401 \ProvideTextCommand{\guillemotright}{OT1}{%
1402   \ifmmode
1403     \gg
1404   \else
1405     \save@sf@q{\nobreak
1406       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bb1@allowhyphens}%
1407   \fi}
```

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

```
1408 \ProvideTextCommandDefault{\guillemotleft}{%
1409   \UseTextSymbol{OT1}{\guillemotleft}}
1410 \ProvideTextCommandDefault{\guillemotright}{%
1411   \UseTextSymbol{OT1}{\guillemotright}}
```

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

```
\guilsinglright 1412 \ProvideTextCommand{\guilsinglleft}{OT1}{%
1413   \ifmmode
1414     <%
1415   \else
1416     \save@sf@q{\nobreak
1417       \raise.2ex\hbox{$\scriptscriptstyle<$}\bb1@allowhyphens}%
1418   \fi}
1419 \ProvideTextCommand{\guilsinglright}{OT1}{%
1420   \ifmmode
```

```

1421 >%
1422 \else
1423 \save@sf@q{\nobreak
1424 \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
1425 \fi}

```

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

```

1426 \ProvideTextCommandDefault{\guilsinglleft}{%
1427 \UseTextSymbol{OT1}{\guilsinglleft}}
1428 \ProvideTextCommandDefault{\guilsinglright}{%
1429 \UseTextSymbol{OT1}{\guilsinglright}}

```

8.11.2 Letters

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

```

1430 \DeclareTextCommand{\ij}{OT1}{%
1431 i\kern-0.02em\bbl@allowhyphens j}
1432 \DeclareTextCommand{\IJ}{OT1}{%
1433 I\kern-0.02em\bbl@allowhyphens J}
1434 \DeclareTextCommand{\ij}{T1}{\char188}
1435 \DeclareTextCommand{\IJ}{T1}{\char156}

```

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

```

1436 \ProvideTextCommandDefault{\ij}{%
1437 \UseTextSymbol{OT1}{\ij}}
1438 \ProvideTextCommandDefault{\IJ}{%
1439 \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).

```

1440 \def\crrtic@{\hrule height0.1ex width0.3em}
1441 \def\crttic@{\hrule height0.1ex width0.33em}
1442 \def\ddj@{%
1443 \setbox0\hbox{d}\dimen@=\ht0
1444 \advance\dimen@1ex
1445 \dimen@.45\dimen@
1446 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1447 \advance\dimen@ii.5ex
1448 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
1449 \def\DDJ@{%
1450 \setbox0\hbox{D}\dimen@=.55\ht0
1451 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1452 \advance\dimen@ii.15ex % correction for the dash position
1453 \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
1454 \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
1455 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
1456 %
1457 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
1458 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}

```

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

```

1459 \ProvideTextCommandDefault{\dj}{%
1460   \UseTextSymbol{OT1}{\dj}}
1461 \ProvideTextCommandDefault{\DJ}{%
1462   \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.

```

1463 \DeclareTextCommand{\SS}{OT1}{SS}
1464 \ProvideTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}

```

8.11.3 Shorthands for quotation marks

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

\glq The ‘german’ single quotes.

```

\grq 1465 \ProvideTextCommandDefault{\glq}{%
      1466   \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}

```

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

```

1467 \ProvideTextCommand{\grq}{T1}{%
1468   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
1469 \ProvideTextCommand{\grq}{TU}{%
1470   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
1471 \ProvideTextCommand{\grq}{OT1}{%
1472   \save@sf@q{\kern-.0125em
1473     \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
1474     \kern.07em\relax}}
1475 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}{\grq}}

```

\glqq The ‘german’ double quotes.

```

\grqq 1476 \ProvideTextCommandDefault{\glqq}{%
      1477   \textormath{\quotedblbase}{\mbox{\quotedblbase}}}

```

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

```

1478 \ProvideTextCommand{\grqq}{T1}{%
1479   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1480 \ProvideTextCommand{\grqq}{TU}{%
1481   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1482 \ProvideTextCommand{\grqq}{OT1}{%
1483   \save@sf@q{\kern-.07em
1484     \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
1485     \kern.07em\relax}}
1486 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}{\grqq}}

```

\flq The ‘french’ single guillemets.

```

\frq 1487 \ProvideTextCommandDefault{\flq}{%
      1488   \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
      1489 \ProvideTextCommandDefault{\frq}{%
      1490   \textormath{\guilsinglright}{\mbox{\guilsinglright}}}

```

`\flqq` The ‘french’ double guillemets.

```
\frqq 1491 \ProvideTextCommandDefault{\flqq}{%
1492   \textormath{\guillemotleft}{\mbox{\guillemotleft}}}
1493 \ProvideTextCommandDefault{\frqq}{%
1494   \textormath{\guillemotright}{\mbox{\guillemotright}}}
```

8.11.4 Umlauts and tremas

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

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

```
1495 \def\umlauthigh{%
1496   \def\bbl@umlauta##1{\leavevmode\bgroup%
1497     \expandafter\accent\csname\fontencoding dqpos\endcsname
1498     ##1\bbl@allowhyphens\egroup}%
1499   \let\bbl@umlaute\bbl@umlauta}
1500 \def\umlautlow{%
1501   \def\bbl@umlauta{\protect\lower@umlaut}}
1502 \def\umlautelower{%
1503   \def\bbl@umlaute{\protect\lower@umlaut}}
1504 \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.

```
1505 \expandafter\ifx\csname U@D\endcsname\relax
1506   \csname newdimen\endcsname\U@D
1507 \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.

```
1508 \def\lower@umlaut#1{%
1509   \leavevmode\bgroup
1510   \U@D 1ex%
1511   {\setbox\z@\hbox{%
1512     \expandafter\char\csname\fontencoding dqpos\endcsname}%
1513     \dimen@ -.45ex\advance\dimen@\ht\z@
1514     \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
1515   \expandafter\accent\csname\fontencoding dqpos\endcsname
1516   \fontdimen5\font\U@D #1%
1517   \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).

```

1518 \AtBeginDocument{%
1519   \DeclareTextCompositeCommand{"}{OT1}{a}{\bbl@umlauta{a}}%
1520   \DeclareTextCompositeCommand{"}{OT1}{e}{\bbl@umlaute{e}}%
1521   \DeclareTextCompositeCommand{"}{OT1}{i}{\bbl@umlaute{i}}%
1522   \DeclareTextCompositeCommand{"}{OT1}{\i}{\bbl@umlaute{i}}%
1523   \DeclareTextCompositeCommand{"}{OT1}{o}{\bbl@umlauta{o}}%
1524   \DeclareTextCompositeCommand{"}{OT1}{u}{\bbl@umlauta{u}}%
1525   \DeclareTextCompositeCommand{"}{OT1}{A}{\bbl@umlauta{A}}%
1526   \DeclareTextCompositeCommand{"}{OT1}{E}{\bbl@umlaute{E}}%
1527   \DeclareTextCompositeCommand{"}{OT1}{I}{\bbl@umlaute{I}}%
1528   \DeclareTextCompositeCommand{"}{OT1}{O}{\bbl@umlauta{O}}%
1529   \DeclareTextCompositeCommand{"}{OT1}{U}{\bbl@umlauta{U}}%
1530 }

```

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

```

1531 \ifx\l@english\@undefined
1532   \chardef\l@english\z@
1533 \fi
1534 \main@language{english}

```

8.12 Layout

Work in progress.

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

```

1535 \providecommand\IfBabelLayout[3]{#3}%
1536 \newcommand\BabelPatchSection[1]{%
1537   \ifundefined{#1}{%
1538     \bbl@exp{\let<bbl@ss@#1>\<#1>}%
1539     \@namedef{#1}{%
1540       \ifstar{\bbl@presec@s{#1}}%
1541       {\@dblarg{\bbl@presec@x{#1}}}}%
1542   \def\bbl@presec@x#1[#2]#3{%
1543     \let\bbl@savetemplang\languagename
1544     \select@language@x\bbl@main@language
1545     \@nameuse{bbl@ss@#1}%
1546     [\foreignlanguage{\bbl@savetemplang}{#2}]%
1547     {\foreignlanguage{\bbl@savetemplang}{#3}}%
1548     \select@language@x\bbl@savetemplang}
1549   \def\bbl@presec@s#1#2{%
1550     \let\bbl@savetemplang\languagename
1551     \select@language@x\bbl@main@language
1552     \@nameuse{bbl@ss@#1}*{\foreignlanguage{\languagename}{#2}}%
1553     \select@language@x\bbl@savetemplang}
1554 \IfBabelLayout{sectioning}% at begin document ???
1555 {\BabelPatchSection{part}%
1556   \BabelPatchSection{chapter}%
1557   \BabelPatchSection{section}%
1558   \BabelPatchSection{subsection}%
1559   \BabelPatchSection{subsubsection}%
1560   \BabelPatchSection{paragraph}%
1561   \BabelPatchSection{subparagraph}}{}

```

Now we load definition files for engines.

```
1562 \ifcase\bbl@engine
1563   \input txtbabel.def
1564 \or
1565   \input luababel.def
1566 \or
1567   \input xebabel.def
1568 \fi
```

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

9.1 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`.

```
1569 {\def\format{lplain}
1570 \ifx\fmtname\format
1571 \else
1572   \def\format{LaTeX2e}
1573   \ifx\fmtname\format
1574   \else
1575     \aftergroup\endinput
1576   \fi
1577 \fi}
```

9.2 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.

```
1578 \newcommand\babelprovide[2][\%
1579   \let\bbl@savelangname\languagenam
1580   \def\languagenam{#2}%
1581   \let\bbl@KVP@captions\@nil
1582   \let\bbl@KVP@import\@nil
1583   \let\bbl@KVP@main\@nil
1584   \let\bbl@KVP@script\@nil
1585   \let\bbl@KVP@language\@nil
1586   \let\bbl@KVP@dir\@nil
1587   \let\bbl@KVP@hyphenrules\@nil
1588   \bbl@forkv{#1}{\bbl@csarg\def{KVP@##1}{##2}}% TODO - error handling
1589   \ifx\bbl@KVP@captions\@nil
1590     \let\bbl@KVP@captions\bbl@KVP@import
1591   \fi
1592   \bbl@ifunset{date#2}%
1593   {\bbl@provide@new{#2}}%
1594   {\bbl@ifblank{#1}%
1595     {\bbl@error
1596       {If you want to modify `#2' you must tell how in\\%
1597       the optional argument. Currently there are three\\%
```

```

1598     options: captions=lang-tag, hyphenrules=lang-list\\%
1599     import=lang-tag}%
1600     {Use this macro as documented}}%
1601     {\bbl@provide@renew{#2}}}%
1602 \bbl@exp{\bbl@babelensure[exclude=\\today]{#2}}%
1603 \bbl@ifunset{\bbl@ensure@\\language}%
1604     {\bbl@exp{%
1605         \\DeclareRobustCommand\<bbl@ensure@\\language>[1]{%
1606             \\foreignlanguage{\\language}%
1607             {###1}}}%
1608     }%
1609 \ifx\bbl@KVP@script\@nil\else
1610     \bbl@csarg\edef{sname@#2}{\bbl@KVP@script}%
1611 \fi
1612 \ifx\bbl@KVP@language\@nil\else
1613     \bbl@csarg\edef{lname@#2}{\bbl@KVP@language}%
1614 \fi
1615 \let\\language\bbl@savelangname}

```

Depending on whether or not the language exists, we define two macros.

```

1616 \def\bbl@provide@new#1{%
1617     \@namedef{date#1}{}% marks lang exists - required by \StartBabelCommands
1618     \@namedef{extras#1}{}%
1619     \@namedef{noextras#1}{}%
1620     \StartBabelCommands*{#1}{captions}%
1621     \ifx\bbl@KVP@captions\@nil % and also if import, implicit
1622         \def\bbl@tempb##1{% elt for \bbl@captionslist
1623             \ifx##1\@empty\else
1624                 \bbl@exp{%
1625                     \\SetString\\##1{%
1626                         \\bbl@nocaption{\bbl@stripslash##1}{\<#1\bbl@stripslash##1>}}%
1627                     \expandafter\bbl@tempb
1628                 \fi}%
1629             \expandafter\bbl@tempb\bbl@captionslist\@empty
1630         \else
1631             \bbl@read@ini{\bbl@KVP@captions}% Here all letters cat = 11
1632             \bbl@after@ini
1633             \bbl@savestrings
1634         \fi
1635     \StartBabelCommands*{#1}{date}%
1636     \ifx\bbl@KVP@import\@nil
1637         \bbl@exp{%
1638             \\SetString\\today{\bbl@nocaption{today}{\<#1today>}}}%
1639         \else
1640             \bbl@savetoday
1641             \bbl@savedate
1642         \fi
1643     \EndBabelCommands
1644     \bbl@exp{%
1645         \def\<#1hyphenmins>{%
1646             {\bbl@ifunset{\bbl@lfthm@#1}{2}{\@nameuse{\bbl@lfthm@#1}}}%
1647             {\bbl@ifunset{\bbl@rgthm@#1}{3}{\@nameuse{\bbl@rgthm@#1}}}%
1648             \bbl@provide@hyphens{#1}%
1649         \ifx\bbl@KVP@main\@nil\else
1650             \expandafter\main@language\expandafter{#1}%
1651         \fi
1652     \def\bbl@provide@renew#1{%
1653         \ifx\bbl@KVP@captions\@nil\else
1654             \StartBabelCommands*{#1}{captions}%

```

```

1655 \bbl@read@ini{\bbl@KVP@captions}% Here all letters cat = 11
1656 \bbl@after@ini
1657 \bbl@savestrings
1658 \EndBabelCommands
1659 \fi
1660 \ifx\bbl@KVP@import\@nil\else
1661 \StartBabelCommands*{#1}{date}%
1662 \bbl@savetoday
1663 \bbl@savestate
1664 \EndBabelCommands
1665 \fi
1666 \bbl@provide@hyphens{#1}}

```

The hyphenrules option is handled with an auxiliary macro.

```

1667 \def\bbl@provide@hyphens#1{%
1668 \let\bbl@tempa\relax
1669 \ifx\bbl@KVP@hyphenrules\@nil\else
1670 \bbl@replace\bbl@KVP@hyphenrules{ }{,}%
1671 \bbl@foreach\bbl@KVP@hyphenrules{%
1672 \ifx\bbl@tempa\relax % if not yet found
1673 \bbl@ifsamestring{##1}{+}%
1674 {\bbl@exp{\addlanguage\<l@##1>}}}%
1675 }%
1676 \bbl@ifunset{l@##1}%
1677 }%
1678 {\bbl@exp{\let\bbl@tempa\<l@##1>}}%
1679 \fi}%
1680 \fi
1681 \ifx\bbl@tempa\relax % if no opt or no language in opt found
1682 \ifx\bbl@KVP@import\@nil\else % if importing
1683 \bbl@exp{% and hyphenrules is not empty
1684 \bbl@ifblank{\@nameuse{bbl@hyphr@#1}}%
1685 }%
1686 {\let\bbl@tempa\<l@\@nameuse{bbl@hyphr@\language}\>}}%
1687 \fi
1688 \fi
1689 \bbl@ifunset{bbl@tempa}% ie, relax or undefined
1690 {\bbl@ifunset{l@#1}% no hyphenrules found - fallback
1691 {\bbl@exp{\adddialect\<l@#1>\language}}%
1692 }% so, l@<lang> is ok - nothing to do
1693 {\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.*

```

1694 \def\bbl@read@ini#1{%
1695 \openin1=babel-#1.ini
1696 \ifeof1
1697 \bbl@error
1698 {There is no ini file for the requested language\%
1699 (#1). Perhaps you misspelled it or your installation\%
1700 is not complete.}%
1701 {Fix the name or reinstall babel.}%
1702 \else
1703 \let\bbl@section\@empty
1704 \let\bbl@savestrings\@empty
1705 \let\bbl@savetoday\@empty
1706 \let\bbl@savestate\@empty
1707 \let\bbl@inireader\bbl@iniskip
1708 \bbl@info{Importing data from babel-#1.ini for \language}%

```

```

1709 \loop
1710 \endlinechar\m@ne
1711 \read1 to \bbl@line
1712 \endlinechar\^^M
1713 \if T\ifeof1F\fi T\relax % Trick, because inside \loop
1714 \ifx\bbl@line\empty\else
1715 \expandafter\bbl@iniline\bbl@line\bbl@iniline
1716 \fi
1717 \repeat
1718 \fi}
1719 \def\bbl@iniline#1\bbl@iniline{%
1720 \@ifnextchar[\bbl@inisecl{\@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.

```

1721 \def\bbl@iniskip#1\@{}% if starts with ;
1722 \def\bbl@inisecl[#1]#2\@{}% if starts with opening bracket
1723 \@nameuse{\bbl@secpost@\bbl@section}% ends previous section
1724 \def\bbl@section{#1}%
1725 \@nameuse{\bbl@secpre@\bbl@section}% starts current section
1726 \bbl@ifunset{\bbl@secline@#1}%
1727 {\let\bbl@inireader\bbl@iniskip}%
1728 {\bbl@exp{\let\\bbl@inireader<\bbl@secline@#1>}}}

```

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

```

1729 \def\bbl@inikv#1=#2\@{}% key=value
1730 \bbl@trim@def\bbl@tempa{#1}%
1731 \bbl@trim\toks@{#2}%
1732 \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.

```

1733 \def\bbl@exportkey#1#2#3{%
1734 \bbl@ifunset{\bbl@kv@#2}%
1735 {\bbl@csarg\gdef{#1@\language}\{#3}}%
1736 {\expandafter\ifx\csname\bbl@kv@#2\endcsname\empty
1737 \bbl@csarg\gdef{#1@\language}\{#3}}%
1738 \else
1739 \bbl@exp{\global\let<\bbl@#1@\language>\<\bbl@kv@#2>}%
1740 \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.

```

1741 \let\bbl@secline@identification\bbl@inikv
1742 \def\bbl@secpost@identification{%
1743 \bbl@exportkey{lname}{identification.name.english}{}%
1744 \bbl@exportkey{lbcpr}{identification.tag.bcp47}{}%
1745 \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
1746 \bbl@exportkey{sname}{identification.script.name}{}%
1747 \bbl@exportkey{sbcpr}{identification.script.tag.bcp47}{}%
1748 \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}}
1749 \let\bbl@secline@typography\bbl@inikv
1750 \def\bbl@after@ini{%
1751 \bbl@exportkey{lftm}{typography.lefthyphenmin}{2}%
1752 \bbl@exportkey{rgtm}{typography.righthyphenmin}{3}%
1753 \bbl@exportkey{hyphr}{typography.hyphenrules}{}%
1754 \def\bbl@tempa{0.9}%
1755 \bbl@csarg\ifx{\kv@identification.version}\bbl@tempa

```

```

1756 \bbl@warning{%
1757     The '\language' date format may not be suitable\\%
1758     for proper typesetting, and therefore it very likely will\\%
1759     change in a future release. Reported}%
1760 \fi
1761 \bbl@toglobal\bbl@savetoday
1762 \bbl@toglobal\bbl@savestate}

```

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

```

1763 \ifcase\bbl@engine
1764 \bbl@csarg\def{secline@captions.licr}#1=#2\@@{%
1765     \bbl@ini@captions@aux{#1}{#2}}
1766 \bbl@csarg\def{secline@date.gregorian}#1=#2\@@{%           for defaults
1767     \bbl@ini@dategreg#1...\relax{#2}}
1768 \bbl@csarg\def{secline@date.gregorian.licr}#1=#2\@@{%      override
1769     \bbl@ini@dategreg#1...\relax{#2}}
1770 \else
1771 \def\bbl@secline@captions#1=#2\@@{%
1772     \bbl@ini@captions@aux{#1}{#2}}
1773 \bbl@csarg\def{secline@date.gregorian}#1=#2\@@{%
1774     \bbl@ini@dategreg#1...\relax{#2}}
1775 \fi

```

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

```

1776 \def\bbl@ini@captions@aux#1#2{%
1777     \bbl@trim@def\bbl@tempa{#1}%
1778     \bbl@ifblank{#2}%
1779     {\bbl@exp{%
1780         \toks{\bbl@nocaption{\bbl@tempa}\<\language\bbl@tempa name>}}}%
1781     {\bbl@trim\toks{#2}}}%
1782 \bbl@exp{%
1783     \bbl@add\bbl@savestrings{%
1784         \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.

```

1785 \newif\ifbbl@usedategroup
1786 \bbl@csarg\def{secpred@date.gregorian.licr}{%
1787     \ifcase\bbl@engine\let\bbl@savestate\empty\fi}
1788 \def\bbl@ini@dategreg#1.#2.#3.#4\relax#5{% TODO - ignore with 'captions'
1789     \bbl@trim@def\bbl@tempa{#1.#2}%
1790     \bbl@ifsamestring{\bbl@tempa}{months.wide}%
1791     {\bbl@trim@def\bbl@tempa{#3}%
1792         \bbl@trim\toks{#5}%
1793         \bbl@exp{%
1794             \bbl@add\bbl@savestate{%
1795                 \SetString\<month\romannumeral\bbl@tempa name>{\the\toks}}}%
1796         {\bbl@ifsamestring{\bbl@tempa}{date.long}%
1797             {\bbl@trim@def\bbl@toreplace{#5}%
1798                 \bbl@TG@date
1799                 \global\bbl@csarg\let{date@\language}\bbl@toreplace
1800                 \bbl@exp{%
1801                     \gdef\<\language date>{\protect\<\language date >}}%
1802                     \gdef\<\language date >####1####2####3{%
1803                         \bbl@usedategroupttrue
1804                         \<\language>{
1805                             \bbl@date@\language>{####1}{####2}{####3}}}%
1806                     \bbl@add\bbl@savetoday}%

```

```

1807          \\SetString\\today{%
1808          <\language name date>{\\the\year}{\\the\month}{\\the\day}}}%
1809      {}

```

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.

```

1810 \newcommand\BabelDateSpace{\nobreakspace}
1811 \newcommand\BabelDateDot{. \@}
1812 \newcommand\BabelDated[1]{\number#1}
1813 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
1814 \newcommand\BabelDateM[1]{\number#1}
1815 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}
1816 \newcommand\BabelDateMMM[1]{%
1817   \csname month\romannumeral#1name\endcsname}%
1818 \newcommand\BabelDateY[1]{\number#1}%
1819 \newcommand\BabelDateyy[1]{%
1820   \ifnum#1<10 0\number#1 %
1821   \else\ifnum#1<100 \number#1 %
1822   \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
1823   \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
1824   \else
1825     \bbl@error
1826     {Currently two-digit years are restricted to the\
1827     range 0-9999.}%
1828     {There is little you can do. Sorry.}%
1829   \fi\fi\fi\fi}
1830 \newcommand\BabelDateyyyy[1]{\number#1}
1831 \def\bbl@replace@finish@iii#1{%
1832   \bbl@exp{\def\#1####1####2####3{\the\toks@}}
1833 \def\bbl@TG@date{%
1834   \bbl@replace\bbl@toreplace{[ ]}{\BabelDateSpace}}%
1835   \bbl@replace\bbl@toreplace{[.]}{\BabelDateDot}}%
1836   \bbl@replace\bbl@toreplace{[d]}{\BabelDated{####3}}%
1837   \bbl@replace\bbl@toreplace{[dd]}{\BabelDatedd{####3}}%
1838   \bbl@replace\bbl@toreplace{[M]}{\BabelDateM{####2}}%
1839   \bbl@replace\bbl@toreplace{[MM]}{\BabelDateMM{####2}}%
1840   \bbl@replace\bbl@toreplace{[MMM]}{\BabelDateMMM{####2}}%
1841   \bbl@replace\bbl@toreplace{[y]}{\BabelDateY{####1}}%
1842   \bbl@replace\bbl@toreplace{[yy]}{\BabelDateyy{####1}}%
1843   \bbl@replace\bbl@toreplace{[yyyy]}{\BabelDateyyyy{####1}}%
1844 % Note after \bbl@replace \toks@ contains the resulting string.
1845 % TODO - Using this implicit behavior doesn't seem a good idea.
1846   \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.

```

1847 \def\bbl@provide@lsys#1{%
1848   \bbl@ifunset{\bbl@lname@#1}%
1849   {\bbl@ini@ids{#1}}%
1850   {}%
1851   \bbl@csarg\let{lsys@#1}\@empty
1852   \bbl@ifunset{\bbl@sname@#1}{\bbl@csarg\gdef{sname@#1}{Default}}}%
1853   \bbl@ifunset{\bbl@sotf@#1}{\bbl@csarg\gdef{sotf@#1}{DFLT}}}%
1854   \bbl@csarg\bbl@add@list{lsys@#1}{Script=\bbl@cs{sname@#1}}%
1855   \bbl@ifunset{\bbl@lname@#1}{}%
1856   {\bbl@csarg\bbl@add@list{lsys@#1}{Language=\bbl@cs{lname@#1}}}%
1857   \bbl@csarg\bbl@to@global{lsys@#1}}%
1858 % \bbl@exp{% TODO - should be global

```

```

1859 % \<keys_if_exist:nnF>{fontspec-opentype/Script}{\bbl@cs{sname@#1}}%
1860 % {\newfontscript{\bbl@cs{sname@#1}}{\bbl@cs{sotf@#1}}}%
1861 % \<keys_if_exist:nnF>{fontspec-opentype/Language}{\bbl@cs{lname@#1}}%
1862 % {\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.

```

1863 \def\bbl@ini@ids#1{%
1864   \def\BabelBeforeIni##1##2{%
1865     \begingroup
1866     \bbl@add\bbl@secpost@identification{%
1867       \def\bbl@inline#####1\bbl@inline{}}%
1868     \catcode`\[=12 \catcode`\]=12 \catcode`\==12
1869     \bbl@read@ini{##1}%
1870   \endgroup}% boxed, to avoid extra spaces:
1871   \setbox\z@\hbox{\InputIfFileExists{babel-#1.tex}{}}}%

```

9.3 Cross referencing macros

The \LaTeX book states:

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

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

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

The only way to accomplish this in most cases is to use the trick described in the \TeX book [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.

```

1872 %\bbl@redefine\newlabel#1#2{%
1873 % \@safe@activestrue\org@newlabel{#1}{#2}\@safe@activesfalse}

```

`\@newl@bel` We need to change the definition of the \LaTeX -internal macro `\@newl@bel`. This is needed because we need to make sure that shorthand characters expand to their non-active version.

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

```

1874 <<More package options>> ≡
1875 \DeclareOption{safe=none}{\let\bbl@opt@safe\@empty}
1876 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
1877 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
1878 <</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.

```

1879 \ifx\bbl@opt@safe\@empty\else
1880   \def\@newl@bel#1#2#3{%

```



```

1881 {\@safe@activetrue
1882 \bbl@ifunset{#1@#2}%
1883 \relax
1884 {\gdef\@multiplelabels{%
1885 \@latex@warning@no@line{There were multiply-defined labels}}%
1886 \@latex@warning@no@line{Label `#2' multiply defined}}%
1887 \global\@namedef{#1@#2}{#3}}

```

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

```

1888 \CheckCommand*\@testdef[3]{%
1889 \def\reserved@a{#3}%
1890 \expandafter\ifx\csname#1@#2\endcsname\reserved@a
1891 \else
1892 \@tempswatrue
1893 \fi}

```

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

```

1894 \def\@testdef#1#2#3{%
1895 \@safe@activetrue

```

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

```

1896 \expandafter\let\expandafter\bbl@tempa\csname #1@#2\endcsname

```

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

```

1897 \def\bbl@tempb{#3}%
1898 \@safe@activesfalse

```

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

```

1899 \ifx\bbl@tempa\relax
1900 \else
1901 \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
1902 \fi

```

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

```

1903 \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%

```

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

```

1904 \ifx\bbl@tempa\bbl@tempb
1905 \else
1906 \@tempswatrue
1907 \fi}
1908 \fi

```

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

```

1909 \bbl@xin@{R}\bbl@opt@safe
1910 \ifin@
1911 \bbl@redefineroobust\ref#1{%
1912 \@safe@activetrue\org@ref{#1}\@safe@activesfalse}
1913 \bbl@redefineroobust\pageref#1{%

```

```

1914 \safe@activetrue\org@pageref{#1}\safe@activesfalse}
1915 \else
1916 \let\org@ref\ref
1917 \let\org@pageref\pageref
1918 \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.

```

1919 \bbl@xin@{B}\bbl@opt@safe
1920 \ifin@
1921 \bbl@redefine\@citex[#1]#2{%
1922 \safe@activetrue\edef\@tempa{#2}\safe@activesfalse
1923 \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.

```

1924 \AtBeginDocument{%
1925 \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.)

```

1926 \def\@citex[#1][#2]#3{%
1927 \safe@activetrue\edef\@tempa{#3}\safe@activesfalse
1928 \org@@citex[#1][#2]{\@tempa}}%
1929 }{}

```

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

```

1930 \AtBeginDocument{%
1931 \ifpackageloaded{cite}{%
1932 \def\@citex[#1]#2{%
1933 \safe@activetrue\org@@citex[#1]{#2}\safe@activesfalse}%
1934 }{}

```

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

```

1935 \bbl@redefine\nocite#1{%
1936 \safe@activetrue\org@nocite{#1}\safe@activesfalse}

```

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

```

1937 \bbl@redefine\bibcite{%

```

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

```
1938 \bbl@cite@choice
1939 \biblecite}
```

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

```
1940 \def\bbl@biblecite#1#2{%
1941 \org@biblecite{#1}{\@safe@activesfalse#2}}
```

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

```
1942 \def\bbl@cite@choice{%
```

First we give `\biblecite` its default definition.

```
1943 \global\let\biblecite\bbl@biblecite
```

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

```
1944 \@ifpackageloaded{natbib}{\global\let\biblecite\org@biblecite}{}%
```

For `cite` we do the same.

```
1945 \@ifpackageloaded{cite}{\global\let\biblecite\org@biblecite}{}%
```

Make sure this only happens once.

```
1946 \global\let\bbl@cite@choice\relax}
```

When a document is run for the first time, no `.aux` file is available, and `\biblecite` will not yet be properly defined. In this case, this has to happen before the document starts.

```
1947 \AtBeginDocument{\bbl@cite@choice}
```

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

```
1948 \bbl@redefine\@bibitem#1{%
1949 \@safe@activetrue\org@@bibitem{#1}\@safe@activesfalse}
1950 \else
1951 \let\org@nocite\nocite
1952 \let\org@@citex\citex
1953 \let\org@biblecite\biblecite
1954 \let\org@@bibitem\@bibitem
1955 \fi
```

9.4 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.

```
1956 \bbl@redefine\markright#1{%
1957 \bbl@ifblank{#1}%
1958 {\org@markright}{}%
1959 {\toks@{#1}}%
1960 \bbl@exp{%
1961 \org@markright{\protect\foreignlanguage{\language}%
1962 {\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`.

```
1963 \ifx\@mkboth\markboth
1964   \def\bbl@tempc{\let\@mkboth\markboth}
1965 \else
1966   \def\bbl@tempc{}
1967 \fi
```

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

```
1968 \bbl@redefine\markboth#1#2{%
1969   \protected@edef\bbl@tempb##1{%
1970     \protect\foreignlanguage{\language}{\protect\bbl@restore@actives##1}}%
1971   \bbl@ifblank{#1}%
1972     {\toks@{}}%
1973     {\toks@\expandafter{\bbl@tempb{#1}}}%
1974   \bbl@ifblank{#2}%
1975     {\@temptokena{}}%
1976     {\@temptokena\expandafter{\bbl@tempb{#2}}}%
1977   \bbl@exp{\@org@markboth{\the\toks@}{\the\@temptokena}}}
```

and copy it to `\@mkboth` if necessary.

```
1978 \bbl@tempc
```

9.5 Preventing clashes with other packages

9.5.1 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.

```
1979 \bbl@xin@{R}\bbl@opt@safe
1980 \ifin@
1981   \AtBeginDocument{%
1982     \@ifpackageloaded{ifthen}{%
```

Then we can redefine `\ifthenelse`:

```
1983   \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.

```
1984     \let\bbl@temp@pref\pageref
1985     \let\pageref\org@pageref
1986     \let\bbl@temp@ref\ref
1987     \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.

```

1988     \@safe@activestrue
1989     \org@ifthenelse{#1}%
1990     {\let\pageref\bbl@temp@pref
1991     \let\ref\bbl@temp@ref
1992     \@safe@activesfalse
1993     #2}%
1994     {\let\pageref\bbl@temp@pref
1995     \let\ref\bbl@temp@ref
1996     \@safe@activesfalse
1997     #3}%
1998     }%
1999     }{}%
2000 }

```

9.5.2 varioref

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

```

\Ref 2001 \AtBeginDocument{%
2002     \@ifpackageloaded{varioref}{%
2003     \bbl@redefine\@vpageref#1[#2]#3{%
2004     \@safe@activestrue
2005     \org@@@vpageref{#1}[#2]{#3}%
2006     \@safe@activesfalse}%

```

The same needs to happen for `\vrefpagemum`.

```

2007     \bbl@redefine\vrefpagemum#1#2{%
2008     \@safe@activestrue
2009     \org\vrefpagemum{#1}{#2}%
2010     \@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.

```

2011     \expandafter\def\csname Ref \endcsname#1{%
2012     \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
2013     }{}%
2014 }
2015 \fi

```

9.5.3 hhline

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

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

```

2016 \AtEndOfPackage{%
2017 \AtBeginDocument{%
2018     \@ifpackageloaded{hhline}%

```

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

```
2019      {\expandafter\ifx\csname normal@char\string:\endcsname\relax
2020      \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.

```
2021      \makeatletter
2022      \def\@currname{hhline}\input{hhline.sty}\makeatother
2023      \fi}%
2024      {}}}
```

9.5.4 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.

```
2025 \AtBeginDocument{%
2026   \ifx\pdfstringdefDisableCommands\@undefined\else
2027     \pdfstringdefDisableCommands{\languageshorthands{system}}%
2028   \fi}
```

9.5.5 fancyhdr

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

```
2029 \DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
2030   \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.

```
2031 \def\substitutefontfamily#1#2#3{%
2032   \lowercase{\immediate\openout15=#1#2.fd\relax}%
2033   \immediate\write15{%
2034     \string\ProvidesFile{#1#2.fd}%
2035     [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
2036     \space generated font description file]^J
2037     \string\DeclareFontFamily{#1}{#2}{^^J
2038     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{^^J
2039     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{^^J
2040     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{^^J
2041     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{^^J
2042     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{^^J
2043     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{^^J
2044     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{^^J
2045     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{^^J
2046     }%
2047   \closeout15
2048 }
```

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

```
2049 \@onlypreamble\substitutefontfamily
```

9.6 Encoding and fonts

Because documents may use non-ASCII font encodings, we make sure that the logos of \TeX and \LaTeX always come out in the right encoding. There is a list of non-ASCII encodings. Unfortunately, fontenc deletes its package options, so we must guess which encodings has been loaded by traversing `\@filelist` to search for `\enc.def`. If a non-ASCII has been loaded, we define versions of \TeX and \LaTeX for them using `\ensureascii`. The default ASCII encoding is set, too (in reverse order): the “main” encoding (when the document begins), the last loaded, or OT1.

`\ensureascii`

```
2050 \newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU,}
2051 \let\org@TeX\TeX
2052 \let\org@LaTeX\LaTeX
2053 \let\ensureascii\@firstofone
2054 \AtBeginDocument{%
2055   \in@false
2056   \bbl@foreach\BabelNonASCII{% is there a non-ascii enc?
2057     \ifin@ \else
2058       \lowercase{\bbl@xin@{,#1enc.def},{, \@filelist,}}%
2059       \fi}%
2060   \ifin@ % if a non-ascii has been loaded
2061     \def\ensureascii#1{{\fontencoding{OT1}\selectfont#1}}%
2062     \DeclareTextCommandDefault{\TeX}{\org@TeX}%
2063     \DeclareTextCommandDefault{\LaTeX}{\org@LaTeX}%
2064     \def\bbl@tempb#1\@@{\uppercase{\bbl@tempc#1}ENC.DEF\empty\@@}%
2065     \def\bbl@tempc#1ENC.DEF#2\@@{%
2066       \ifx\@empty#2\else
2067         \bbl@ifunset{T#1}%
2068         {}%
2069         {\bbl@xin@{,#1,},{, \BabelNonASCII,}}%
2070         \ifin@
2071           \DeclareTextCommand{\TeX}{#1}{\ensureascii{\org@TeX}}%
2072           \DeclareTextCommand{\LaTeX}{#1}{\ensureascii{\org@LaTeX}}%
2073         \else
2074           \def\ensureascii#1{{\fontencoding{#1}\selectfont#1}}%
2075           \fi}%
2076     \fi}%
2077   \bbl@foreach\@filelist{\bbl@tempb#1\@@}% TODO - \@@ de mas??
2078   \bbl@xin@{\cf@encoding,}{, \BabelNonASCII,}%
2079   \ifin@ \else
2080     \edef\ensureascii#1{%
2081       \noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}%
2082     \fi
2083   \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.

```
2084 \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.

```

2085 \AtBeginDocument{%
2086   \ifpackageloaded{fontspec}%
2087     {\xdef\latinencoding{%
2088       \ifx\UTFencname\@undefined
2089         EU\ifcase\bbl@engine\or2\or1\fi
2090       \else
2091         \UTFencname
2092       \fi}}%
2093   {\gdef\latinencoding{OT1}%
2094     \ifx\cf@encoding\bbl@t@one
2095       \xdef\latinencoding{\bbl@t@one}%
2096     \else
2097       \@ifl@aded{def}{t1enc}{\xdef\latinencoding{\bbl@t@one}}{}%
2098     \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.

```

2099 \DeclareRobustCommand{\latintext}{%
2100   \fontencoding{\latinencoding}\selectfont
2101   \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.

```

2102 \ifx\@undefined\DeclareTextFontCommand
2103   \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
2104 \else
2105   \DeclareTextFontCommand{\textlatin}{\latintext}
2106 \fi

```

9.7 Basic bidi support

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

- `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>>).

```

2107 \def\bbl@alscripts{,Arabic,Syriac,Thaana,}
2108 \def\bbl@rscripts{%
2109   ,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
2110   Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaeen,%
2111   Manichaeen,Meroitic Cursive,Meroitic,Old North Arabian,%
2112   Nabataean,N'Ko,Orkhon,Palmyrene,Inscriptional Pahlavi,%
2113   Psalter Pahlavi,Phoenician,Inscriptional Parthian,Samaritan,%

```



```

2114 Old South Arabian,}%
2115 \def\bbl@provide@dirs#1{%
2116   \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts\bbl@rscripts}%
2117   \ifin@
2118     \global\bbl@csarg\chardef{wdir@#1}\@ne
2119     \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts}%
2120     \ifin@
2121       \global\bbl@csarg\chardef{wdir@#1}\tw@ % useless in xetex
2122       \fi
2123     \else
2124       \global\bbl@csarg\chardef{wdir@#1}\z@
2125     \fi}
2126 \def\bbl@switchdir{%
2127   \bbl@ifunset{bbl@lsys@\language}{\bbl@provide@lsys{\language}}{}%
2128   \bbl@ifunset{bbl@wdir@\language}{\bbl@provide@dirs{\language}}{}%
2129   \bbl@exp{\bbl@setdirs\bbl@cs{wdir@\language}}%
2130 \def\bbl@setdirs#1{% TODO - math
2131   \ifcase\bbl@select@type % TODO - strictly, not the right test
2132     \bbl@bodydir{#1}%
2133     \bbl@pardir{#1}%
2134   \fi
2135   \bbl@textdir{#1}}
2136 \ifodd\bbl@engine % luatex=1
2137   \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
2138   \DisableBabelHook{babel-bidi}
2139   \def\bbl@getluadir#1{%
2140     \directlua{
2141       if tex.#1dir == 'TLT' then
2142         tex.sprint('0')
2143       elseif tex.#1dir == 'TRT' then
2144         tex.sprint('1')
2145       end}}
2146   \def\bbl@setdir#1#2#3{% 1=text/par.. 2=\textdir.. 3=0 lr/1 rl
2147     \ifcase#3\relax
2148       \ifcase\bbl@getluadir{#1}\relax\else
2149         #2 TLT\relax
2150       \fi
2151     \else
2152       \ifcase\bbl@getluadir{#1}\relax
2153         #2 TRT\relax
2154       \fi
2155     \fi}
2156   \def\bbl@textdir#1{%
2157     \bbl@setdir{text}\textdir{#1}% TODO - ?\linedir
2158     \setattribute\bbl@attr@dir{#1}}
2159   \def\bbl@pardir{\bbl@setdir{par}\pardir}
2160   \def\bbl@bodydir{\bbl@setdir{body}\bodydir}
2161   \def\bbl@pagedir{\bbl@setdir{page}\pagedir}
2162   \def\bbl@dirparastext{\pardir\the\textdir\relax}% %%%
2163 \else % pdftex=0, xetex=2
2164   \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
2165   \DisableBabelHook{babel-bidi}
2166   \newcount\bbl@dirlevel
2167   \chardef\bbl@thetextdir\z@
2168   \chardef\bbl@thepardir\z@
2169   \def\bbl@textdir#1{%
2170     \ifcase#1\relax
2171       \chardef\bbl@thetextdir\z@
2172       \bbl@textdir@i\beginL\endL

```

```

2173 \else
2174 \chardef\bbl@thetextdir\@ne
2175 \bbl@textdir@i\beginR\endR
2176 \fi}
2177 \def\bbl@textdir@i#1#2{%
2178 \ifhmode
2179 \ifnum\currentgrouplevel>\z@
2180 \ifnum\currentgrouplevel=\bbl@dirlevel
2181 \bbl@error{Multiple bidi settings inside a group}%
2182 {I'll insert a new group, but expect wrong results.}%
2183 \bgroup\aftergroup#2\aftergroup\egroup
2184 \else
2185 \ifcase\currentgrouptype\or % 0 bottom
2186 \aftergroup#2% 1 simple {}
2187 \or
2188 \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
2189 \or
2190 \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
2191 \or\or\or % vbox vtop align
2192 \or
2193 \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
2194 \or\or\or\or\or\or % output math disc insert vcent mathchoice
2195 \or
2196 \aftergroup#2% 14 \begingroup
2197 \else
2198 \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
2199 \fi
2200 \fi
2201 \bbl@dirlevel\currentgrouplevel
2202 \fi
2203 #1%
2204 \fi}
2205 \def\bbl@pardir#1{\chardef\bbl@thepardir#1\relax}
2206 \let\bbl@bodydir\@gobble
2207 \let\bbl@pagedir\@gobble
2208 \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).

```

2209 \def\bbl@xebidipar{%
2210 \let\bbl@xebidipar\relax
2211 \TeXeTstate\@ne
2212 \def\bbl@xeeverypar{%
2213 \ifcase\bbl@thepardir
2214 \ifcase\bbl@thetextdir\else\beginR\fi
2215 \else
2216 {\setbox\z@\lastbox\beginR\box\z@}%
2217 \fi}%
2218 \let\bbl@severypar\everypar
2219 \newtoks\everypar
2220 \everypar=\bbl@severypar
2221 \bbl@severypar{\bbl@xeeverypar\the\everypar}}
2222 \fi

```

A tool for weak L (mainly digits).

```

2223 \DeclareRobustCommand\babelsublr[1]{\leavevmode{\bbl@textdir\z@#1}}

```

9.8 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`.

```
2224 \ifx\loadlocalcfg\undefined
2225   \ifpackagewith{babel}{noconfigs}%
2226   {\let\loadlocalcfg@gobble}%
2227   {\def\loadlocalcfg#1{%
2228     \InputIfFileExists{#1.cfg}%
2229     {\typeout{*****^J%
2230               * Local config file #1.cfg used^J%
2231               *}}%
2232     \@empty}}
2233 \fi
```

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

```
2234 \ifx@unexpandable@protect\undefined
2235   \def@unexpandable@protect{\noexpand\protect\noexpand}
2236   \long\def\protected@write#1#2#3{%
2237     \begingroup
2238       \let\thepage\relax
2239       #2%
2240       \let\protect@unexpandable@protect
2241       \edef\reserved@a{\write#1{#3}}%
2242       \reserved@a
2243     \endgroup
2244     \if@nobreak\ifvmode\nobreak\fi\fi}
2245 \fi
2246 \core
```

10 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.

```
2247 <kernel>
2248 <<Make sure ProvidesFile is defined>>
2249 \ProvidesFile{switch.def}[\<date>] [\<version>] Babel switching mechanism]
2250 <<Load macros for plain if not LaTeX>>
2251 <<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.

```
2252 \def\bbl@version{\<version>}
2253 \def\bbl@date{\<date>}
2254 \def\adddialect#1#2{%
2255   \global\chardef#1#2\relax
2256   \bbl@usehooks{adddialect}{#1}{#2}%
2257   \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.

```

2258 \def\bbl@fixname#1{%
2259   \begingroup
2260   \def\bbl@tempe{l@}%
2261   \edef\bbl@tempd{\noexpand\@ifundefined{\noexpand\bbl@tempe#1}}%
2262   \bbl@tempd
2263     {\lowercase\expandafter{\bbl@tempd}%
2264      {\uppercase\expandafter{\bbl@tempd}}%
2265      \@empty
2266      {\edef\bbl@tempd{\def\noexpand#1{#1}}%
2267       \uppercase\expandafter{\bbl@tempd}}}%
2268     {\edef\bbl@tempd{\def\noexpand#1{#1}}%
2269      \lowercase\expandafter{\bbl@tempd}}}%
2270   \@empty
2271   \edef\bbl@tempd{\endgroup\def\noexpand#1{#1}}%
2272   \bbl@tempd}
2273 \def\bbl@iflanguage#1{%
2274   \@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.

```

2275 \def\iflanguage#1{%
2276   \bbl@iflanguage{#1}{%
2277     \ifnum\csname l@#1\endcsname=\language
2278       \expandafter\@firstoftwo
2279     \else
2280       \expandafter\@secondoftwo
2281     \fi}}

```

10.1 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.

```

2282 \let\bbl@select@type\z@
2283 \edef\selectlanguage{%
2284   \noexpand\protect
2285   \expandafter\noexpand\csname selectlanguage \endcsname}

```

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

```

2286 \ifx\@undefined\protect\let\protect\relax\fi

```

As \LaTeX 2.09 writes to files *expanded* whereas \LaTeX 2_ε takes care *not* to expand the arguments of `\write` statements we need to be a bit clever about the way we add information to .aux files. Therefore we introduce the macro `\xstring` which should expand to the right amount of `\string`'s.

```

2287 \ifx\documentclass\@undefined
2288   \def\xstring{\string\string\string}
2289 \else
2290   \let\xstring\string
2291 \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.

```

2292 \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`

```

2293 \def\bbl@push@language{%
2294   \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.

```

2295 \def\bbl@pop@lang#1+#2-#3{%
2296   \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.

```

2297 \let\bbl@ifrestoring\@secondoftwo
2298 \def\bbl@pop@language{%
2299   \expandafter\bbl@pop@lang\bbl@language@stack-\bbl@language@stack
2300   \let\bbl@ifrestoring\@firstoftwo
2301   \expandafter\bbl@set@language\expandafter{\language}%
2302   \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.

```

2303 \expandafter\def\csname selectlanguage \endcsname#1{%
2304   \ifnum\bbl@hymapsel=\@ccclv\let\bbl@hymapsel\tw\@fi
2305   \bbl@push@language
2306   \aftergroup\bbl@pop@language
2307   \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.

```

2308 \def\BabelContentsFiles{toc,lof,lot}
2309 \def\bbl@set@language#1{%
2310   \edef\language{%
2311     \ifnum\escapechar=\expandafter`\string#1\@empty
2312     \else\string#1\@empty\fi}%
2313   \select@language{\language}%
2314   \expandafter\ifx\csname date\language\endcsname\relax\else
2315     \if@filesw
2316       \protected@write\auxout{{\string\babel@aux{\language}}}%
2317       \bbl@usehooks{write}}}%
2318   \fi
2319   \fi}
2320 \def\select@language#1{%
2321   \ifnum\bbl@hymapsel=\@ccclv\chardef\bbl@hymapsel4\relax\fi
2322   \edef\language{#1}%
2323   \bbl@fixname\language
2324   \bbl@iflanguage\language{%
2325     \expandafter\ifx\csname date\language\endcsname\relax
2326       \bbl@error
2327       {Unknown language `#1'. Either you have\\%
2328        misspelled its name, it has not been installed,\\%
2329        or you requested it in a previous run. Fix its name,\\%
2330        install it or just rerun the file, respectively}%
2331       {You may proceed, but expect wrong results}%
2332     \else
2333       \let\bbl@select@type\z@
2334       \expandafter\bbl@switch\expandafter{\language}%
2335       \fi}}
2336 \def\babel@aux#1#2{%
2337   \select@language{#1}%
2338   \bbl@foreach\BabelContentsFiles{%
2339     \@writefile{##1}{\babel@toc{#1}{#2}}} % TODO - ok in plain? \string?

```

```

2340 \def\babel@toc#1#2{%
2341   \select@language{#1}{#2}}

```

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

```

2342 \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.

```

2343 \def\bbbl@switch#1{%
2344   \originalTeX
2345   \expandafter\def\expandafter\originalTeX\expandafter{%
2346     \csname noextras#1\endcsname
2347     \let\originalTeX\@empty
2348     \babel@beginsave}%
2349   \bbbl@usehooks{afterreset}}}%
2350   \languageshorthands{none}%
2351   \ifcase\bbbl@select@type
2352     \ifhmode
2353       \hskip\z@skip % trick to ignore spaces
2354       \csname captions#1\endcsname\relax
2355       \csname date#1\endcsname\relax
2356       \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
2357     \else
2358       \csname captions#1\endcsname\relax
2359       \csname date#1\endcsname\relax
2360     \fi
2361   \else\ifbbbl@usedategroup
2362     \bbbl@usedategroupfalse
2363     \ifhmode
2364       \hskip\z@skip % trick to ignore spaces
2365       \csname date#1\endcsname\relax
2366       \loop\ifdim\lastskip>\z@\unskip\repeat\unskip
2367     \else
2368       \csname date#1\endcsname\relax
2369     \fi
2370   \fi\fi
2371   \bbbl@usehooks{beforeextras}}}%
2372   \csname extras#1\endcsname\relax
2373   \bbbl@usehooks{afterextras}}}%
2374   \ifcase\bbbl@opt@hyphenmap\or
2375     \def\BabelLower##1##2{\lccode##1=##2\relax}%
2376     \ifnum\bbbl@hymapsel>4\else
2377       \csname\language @bbbl@hyphenmap\endcsname
2378     \fi

```

```

2379 \chardef\bbl@opt@hyphenmap\z@
2380 \else
2381 \ifnum\bbl@hymapsel>\bbl@opt@hyphenmap\else
2382 \csname\language @bbl@hyphenmap\endcsname
2383 \fi
2384 \fi
2385 \global\let\bbl@hymapsel@cclv
2386 \bbl@patterns{#1}%
2387 \babel@savevariable\lefthyphenmin
2388 \babel@savevariable\righthyphenmin
2389 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2390 \set@hyphenmins\tw@\thr@@\relax
2391 \else
2392 \expandafter\expandafter\expandafter\set@hyphenmins
2393 \csname #1hyphenmins\endcsname\relax
2394 \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.

```

2395 \long\def\otherlanguage#1{%
2396 \ifnum\bbl@hymapsel=\cclv\let\bbl@hymapsel\thr@@\fi
2397 \csname selectlanguage \endcsname{#1}%
2398 \ignorespaces}

```

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

```

2399 \long\def\endotherlanguage{%
2400 \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`.

```

2401 \expandafter\def\csname otherlanguage*\endcsname#1{%
2402 \ifnum\bbl@hymapsel=\cclv\chardef\bbl@hymapsel4\relax\fi
2403 \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”.

```

2404 \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 behaviour 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.

```

2405 \let\bbl@beforeforeign\@empty
2406 \edef\foreignlanguage{%
2407   \noexpand\protect
2408   \expandafter\noexpand\csname foreignlanguage \endcsname}
2409 \expandafter\def\csname foreignlanguage \endcsname{%
2410   \@ifstar\bbl@foreign@s\bbl@foreign@x}
2411 \def\bbl@foreign@x#1#2{%
2412   \begingroup
2413     \let\BabelText\@firstofone
2414     \bbl@beforeforeign
2415     \foreign@language{#1}%
2416     \bbl@usehooks{foreign}{}%
2417     \BabelText{#2}% Now in horizontal mode!
2418   \endgroup}
2419 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \setpar, ?\@par
2420   \begingroup
2421     {\par}%
2422     \let\BabelText\@firstofone
2423     \foreign@language{#1}%
2424     \bbl@usehooks{foreign*}{}%
2425     \bbl@dirparastext
2426     \BabelText{#2}% Still in vertical mode!
2427     {\par}%
2428   \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`.

```

2429 \def\foreign@language#1{%
2430   \edef\language#1}%
2431   \bbl@fixname\language
2432   \bbl@iflanguage\language
2433   \expandafter\ifx\csname date\language\endcsname\relax
2434     \bbl@warning
2435     {Unknown language `#1'. Either you have\\%
2436       misspelled its name, it has not been installed,\\%
2437       or you requested it in a previous run. Fix its name,\\%
2438       install it or just rerun the file, respectively.\\%
2439       I'll proceed, but expect wrong results.\\%
2440       Reported}%
2441   \fi
2442   \let\bbl@select@type\@ne
2443   \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.

```

2444 \let\bbl@hyphlist\@empty
2445 \let\bbl@hyphenation@\relax
2446 \let\bbl@pttnlist\@empty
2447 \let\bbl@patterns@\relax
2448 \let\bbl@hymapset=\ccclv
2449 \def\bbl@patterns#1{%
2450   \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
2451     \csname l@#1\endcsname
2452     \edef\bbl@tempa{#1}%
2453   \else
2454     \csname l@#1:\f@encoding\endcsname
2455     \edef\bbl@tempa{#1:\f@encoding}%
2456   \fi
2457   \@expandtwoargs\bbl@usehooks{patterns}{#1}{\bbl@tempa}%
2458   \@ifundefined{bbl@hyphenation@}{#1}{% Can be \relax!
2459     \begingroup
2460       \bbl@xin@{,\number\language,}{,\bbl@hyphlist}%
2461       \ifin@else
2462         \@expandtwoargs\bbl@usehooks{hyphenation}{#1}{\bbl@tempa}%
2463         \hyphenation{%
2464           \bbl@hyphenation@
2465           \@ifundefined{bbl@hyphenation@#1}%
2466             \@empty
2467             {\space\csname bbl@hyphenation@#1\endcsname}}%
2468         \xdef\bbl@hyphlist{\bbl@hyphlist\number\language,}%
2469       \fi
2470     \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*.

```

2471 \def\hyphenrules#1{%
2472   \edef\bbl@tempf{#1}%
2473   \bbl@fixname\bbl@tempf
2474   \bbl@iflanguage\bbl@tempf{%
2475     \expandafter\bbl@patterns\expandafter{\bbl@tempf}%
2476     \languageshortands{none}%
2477     \bbl@ifunset{\bbl@tempf hyphenmins}%
2478       {\set@hyphenmins\tw@\thr@\relax}%
2479       {\bbl@exp{\set@hyphenmins\@nameuse{\bbl@tempf hyphenmins}}}}
2480 \let\endhyphenrules\@empty

```

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

```

2481 \def\providehyphenmins#1#2{%
2482   \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2483     \@namedef{#1hyphenmins}{#2}%
2484   \fi}

```

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

```
2485 \def\set@hyphenmins#1#2{%
2486   \lefthyphenmin#1\relax
2487   \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.

```
2488 \ifx\ProvidesFile\@undefined
2489   \def\ProvidesLanguage#1[#2 #3 #4]{%
2490     \wlog{Language: #1 #4 #3 <#2>}%
2491   }
2492 \else
2493   \def\ProvidesLanguage#1{%
2494     \begingroup
2495       \catcode`\ 10 %
2496       \@makeother\/%
2497       \@ifnextchar[%]
2498         {@provideslanguage{#1}}{@provideslanguage{#1}[]}
2499   \def{@provideslanguage#1[#2]}{%
2500     \wlog{Language: #1 #2}%
2501     \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
2502     \endgroup}
2503 \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`

```
2504 \def\LdfInit{%
2505   \chardef\atcatcode=\catcode`\@
2506   \catcode`\@=11\relax
2507   \input babel.def\relax
2508   \catcode`\@=\atcatcode \let\atcatcode\relax
2509   \LdfInit}
```

`\originalTeX` The macro `\originalTeX` should be known to T_EX at this moment. As it has to be expandable we `\let` it to `\@empty` instead of `\relax`.

```
2510 \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.

```
2511 \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’:

```
2512 \providecommand\setlocale{%
2513   \bbl@error
2514   {Not yet available}%
2515   {Find an armchair, sit down and wait}}
2516 \let\uselocale\setlocale
```

```

2517 \let\locale\setlocale
2518 \let\selectlocale\setlocale
2519 \let\textlocale\setlocale
2520 \let\textlanguage\setlocale
2521 \let\languagetext\setlocale

```

10.2 Errors

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

`\@noopterr` When the package was loaded without options not everything will work as expected. An error message is issued in that case.
When the format knows about `\PackageError` it must be $\text{\LaTeX 2}_{\epsilon}$, so we can safely use its error handling interface. Otherwise we'll have to 'keep it simple'.

```

2522 \edef\bb1@nulllanguage{\string\language=0}
2523 \ifx\PackageError\@undefined
2524   \def\bb1@error#1#2{%
2525     \begingroup
2526       \newlinechar=`^^J
2527       \def\{^^J(babel) }%
2528       \errhelp{#2}\errmessage{\{#1}%
2529     \endgroup}
2530   \def\bb1@warning#1{%
2531     \begingroup
2532       \newlinechar=`^^J
2533       \def\{^^J(babel) }%
2534       \message{\{#1}%
2535     \endgroup}
2536   \def\bb1@info#1{%
2537     \begingroup
2538       \newlinechar=`^^J
2539       \def\{^^J}%
2540       \wlog{#1}%
2541     \endgroup}
2542 \else
2543   \def\bb1@error#1#2{%
2544     \begingroup
2545       \def\{\MessageBreak}%
2546       \PackageError{babel}{#1}{#2}%
2547     \endgroup}
2548   \def\bb1@warning#1{%
2549     \begingroup
2550       \def\{\MessageBreak}%
2551       \PackageWarning{babel}{#1}%
2552     \endgroup}
2553   \def\bb1@info#1{%
2554     \begingroup
2555       \def\{\MessageBreak}%
2556       \PackageInfo{babel}{#1}%
2557     \endgroup}
2558 \fi
2559 \@ifpackagewith{babel}{silent}
2560   {\let\bb1@info\@gobble
2561    \let\bb1@warning\@gobble}

```

```

2562 {}
2563 \def\bbl@nocaption#1#2{% 1: text to be printed 2: caption macro \langXname
2564 \gdef#2{\textbf{?#1?}}}%
2565 #2%
2566 \bbl@warning{%
2567 \string#2 not set. Please, define\\%
2568 it in the preamble with something like:\\%
2569 \string\renewcommand\string#2{..}\\%
2570 Reported}}
2571 \def\@nolanerr#1{%
2572 \bbl@error
2573 {You haven't defined the language #1\space yet}%
2574 {Your command will be ignored, type <return> to proceed}}
2575 \def\@nopatterns#1{%
2576 \bbl@warning
2577 {No hyphenation patterns were preloaded for\\%
2578 the language `#1' into the format.\\%
2579 Please, configure your TeX system to add them and\\%
2580 rebuild the format. Now I will use the patterns\\%
2581 preloaded for \bbl@nulllanguage\space instead}}
2582 \let\bbl@usehooks\@gobbletwo
2583 \end{kernel}

```

11 Loading hyphenation patterns

The following code is meant to be read by \LaTeX because it should instruct \LaTeX 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. `toks8` stores info to be shown when the program is run.

We want to add a message to the message \LaTeX 2.09 puts in the `\everyjob` register. This could be done by the following code:

```

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

```

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

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

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

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

\@preamblecmds. But we can only do that after it has been defined, so we add this piece of code to \dump.

This new definition starts by adding an instruction to write a message on the terminal and in the transcript file to inform the user of the preloaded hyphenation patterns.

Then everything is restored to the old situation and the format is dumped.

```

2584 <*patterns>
2585 <<Make sure ProvidesFile is defined>>
2586 \ProvidesFile{hyphen.cfg}[<<date>> <<version>> Babel hyphens]
2587 \xdef\bbl@format{\jobname}
2588 \ifx\AtBeginDocument\@undefined
2589   \def\@empty{}
2590   \let\orig@dump\dump
2591   \def\dump{%
2592     \ifx\@ztryfc\@undefined
2593       \else
2594         \toks0=\expandafter{\@preamblecmds}%
2595         \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
2596         \def\@begindocumenthook{}%
2597       \fi
2598       \let\dump\orig@dump\let\orig@dump\@undefined\dump}
2599 \fi
2600 <<Define core switching macros>>
2601 \toks8{Babel <<@version@>> and hyphenation patterns for }%

```

\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.

```

2602 \def\process@line#1#2 #3 #4 {%
2603   \ifx=#1%
2604     \process@synonym{#2}%
2605   \else
2606     \process@language{#1#2}{#3}{#4}%
2607   \fi
2608   \ignorespaces}

```

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

```

2609 \toks@{}
2610 \def\bbl@languages{}

```

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

Otherwise the name will be a synonym for the language loaded last.

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

```

2611 \def\process@synonym#1{%
2612   \ifnum\last@language=\m@ne
2613     \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
2614   \else
2615     \expandafter\chardef\curname l@#1\endcurname\last@language
2616     \wlog{\string\l@#1=\string\language\the\last@language}%
2617     \expandafter\let\curname #1hyphenmins\expandafter\endcurname
2618     \curname\language hyphenmins\endcurname
2619     \let\bbl@elt\relax
2620     \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}}}%

```

`\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 ‘name’ of the language that will be loaded now is added to the token register `\toks8`. and finally 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 behaviour depending on the given encoding (it is set to empty if no encoding is given).

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

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

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

When the hyphenation patterns have been processed we need to see if a file with hyphenation exceptions needs to be read. This is the case when the third argument is not empty and when it does not contain a space token. (Note however there is no need to save hyphenation exceptions into the format.)

`\bbl@languages` saves a snapshot of the loaded languages in the form `\bbl@elt{<language-name>}{<number>}{<patterns-file>}{<exceptions-file>}`. Note the last 2 arguments are empty in ‘dialects’ defined in `language.dat` with =. Note also the language name can have encoding info.

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

```

2622 \def\process@language#1#2#3{%
2623   \expandafter\addlanguage\csname l@#1\endcsname
2624   \expandafter\language\csname l@#1\endcsname
2625   \edef\language{#1}%
2626   \bbl@hook@everylanguage{#1}%
2627   \bbl@get@enc#1::@@@
2628   \begingroup
2629     \lefthyphenmin\m@ne
2630     \bbl@hook@loadpatterns{#2}%
2631     \ifnum\lefthyphenmin=\m@ne
2632     \else
2633       \expandafter\xdef\csname #1hyphenmins\endcsname{%
2634         \the\lefthyphenmin\the\righthyphenmin}%
2635     \fi
2636   \endgroup
2637   \def\bbl@tempa{#3}%
2638   \ifx\bbl@tempa\@empty\else
2639     \bbl@hook@loadexceptions{#3}%
2640   \fi
2641   \let\bbl@elt\relax
2642   \edef\bbl@languages{%
2643     \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
2644   \ifnum\the\language=\z@

```

```

2645 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2646 \set@hyphenmins\tw@thr@@\relax
2647 \else
2648 \expandafter\expandafter\expandafter\set@hyphenmins
2649 \csname #1hyphenmins\endcsname
2650 \fi
2651 \the\toks@
2652 \toks@{}%
2653 \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.

```

2654 \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.

```

2655 \def\bbl@hook@everylanguage#1{}
2656 \def\bbl@hook@loadpatterns#1{\input #1\relax}
2657 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
2658 \let\bbl@hook@loadkernel\bbl@hook@loadpatterns
2659 \begingroup
2660 \def\AddBabelHook#1#2{%
2661 \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
2662 \def\next{\toks1}%
2663 \else
2664 \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname####1}%
2665 \fi
2666 \next}
2667 \ifx\directlua\undefined
2668 \ifx\XeTeXinputencoding\undefined\else
2669 \input xebabel.def
2670 \fi
2671 \else
2672 \input luababel.def
2673 \fi
2674 \openin1 = babel-\bbl@format.cfg
2675 \ifeof1
2676 \else
2677 \input babel-\bbl@format.cfg\relax
2678 \fi
2679 \closein1
2680 \endgroup
2681 \bbl@hook@loadkernel{switch.def}

```

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

```

2682 \openin1 = language.dat

```

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

```

2683 \def\language{english}%
2684 \ifeof1
2685 \message{I couldn't find the file language.dat,\space
2686 I will try the file hyphen.tex}
2687 \input hyphen.tex\relax
2688 \chardef\l@english\z@
2689 \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`.

```
2690 \last@language\m@ne
```

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

```
2691 \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.

```
2692 \endlinechar\m@ne
2693 \read1 to \bbl@line
2694 \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.

```
2695 \if T\ifeof1F\fi T\relax
2696 \ifx\bbl@line\@empty\else
2697 \edef\bbl@line{\bbl@line\space\space\space}%
2698 \expandafter\process@line\bbl@line\relax
2699 \fi
2700 \repeat
```

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

```
2701 \begingroup
2702 \def\bbl@elt#1#2#3#4{%
2703 \global\language=#2\relax
2704 \gdef\languagename{#1}%
2705 \def\bbl@elt##1##2##3##4{}}%
2706 \bbl@languages
2707 \endgroup
2708 \fi
```

and close the configuration file.

```
2709 \closein1
```

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

```
2710 \if/\the\toks@/\else
2711 \errhelp{language.dat loads no language, only synonyms}
2712 \errmessage{Orphan language synonym}
2713 \fi
2714 \advance\last@language\@ne
2715 \edef\bbl@tempa{%
2716 \everyjob{%
2717 \the\everyjob
2718 \ifx\typeout\@undefined
2719 \immediate\write16%
2720 \else
2721 \noexpand\typeout
2722 \fi
2723 {\the\toks8 \the\last@language\space language(s) loaded.}}}
2724 \advance\last@language\m@ne
2725 \bbl@tempa
```

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.

```

2726 \let\bbl@line\@undefined
2727 \let\process@line\@undefined
2728 \let\process@synonym\@undefined
2729 \let\process@language\@undefined
2730 \let\bbl@get@enc\@undefined
2731 \let\bbl@hyph@enc\@undefined
2732 \let\bbl@tempa\@undefined
2733 \let\bbl@hook@loadkernel\@undefined
2734 \let\bbl@hook@everylanguage\@undefined
2735 \let\bbl@hook@loadpatterns\@undefined
2736 \let\bbl@hook@loadexceptions\@undefined
2737 \end{patterns}

```

Here the code for `iniTEX` ends.

12 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].

```

2738 <<*More package options>> ≡
2739 \ifodd\bbl@engine
2740   \DeclareOption{bidi=basic-r}%
2741   {\let\bbl@beforeforeign\leavevmode
2742     \newattribute\bbl@attr@dir
2743     \bbl@exp{\output{\bodydir\pagedir\the\output}}%
2744     \AtEndOfPackage{\EnableBabelHook{babel-bidi}}}
2745 \else
2746   \DeclareOption{bidi=basic-r}%
2747   {\bbl@error
2748     {The bidi method 'basic-r' is available only in\\%
2749     luatex. I'll continue with 'bidi=default', so\\%
2750     expect wrong results}%
2751     {See the manual for further details.}%
2752     \let\bbl@beforeforeign\leavevmode
2753     \AtEndOfPackage{%
2754       \EnableBabelHook{babel-bidi}%
2755       \bbl@xebidipar}}
2756 \fi
2757 \DeclareOption{bidi=default}%
2758 {\let\bbl@beforeforeign\leavevmode
2759   \ifodd\bbl@engine
2760     \newattribute\bbl@attr@dir
2761     \bbl@exp{\output{\bodydir\pagedir\the\output}}%
2762   \fi
2763   \AtEndOfPackage{%
2764     \EnableBabelHook{babel-bidi}%
2765     \ifodd\bbl@engine\else
2766       \bbl@xebidipar
2767     \fi}}
2768 <</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.

```

2769 <<{*Font selection}>> ≡
2770 \@onlypreamble\babelfont
2771 \newcommand\babelfont[2][]{% 1=langs/scripts 2=fam
2772   \edef\bb1@tempa{#1}%
2773   \def\bb1@tempb{#2}%
2774   \ifx\fontspec\undefined
2775     \usepackage{fontspec}%
2776   \fi
2777   \EnableBabelHook{babel-fontspec}%
2778   \bb1@babelfont}
2779 \newcommand\bb1@babelfont[2][]{% 1=features 2=fontname
2780   \bb1@ifunset{\bb1@tempb family}{\bb1@providefam{\bb1@tempb}}{}}%
2781   \bb1@ifunset{\bb1@sys@\language}{\bb1@provide@sys{\language}}{}}%
2782   \expandafter\bb1@ifblank\expandafter{\bb1@tempa}%
2783   {\bb1@csarg\edef{\bb1@tempb dflt@}{<{#1}{#2}}% save \bb1@rmdflt@
2784   \bb1@exp{%
2785     \let\<\bb1@tempb dflt@\language>\<\bb1@tempb dflt@>%
2786     \\\bb1@font@set\<\bb1@tempb dflt@\language>%
2787     \<\bb1@tempb default>\<\bb1@tempb family>}}%
2788   {\bb1@foreach\bb1@tempa{% ie \bb1@rmdflt@lang / *scrt
2789     \bb1@csarg\def{\bb1@tempb dflt@##1}{<{#1}{#2}}}}}%

```

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

```

2790 \def\bb1@providefam#1{%
2791   \bb1@exp{%
2792     \\\newcommand\<#1default>{}% Just define it
2793     \\\bb1@add@list\bb1@font@fams{#1}%
2794     \\\DeclareRobustCommand\<#1family>{%
2795       \\\not@math@alphabet\<#1family>\relax
2796       \\\fontfamily\<#1default>\selectfont}%
2797     \\\DeclareTextFontCommand{\<text#1>}{\<#1family>}}%

```

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

```

2798 \def\bb1@switchfont{%
2799   \bb1@ifunset{\bb1@sys@\language}{\bb1@provide@sys{\language}}{}}%
2800   \bb1@exp{% eg Arabic -> arabic
2801     \lowercase{\edef\bb1@tempa{\bb1@cs{sname@\language}}}}%
2802   \bb1@foreach\bb1@font@fams{%
2803     \bb1@ifunset{\bb1@##1dflt@\language}% (1) language?
2804     {\bb1@ifunset{\bb1@##1dflt@*\bb1@tempa}% (2) from script?
2805       {\bb1@ifunset{\bb1@##1dflt@}% 2=F - (3) from generic?
2806         {}% 123=F - nothing!
2807         {\bb1@exp{% 3=T - from generic
2808           \global\let\<\bb1@##1dflt@\language>%
2809           \<\bb1@##1dflt@>}}}%
2810         {\bb1@exp{% 2=T - from script
2811           \global\let\<\bb1@##1dflt@\language>%
2812           \<\bb1@##1dflt@*\bb1@tempa>}}}%
2813         {}% 1=T - language, already defined
2814   \def\bb1@tempa{%
2815     \bb1@warning{The current font is not a standard family.\\%
2816       Script and Language are not applied. Consider defining\\%
2817       a new family with \string\babelfont,}}%
2818   \bb1@foreach\bb1@font@fams{% don't gather with prev for
2819     \bb1@ifunset{\bb1@##1dflt@\language}%
2820     {\bb1@cs{famrst@##1}%
2821     \global\bb1@csarg\let{famrst@##1}\relax}%
2822     {\bb1@exp{% order is relevant
2823       \\\bb1@add\originalTeX{%

```

```

2824      \\bbl@font@rst{\bbl@cs{##1dflt@\languagename}}%
2825      \<##1default>\<##1family>{##1}}%
2826      \\bbl@font@set\<bbl@##1dflt@\languagename>% the main part!
2827      \<##1default>\<##1family>}}}%
2828      \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.

```

2829 \def\bbl@font@set#1#2#3{%
2830   \bbl@xin@{<>}{#1}%
2831   \ifin@
2832     \bbl@exp{\bbl@fontspec@set\#1\expandafter\@gobbletwo#1}%
2833   \fi
2834   \bbl@exp{%
2835     \def\#2{#1}%          eg, \rmdefault{\bbl@rm1dflt@lang}
2836     \\bbl@ifsamestring{#2}{\f@family}{\#3\let\bbl@tempa\relax}}}%
2837 \def\bbl@fontspec@set#1#2#3{%
2838   \bbl@exp{\<fontspec_set_family:Nnn>\#1%
2839     {\bbl@cs{lsys@\languagename},#2}}{#3}%
2840   \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.

```

2841 \def\bbl@font@rst#1#2#3#4{%
2842   \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.

```

2843 \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 :-).

```

2844 \newcommand\babelFSstore[2][{%
2845   \bbl@ifblank{#1}%
2846   {\bbl@csarg\def{sname@#2}{Latin}}}%
2847   {\bbl@csarg\def{sname@#2}{#1}}}%
2848 \bbl@provide@dirs{#2}%
2849 \bbl@csarg\ifnum{wdir@#2}>\z@
2850   \let\bbl@beforeforeign\leavevmode
2851   \EnableBabelHook{babel-bidi}%
2852 \fi
2853 \bbl@foreach{#2}{%
2854   \bbl@FSstore{##1}{rm}\rmdefault\bbl@save@rmdefault
2855   \bbl@FSstore{##1}{sf}\sfdefault\bbl@save@sfdefault
2856   \bbl@FSstore{##1}{tt}\ttdefault\bbl@save@ttdefault}}
2857 \def\bbl@FSstore#1#2#3#4{%
2858   \bbl@csarg\edef{#2default#1}{#3}%
2859   \expandafter\addto\csname extras#1\endcsname{%
2860     \let#4#3%
2861     \ifx#3\f@family
2862       \edef#3{\csname bbl@#2default#1\endcsname}%
2863       \fontfamily{#3}\selectfont
2864     \else
2865       \edef#3{\csname bbl@#2default#1\endcsname}%
2866       \fi}%
2867   \expandafter\addto\csname noextras#1\endcsname{%

```

```

2868 \ifx#3\f@family
2869 \fontfamily{#4}\selectfont
2870 \fi
2871 \let#3#4}}
2872 \let\bbl@langfeatures\empty
2873 \def\babelFSfeatures{% make sure \fontspec is redefined once
2874 \let\bbl@ori@fontspec\fontspec
2875 \renewcommand\fontspec[1][{}]{%
2876 \bbl@ori@fontspec[\bbl@langfeatures##1]}
2877 \let\babelFSfeatures\bbl@FSfeatures
2878 \babelFSfeatures}
2879 \def\bbl@FSfeatures#1#2{%
2880 \expandafter\addto\csname extras#1\endcsname{%
2881 \babel@save\bbl@langfeatures
2882 \edef\bbl@langfeatures{#2,}}
2883 <</Font selection>>

```

13 Hooks for XeTeX and LuaTeX

13.1 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.

```

2884 <<(*Restore Unicode catcodes before loading patterns)>> ≡
2885 \begingroup
2886 % Reset chars "80-"C0 to category "other", no case mapping:
2887 \catcode\@=11 \count@=128
2888 \loop\ifnum\count@<192
2889 \global\uccode\count@=0 \global\lccode\count@=0
2890 \global\catcode\count@=12 \global\sffcode\count@=1000
2891 \advance\count@ by 1 \repeat
2892 % Other:
2893 \def\O ##1 {%
2894 \global\uccode"##1=0 \global\lccode"##1=0
2895 \global\catcode"##1=12 \global\sffcode"##1=1000 }%
2896 % Letter:
2897 \def\L ##1 ##2 ##3 {\global\catcode"##1=11
2898 \global\uccode"##1="##2
2899 \global\lccode"##1="##3
2900 % Uppercase letters have sffcode=999:
2901 \ifnum"##1="##3 \else \global\sffcode"##1=999 \fi }%
2902 % Letter without case mappings:
2903 \def\l ##1 {\L ##1 ##1 ##1 }%
2904 \l 00AA
2905 \L 00B5 039C 00B5
2906 \l 00BA
2907 \O 00D7
2908 \l 00DF
2909 \O 00F7
2910 \L 00FF 0178 00FF
2911 \endgroup
2912 \input #1\relax
2913 <</Restore Unicode catcodes before loading patterns>>

```

Now, the code.

```
2914 (*xetex)
2915 \def\BabelStringsDefault{unicode}
2916 \let\xebbl@stop\relax
2917 \AddBabelHook{xetex}{encodedcommands}{%
2918   \def\bbl@tempa{#1}%
2919   \ifx\bbl@tempa\@empty
2920     \XeTeXinputencoding"bytes"%
2921   \else
2922     \XeTeXinputencoding"#1"%
2923   \fi
2924   \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
2925 \AddBabelHook{xetex}{stopcommands}{%
2926   \xebbl@stop
2927   \let\xebbl@stop\relax}
2928 \AddBabelHook{xetex}{loadkernel}{%
2929   <<Restore Unicode catcodes before loading patterns>>}
2930 \ifx\DisableBabelHook\@undefined\endinput\fi
2931 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
2932 \DisableBabelHook{babel-fontspec}
2933 <<Font selection>>
2934 \input txtbabel.def
2935 </xetex>
```

13.2 Layout

In progress.

Unfortunately, for proper support for xetex lots of macros and packages must be patched somehow. At least at this stage, babel will not do it and therefore a package similar to bidi will be required. Any help in making babel and bidi collaborate will be welcome. Note as well, elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titleps, and geometry.

\bbl@startskip and \bbl@endskip are available to package authors. Thanks to the 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*.

```
2936 (*texxet)
2937 \ifx\bbl@opt@layout\@nnil\endinput\fi % No layout
2938 \def\bbl@startskip{\ifcase\bbl@thepardir\leftskip\else\rightskip\fi}
2939 \def\bbl@endskip{\ifcase\bbl@thepardir\rightskip\else\leftskip\fi}
2940 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
2941   \def\@hangfrom#1{%
2942     \setbox\@tempboxa\hbox{#1}%
2943     \hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
2944     \noindent\box\@tempboxa}
2945   \def\raggedright{%
2946     \let\@centercr
2947     \bbl@startskip\z@skip
2948     \@rightskip\@flushglue
2949     \bbl@endskip\@rightskip
2950     \parindent\z@
2951     \parfillskip\bbl@startskip}
2952   \def\raggedleft{%
2953     \let\@centercr
2954     \bbl@startskip\@flushglue
2955     \bbl@endskip\z@skip
2956     \parindent\z@
```

```

2957 \parfillskip\bbl@endskip}
2958 \fi
2959 \IfBabelLayout{lists}
2960 {\def\list#1#2{%
2961 \ifnum \@listdepth >5\relax
2962 \@toodeep
2963 \else
2964 \global\advance\@listdepth\@ne
2965 \fi
2966 \rightmargin\z@
2967 \listparindent\z@
2968 \itemindent\z@
2969 \csname @list\romannumeral\the\@listdepth\endcsname
2970 \def\@itemlabel{#1}%
2971 \let\makelabel\@mklab
2972 \@nmbrlistfalse
2973 #2\relax
2974 \@trivlist
2975 \parskip\parsep
2976 \parindent\listparindent
2977 \advance\linewidth-\rightmargin
2978 \advance\linewidth-\leftmargin
2979 \advance\@totalleftmargin
2980 \ifcase\bbl@thepardir\leftmargin\else\rightmargin\fi
2981 \parshape\@ne\@totalleftmargin\linewidth
2982 \ignorespaces}%
2983 \def\labelenumii{}\theenumii{%
2984 \def\p@enumiii{\p@enumii}\theenumii{}}
2985 {}
2986 \IfBabelLayout{contents}
2987 {\def\@dottedtocline#1#2#3#4#5{%
2988 \ifnum#1>\c@tocdepth\else
2989 \vskip \z@ \@plus.2\p@
2990 {\bbl@startskip#2\relax
2991 \bbl@endskip\@tocmarg
2992 \parfillskip-\bbl@endskip
2993 \parindent#2\relax
2994 \@afterindenttrue
2995 \interlinepenalty\@M
2996 \leavevmode
2997 \@tempdima#3\relax
2998 \advance\bbl@startskip\@tempdima
2999 \null\nobreak\hskip-\bbl@startskip
3000 {#4}\nobreak
3001 \leaders\hbox{%
3002 $m@th\mkern\@dotsep mu\hbox{.}\mkern\@dotsep mu$}%
3003 \hfill\nobreak
3004 \hb@xt@\@pnumwidth{\hfil\normalfont\normalcolor#5}%
3005 \par}%
3006 \fi}%
3007 \def\babel@toc#1{%
3008 \select@language{#1}%
3009 \bbl@ifunset{bbl@wdir@\bbl@main@language}%
3010 {\bbl@provide@dirs{\bbl@main@language}}%
3011 {}%
3012 \bbl@exp{%
3013 \\bbl@pardir\bbl@cs{wdir@\bbl@main@language}%
3014 \\bbl@textdir\bbl@cs{wdir@\bbl@main@language}}}}
3015 {}

```

```

3016 \IfBabelLayout{columns}
3017   {\def\@outputdblcol{%
3018     \if@firstcolumn
3019       \global\@firstcolumnfalse
3020       \global\setbox\@leftcolumn\copy\@outputbox
3021       \splitmaxdepth\maxdimen
3022       \vbadness\maxdimen
3023       \setbox\@outputbox\vbox{\unvbox\@outputbox\unskip}%
3024       \setbox\@outputbox\vsplit\@outputbox to\maxdimen
3025       \toks@ \expandafter{\topmark}%
3026       \xdef\@firstcoltopmark{\the\toks@}%
3027       \toks@ \expandafter{\splitfirstmark}%
3028       \xdef\@firstcolfirstmark{\the\toks@}%
3029       \ifx\@firstcolfirstmark\@empty
3030         \global\let\@setmarks\relax
3031       \else
3032         \gdef\@setmarks{%
3033           \let\firstmark\@firstcolfirstmark
3034           \let\topmark\@firstcoltopmark}%
3035       \fi
3036     \else
3037       \global\@firstcolumntrue
3038       \setbox\@outputbox\vbox{%
3039         \hb@xt@\textwidth{%
3040           \hskip\columnwidth
3041           \hfil
3042           {\normalcolor\vrule \@width\columnseprule}%
3043           \hfil
3044           \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
3045           \hskip-\textwidth
3046           \hb@xt@\columnwidth{\box\@outputbox \hss}%
3047           \hskip\columnsep
3048           \hskip\columnwidth}}%
3049       \@combinedblfloats
3050       \@setmarks
3051       \@outputpage
3052       \begingroup
3053         \@dblfloatplacement
3054         \@startdblcolumn
3055         \@whiles\if@fcolmade \fi{\@outputpage
3056           \@startdblcolumn}%
3057       \endgroup
3058     \fi}}%
3059   {}

```

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

```

3060 \IfBabelLayout{counters}% Global or language dependent? At begin doc?
3061   {\let\bbl@latinarabic=\@arabic
3062     \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
3063     \let\bbl@asciroman=\@roman
3064     \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
3065     \let\bbl@asciiRoman=\@Roman
3066     \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}}%
3067 \</texet>

```


13.3 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, they are added, but note if the language patterns have not been preloaded they won't at run time.

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

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

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

We need catcode tables, but no format (targeted by `babel`) provide a command to allocate them (although there are packages like `ctablestack`). For the moment, a dangerous approach is used – just allocate a high random number and cross the fingers. To complicate things, `etex.sty` changes the way languages are allocated.

```
3068 (*luatex)
3069 \ifx\AddBabelHook\undefined
3070 \begingroup
3071   \toks@{}
3072   \count@ \z@ % 0=start, 1=0th, 2=normal
3073   \def\bbl@process@line#1#2 #3 #4 {%
3074     \ifx=#1%
3075       \bbl@process@synonym{#2}%
3076     \else
3077       \bbl@process@language{#1#2}{#3}{#4}%
3078     \fi
3079     \ignorespaces}
3080   \def\bbl@manylang{%
3081     \ifnum\bbl@last>\@ne
3082       \bbl@info{Non-standard hyphenation setup}%
3083     \fi
3084     \let\bbl@manylang\relax}
3085   \def\bbl@process@language#1#2#3{%
3086     \ifcase\count@
3087       \ifundefined{zth#1}{\count@ \tw@}{\count@ \@ne}%
3088     \or
3089       \count@ \tw@
3090     \fi
3091     \ifnum\count@=\tw@
```

```

3092 \expandafter\addlanguage\csname l@#1\endcsname
3093 \language\allocationnumber
3094 \chardef\bbl@last\allocationnumber
3095 \bbl@manylang
3096 \let\bbl@elt\relax
3097 \xdef\bbl@languages{%
3098   \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
3099 \fi
3100 \the\toks@
3101 \toks@{}}
3102 \def\bbl@process@synonym@aux#1#2{%
3103   \global\expandafter\chardef\csname l@#1\endcsname#2\relax
3104   \let\bbl@elt\relax
3105   \xdef\bbl@languages{%
3106     \bbl@languages\bbl@elt{#1}{#2}{}}}%
3107 \def\bbl@process@synonym#1{%
3108   \ifcase\count@
3109     \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
3110   \or
3111     \ifundefined{zth@#1}{\bbl@process@synonym@aux{#1}{0}}}%
3112   \else
3113     \bbl@process@synonym@aux{#1}{\the\bbl@last}%
3114   \fi}
3115 \ifx\bbl@languages\@undefined % Just a (sensible?) guess
3116   \chardef\l@english\z@
3117   \chardef\l@USenglish\z@
3118   \chardef\bbl@last\z@
3119   \global\@namedef{bbl@hyphendata@0}{\hyphen.tex}{}
3120   \gdef\bbl@languages{%
3121     \bbl@elt{english}{0}{\hyphen.tex}{}%
3122     \bbl@elt{USenglish}{0}{}}
3123 \else
3124   \global\let\bbl@languages@format\bbl@languages
3125   \def\bbl@elt#1#2#3#4{% Remove all except language 0
3126     \ifnum#2>\z@\else
3127       \noexpand\bbl@elt{#1}{#2}{#3}{#4}%
3128     \fi}%
3129   \xdef\bbl@languages{\bbl@languages}%
3130 \fi
3131 \def\bbl@elt#1#2#3#4{\@namedef{zth@#1}{}} % Define flags
3132 \bbl@languages
3133 \openin1=language.dat
3134 \ifeof1
3135   \bbl@warning{I couldn't find language.dat. No additional\%
3136     patterns loaded. Reported}%
3137 \else
3138   \loop
3139     \endlinechar\m@ne
3140     \read1 to \bbl@line
3141     \endlinechar\^^M
3142     \if T\ifeof1F\fi T\relax
3143     \ifx\bbl@line\@empty\else
3144       \edef\bbl@line{\bbl@line\space\space\space}%
3145       \expandafter\bbl@process@line\bbl@line\relax
3146     \fi
3147   \repeat
3148 \fi
3149 \endgroup
3150 \def\bbl@get@enc#1:#2:#3\@@@{\def\bbl@hyph@enc{#2}}

```

```

3151 \ifx\babelcatcodetablenum\@undefined
3152   \def\babelcatcodetablenum{5211}
3153 \fi
3154 \def\bbl@luapatterns#1#2{%
3155   \bbl@get@enc#1::\@@@
3156   \setbox\z@\hbox\bgroup
3157     \begingroup
3158       \ifx\catcodetable\@undefined
3159         \let\savecatcodetable\luatexsavecatcodetable
3160         \let\initcatcodetable\luatexinitcatcodetable
3161         \let\catcodetable\luatexcatcodetable
3162       \fi
3163       \savecatcodetable\babelcatcodetablenum\relax
3164       \initcatcodetable\numexpr\babelcatcodetablenum+1\relax
3165       \catcodetable\numexpr\babelcatcodetablenum+1\relax
3166       \catcode`\#=6 \catcode`\$=3 \catcode`\&=4 \catcode`\^=7
3167       \catcode`\_ =8 \catcode`\{=1 \catcode`\}=2 \catcode`\~ =13
3168       \catcode`\@=11 \catcode`\^^I=10 \catcode`\^^J=12
3169       \catcode`\<=12 \catcode`\>=12 \catcode`\*=12 \catcode`\.=12
3170       \catcode`\-=12 \catcode`\/=12 \catcode`\[=12 \catcode`\]=12
3171       \catcode`\`=12 \catcode`\'=12 \catcode`\`=12
3172       \input #1\relax
3173       \catcodetable\babelcatcodetablenum\relax
3174     \endgroup
3175   \def\bbl@tempa{#2}%
3176   \ifx\bbl@tempa\@empty\else
3177     \input #2\relax
3178   \fi
3179 \egroup}%
3180 \def\bbl@patterns@lua#1{%
3181   \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
3182     \csname l@#1\endcsname
3183     \edef\bbl@tempa{#1}%
3184   \else
3185     \csname l@#1:\f@encoding\endcsname
3186     \edef\bbl@tempa{#1:\f@encoding}%
3187   \fi\relax
3188   \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
3189   \@ifundefined{bbl@hyphendata@the\language}%
3190     {\def\bbl@elt##1##2##3##4{%
3191       \ifnum##2=\csname l@bbl@tempa\endcsname % #2=spanish, dutch:OT1...
3192       \def\bbl@tempb{##3}%
3193       \ifx\bbl@tempb\@empty\else % if not a synonymous
3194         \def\bbl@tempc{##3}{##4}%
3195       \fi
3196       \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
3197     \fi}%
3198   \bbl@languages
3199   \@ifundefined{bbl@hyphendata@the\language}%
3200     {\bbl@info{No hyphenation patterns were set for\%
3201       language '\bbl@tempa'. Reported}}%
3202     {\expandafter\expandafter\expandafter\bbl@luapatterns
3203       \csname bbl@hyphendata@the\language\endcsname}}}%
3204 \endinput\fi
3205 \begingroup
3206 \catcode`\%=12
3207 \catcode`\'=12
3208 \catcode`\`=12
3209 \catcode`\:=12

```

```

3210 \directlua{
3211   Babel = Babel or {}
3212   function Babel.bytes(line)
3213     return line:gsub("(.)",
3214       function (chr) return unicode.utf8.char(string.byte(chr)) end)
3215   end
3216   function Babel.begin_process_input()
3217     if luatexbase and luatexbase.add_to_callback then
3218       luatexbase.add_to_callback('process_input_buffer',
3219         Babel.bytes, 'Babel.bytes')
3220     else
3221       Babel.callback = callback.find('process_input_buffer')
3222       callback.register('process_input_buffer', Babel.bytes)
3223     end
3224   end
3225   function Babel.end_process_input ()
3226     if luatexbase and luatexbase.remove_from_callback then
3227       luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
3228     else
3229       callback.register('process_input_buffer', Babel.callback)
3230     end
3231   end
3232   function Babel.addpatterns(pp, lg)
3233     local lg = lang.new(lg)
3234     local pats = lang.patterns(lg) or ''
3235     lang.clear_patterns(lg)
3236     for p in pp:gmatch('[^%s]+') do
3237       ss = ''
3238       for i in string.utfcharacters(p:gsub('%d', '')) do
3239         ss = ss .. '%d?' .. i
3240       end
3241       ss = ss:gsub('^%%d%?%', '%%.') .. '%d?'
3242       ss = ss:gsub('%.%%d%?$%', '%%.')
3243       pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
3244       if n == 0 then
3245         tex.sprint(
3246           [[\string\csname\space bbl@info\endcsname{New pattern: }]]
3247           .. p .. [[{}]])
3248         pats = pats .. ' ' .. p
3249       else
3250         tex.sprint(
3251           [[\string\csname\space bbl@info\endcsname{Renew pattern: }]]
3252           .. p .. [[{}]])
3253       end
3254     end
3255     lang.patterns(lg, pats)
3256   end
3257 }
3258 \endgroup
3259 \def\BabelStringsDefault{unicode}
3260 \let\luabbl@stop\relax
3261 \AddBabelHook{luatex}{encodedcommands}{%
3262   \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
3263   \ifx\bbl@tempa\bbl@tempb\else
3264     \directlua{Babel.begin_process_input()}%
3265     \def\luabbl@stop{%
3266       \directlua{Babel.end_process_input()}}%
3267   \fi}%
3268 \AddBabelHook{luatex}{stopcommands}{%

```

```

3269 \luabbl@stop
3270 \let\luabbl@stop\relax}
3271 \AddBabelHook{luatex}{patterns}{%
3272   \@ifundefined{bbl@hyphendata@the\language}%
3273   {\def\bbl@elt##1##2##3##4{%
3274     \ifnum##2=\csname l@#2\endcsname % #2=spanish, dutch:OT1...
3275     \def\bbl@tempb{##3}%
3276     \ifx\bbl@tempb\@empty\else % if not a synonymous
3277       \def\bbl@tempc{##3}{##4}}%
3278     \fi
3279     \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
3280     \fi}%
3281   \bbl@languages
3282   \@ifundefined{bbl@hyphendata@the\language}%
3283   {\bbl@info{No hyphenation patterns were set for\%
3284     language '#2'. Reported}}%
3285   {\expandafter\expandafter\expandafter\bbl@luapatterns
3286     \csname bbl@hyphendata@the\language\endcsname}}}%
3287 \@ifundefined{bbl@patterns@}{%
3288   \begingroup
3289   \bbl@xin@{,\number\language,}{,\bbl@pttnlist}%
3290   \ifin@else
3291     \ifx\bbl@patterns@\@empty\else
3292       \directlua{ Babel.addpatterns(
3293         [[\bbl@patterns@]], \number\language) }%
3294       \fi
3295       \@ifundefined{bbl@patterns@#1}%
3296       {\@empty
3297         {\directlua{ Babel.addpatterns(
3298           [[\space\csname bbl@patterns@#1\endcsname]],
3299           \number\language) }}%
3300         \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
3301         \fi
3302       \endgroup}}
3303 \AddBabelHook{luatex}{everylanguage}{%
3304   \def\process@language##1##2##3{%
3305     \def\process@line####1####2 ####3 ####4 {}}}
3306 \AddBabelHook{luatex}{loadpatterns}{%
3307   \input #1\relax
3308   \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
3309     {##1}{}}
3310 \AddBabelHook{luatex}{loadexceptions}{%
3311   \input #1\relax
3312   \def\bbl@tempb##1##2{##1}{##1}%
3313   \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
3314     {\expandafter\expandafter\expandafter\bbl@tempb
3315       \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.

```

3316 \@onlypreamble\babelpatterns
3317 \AtEndOfPackage{%
3318   \newcommand\babelpatterns[2][\@empty]{%
3319     \ifx\bbl@patterns@\relax
3320       \let\bbl@patterns@\@empty
3321     \fi
3322     \ifx\bbl@pttnlist\@empty\else

```

```

3323 \bbl@warning{%
3324   You must not intermingle \string\selectlanguage\space and\\%
3325   \string\babelpatterns\space or some patterns will not\\%
3326   be taken into account. Reported}%
3327 \fi
3328 \ifx\@empty#1%
3329   \protected@edef\bbl@patterns@\bbl@patterns@\space#2}%
3330 \else
3331   \edef\bbl@tempb{\zap@space#1 \@empty}%
3332   \bbl@for\bbl@tempa\bbl@tempb{%
3333     \bbl@fixname\bbl@tempa
3334     \bbl@iflanguage\bbl@tempa{%
3335       \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
3336         \@ifundefined{bbl@patterns@\bbl@tempa}%
3337         \@empty
3338         {\csname bbl@patterns@\bbl@tempa\endcsname\space}%
3339         #2}}}%
3340 \fi}}

```

Common stuff.

```

3341 \AddBabelHook{luatex}{loadkernel}{%
3342 <<Restore Unicode catcodes before loading patterns>>}
3343 \ifx\DisableBabelHook\undefined\endinput\fi
3344 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
3345 \DisableBabelHook{babel-fontspec}
3346 <<Font selection>>

```

13.4 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.

```

3347 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
3348 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
3349   \def\@hangfrom#1{%
3350     \setbox\@tempboxa\hbox{#1}%
3351     \hangindent\wd\@tempboxa
3352     \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
3353       \shapemode\@ne
3354     \fi
3355     \noindent\box\@tempboxa}
3356 \fi
3357 \IfBabelLayout{lists}
3358   {\def\list#1#2{%
3359     \ifnum \@listdepth >5\relax
3360       \@toodeep
3361     \else
3362       \global\advance\@listdepth\@ne
3363     \fi
3364     \rightmargin\z@
3365     \listparindent\z@

```

```

3366 \itemindent\z@
3367 \csname @list\romannumeral\the\@listdepth\endcsname
3368 \def\@itemlabel{#1}%
3369 \let\makelabel\@mklab
3370 \@nmbrlistfalse
3371 #2\relax
3372 \@trivlist
3373 \parskip\parsep
3374 \parindent\listparindent
3375 \advance\linewidth -\rightmargin
3376 \advance\linewidth -\leftmargin
3377 \advance\@totalleftmargin \leftmargin
3378 \parshape \@ne
3379 \@totalleftmargin \linewidth
3380 \ifnum\bbbl@getluadir{page}=\bbbl@getluadir{par}\else
3381 \shapemode\tw@
3382 \fi
3383 \ignorespaces}}
3384 {}
3385 \IfBabelLayout{contents}
3386 {\def\babel@toc#1{%
3387 \select@language{#1}%
3388 \bbbl@ifunset{bbbl@wdir@\bbbl@main@language}%
3389 {\bbbl@provide@dirs{\bbbl@main@language}}%
3390 }%
3391 \bbbl@exp{%
3392 \\bbbl@pdir\bbbl@cs{wdir@\bbbl@main@language}%
3393 \\bbbl@texdir\bbbl@cs{wdir@\bbbl@main@language}}}}
3394 {}

```

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.

```

3395 \IfBabelLayout{counters}% Global or language dependent? At begin doc?
3396 {\let\bbbl@latinarabic=\@arabic
3397 \def\@arabic#1{\babelsublr{\bbbl@latinarabic#1}}%
3398 \@ifpackagewith{babel}{bidi=default}
3399 {\let\bbbl@asciroman=\@roman
3400 \def\@roman#1{\babelsublr{\ensureascii{\bbbl@asciroman#1}}}%
3401 \let\bbbl@asciiRoman=\@Roman
3402 \def\@Roman#1{\babelsublr{\ensureascii{\bbbl@asciiRoman#1}}}}}%
3403 % only if bidi=default %%%%%%%%%%% :
3404 % \def\labelenumii{}\theenumii{} % luas, pdf, no xe, luar
3405 % \def\p@enumiii{\p@enumii}\theenumii{} % luas, pdf, no xe, luar
3406 \end{lua}

```

13.5 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):

Arrrrh!! 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.

```
3407 (*basic-r)
3408 Babel = Babel or {}
3409
3410 require('babel-bidi.lua')
3411
3412 local characters = Babel.characters
3413 local ranges = Babel.ranges
3414
3415 local DIR = node.id("dir")
3416
3417 local function dir_mark(head, from, to, outer)
3418   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
3419   local d = node.new(DIR)
3420   d.dir = '+' .. dir
3421   node.insert_before(head, from, d)
3422   d = node.new(DIR)
3423   d.dir = '-' .. dir
3424   node.insert_after(head, to, d)
3425 end
3426
3427 function Babel.pre_otfload(head)
3428   local first_n, last_n      -- first and last char with nums
3429   local last_es              -- an auxiliary 'last' used with nums
3430   local first_d, last_d      -- first and last char in L/R block
3431   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):

```
3432   local strong = ('TRT' == tex.pardir) and 'r' or 'l'
3433   local strong_lr = (strong == 'l') and 'l' or 'r'
3434   local outer = strong
3435
3436   local new_dir = false
3437   local first_dir = false
3438
3439   local last_lr
3440
```



```

3441 local type_n = ''
3442
3443 for item in node.traverse(head) do
3444
3445     -- three cases: glyph, dir, otherwise
3446     if item.id == node.id'glyph' then
3447
3448         local chardata = characters[item.char]
3449         dir = chardata and chardata.d or nil
3450         if not dir then
3451             for nn, et in ipairs(ranges) do
3452                 if item.char < et[1] then
3453                     break
3454                 elseif item.char <= et[2] then
3455                     dir = et[3]
3456                     break
3457                 end
3458             end
3459         end
3460         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.

```

3461     if new_dir then
3462         attr_dir = 0
3463         for at in node.traverse(item.attr) do
3464             if at.number == luatexbase.registernumber'bbl@attr@dir' then
3465                 attr_dir = at.value
3466             end
3467         end
3468         if attr_dir == 1 then
3469             strong = 'r'
3470         elseif attr_dir == 2 then
3471             strong = 'al'
3472         else
3473             strong = 'l'
3474         end
3475         strong_lr = (strong == 'l') and 'l' or 'r'
3476         outer = strong_lr
3477         new_dir = false
3478     end
3479     if dir == 'nsm' then dir = strong end -- W1

```

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

```

3480     dir_real = dir -- We need dir_real to set strong below
3481     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:

```

3482     if strong == 'al' then
3483         if dir == 'en' then dir = 'an' end -- W2
3484         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
3485         strong_lr = 'r' -- W3
3486     end

```

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

```

3487 elseif item.id == node.id'dir' then
3488     new_dir = true
3489     dir = nil
3490 else
3491     dir = nil          -- Not a char
3492 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 behaviour could be changed with a switch in the future. Note in the first branch only <an> is relevant if <al>.

```

3493 if dir == 'en' or dir == 'an' or dir == 'et' then
3494     if dir ~= 'et' then
3495         type_n = dir
3496     end
3497     first_n = first_n or item
3498     last_n = last_es or item
3499     last_es = nil
3500 elseif dir == 'es' and last_n then -- W3+W6
3501     last_es = item
3502 elseif dir == 'cs' then          -- it's right - do nothing
3503 elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
3504     if strong_lr == 'r' and type_n ~= '' then
3505         dir_mark(head, first_n, last_n, 'r')
3506     elseif strong_lr == 'l' and first_d and type_n == 'an' then
3507         dir_mark(head, first_n, last_n, 'r')
3508         dir_mark(head, first_d, last_d, outer)
3509         first_d, last_d = nil, nil
3510     elseif strong_lr == 'l' and type_n ~= '' then
3511         last_d = last_n
3512     end
3513     type_n = ''
3514     first_n, last_n = nil, nil
3515 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:

```

3516 if dir == 'l' or dir == 'r' then
3517     if dir ~= outer then
3518         first_d = first_d or item
3519         last_d = item
3520     elseif first_d and dir ~= strong_lr then
3521         dir_mark(head, first_d, last_d, outer)
3522         first_d, last_d = nil, nil
3523     end
3524 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>, resptly, 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.

```

3525 if dir and not last_lr and dir ~= 'l' and outer == 'r' then
3526     item.char = characters[item.char] and

```

```

3527         characters[item.char].m or item.char
3528     elseif (dir or new_dir) and last_lr ~= item then
3529         local mir = outer .. strong_lr .. (dir or outer)
3530         if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
3531             for ch in node.traverse(node.next(last_lr)) do
3532                 if ch == item then break end
3533                 if ch.id == node.id'glyph' then
3534                     ch.char = characters[ch.char].m or ch.char
3535                 end
3536             end
3537         end
3538     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).

```

3539     if dir == 'l' or dir == 'r' then
3540         last_lr = item
3541         strong = dir_real          -- Don't search back - best save now
3542         strong_lr = (strong == 'l') and 'l' or 'r'
3543     elseif new_dir then
3544         last_lr = nil
3545     end
3546 end

```

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

```

3547 if last_lr and outer == 'r' then
3548     for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
3549         ch.char = characters[ch.char].m or ch.char
3550     end
3551 end
3552 if first_n then
3553     dir_mark(head, first_n, last_n, outer)
3554 end
3555 if first_d then
3556     dir_mark(head, first_d, last_d, outer)
3557 end

```

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

```

3558 return node.prev(head) or head
3559 end
3560 </basic-r>

```

14 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.

```

3561 <nil>
3562 \ProvidesLanguage{nil}[<<date>> <<version>> Nil language]
3563 \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.

```

3564 \ifx\l@nohyphenation\@undefined
3565     \@nopatterns{nil}

```

```

3566 \adddialect\l@nil0
3567 \else
3568 \let\l@nil\l@nohyphenation
3569 \fi

```

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

```

3570 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}

```

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

```

\captionnil
\datenil 3571 \let\captionnil\@empty
3572 \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.

```

3573 \ldf@finish{nil}
3574 \</nil>

```

15 Support for Plain \TeX (plain.def)

15.1 Not renaming hyphen.tex

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

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

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

The files `bplain.tex` and `blplain.tex` can be used as replacement wrappers around `plain.tex` and `lplain.tex` to achieve the desired effect, based on the `babel` package. If you load each of them with \TeX , you will get a file called either `bplain.fmt` or `blplain.fmt`, which you can use as replacements for `plain.fmt` and `lplain.fmt`. As these files are going to be read as the first thing \TeX sees, we need to set some category codes just to be able to change the definition of `\input`

```

3575 \< *bplain | blplain>
3576 \catcode`\{=1 % left brace is begin-group character
3577 \catcode`\}=2 % right brace is end-group character
3578 \catcode`\#=6 % hash mark is macro parameter character

```

Now let’s see if a file called `hyphen.cfg` can be found somewhere on \TeX ’s input path by trying to open it for reading...

```

3579 \openin 0 hyphen.cfg

```

If the file wasn’t found the following test turns out true.

```

3580 \ifeof0
3581 \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).

```
3582 \let\input
```

Then `\input` is defined to forget about its argument and load `hyphen.cfg` instead.

```
3583 \def\input #1 {%
3584   \let\input\input
3585   \input{hyphen.cfg}
```

Once that's done the original meaning of `\input` can be restored and the definition of `\input` can be forgotten.

```
3586 \let\input\undefined
3587 }
3588 \fi
3589 </bplain | bplain>
```

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

```
3590 <bplain>\input{plain.tex}
3591 <bplain>\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.

```
3592 <bplain>\def\fmtname{babel-plain}
3593 <bplain>\def\fmtname{babel-lplain}
```

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

15.2 Emulating some \LaTeX features

The following code duplicates or emulates parts of $\text{\LaTeX} 2_{\epsilon}$ that are needed for `babel`.

```
3594 <*plain>
3595 \def\@empty{}
3596 \def\loadlocalcfg#1{%
3597   \openin0#1.cfg
3598   \ifeof0
3599     \closein0
3600   \else
3601     \closein0
3602     {\immediate\write16{*****}%
3603      \immediate\write16{* Local config file #1.cfg used}%
3604      \immediate\write16{*}%
3605     }
3606     \input #1.cfg\relax
3607   \fi
3608   \@endofldf}
```

15.3 General tools

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

```
3609 \long\def\@firstofone#1{#1}
3610 \long\def\@firstoftwo#1#2{#1}
3611 \long\def\@secondoftwo#1#2{#2}
```

```

3612 \def\@nnil{\@nil}
3613 \def\@gobbletwo#1#2{}
3614 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
3615 \def\@star@or@long#1{%
3616   \@ifstar
3617   {\let\l@ngrel@x\relax#1}%
3618   {\let\l@ngrel@x\long#1}}
3619 \let\l@ngrel@x\relax
3620 \def\@car#1#2\@nil{#1}
3621 \def\@cdr#1#2\@nil{#2}
3622 \let\@typeset@protect\relax
3623 \let\protected@edef\edef
3624 \long\def\@gobble#1{}
3625 \edef\@backslashchar{\expandafter\@gobble\string\}
3626 \def\strip@prefix#1>{}
3627 \def\g@addto@macro#1#2{%
3628   \toks@{\expandafter{#1#2}}%
3629   \xdef#1{\the\toks@}}
3630 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
3631 \def\@nameuse#1{\csname #1\endcsname}
3632 \def\@ifundefined#1{%
3633   \expandafter\ifx\csname#1\endcsname\relax
3634     \expandafter\@firstoftwo
3635   \else
3636     \expandafter\@secondoftwo
3637   \fi}
3638 \def\@expandtwoargs#1#2#3{%
3639   \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
3640 \def\zap@space#1 #2{%
3641   #1%
3642   \ifx#2\@empty\else\expandafter\zap@space\fi
3643   #2}

```

$\LaTeX_2\epsilon$ has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

3644 \ifx\@preamblecmds\@undefined
3645   \def\@preamblecmds{}
3646 \fi
3647 \def\@onlypreamble#1{%
3648   \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
3649     \@preamblecmds\do#1}}
3650 \@onlypreamble\@onlypreamble

```

Mimick \LaTeX 's `\AtBeginDocument`; for this to work the user needs to add `\begindocument` to his file.

```

3651 \def\begindocument{%
3652   \@begindocumenthook
3653   \global\let\@begindocumenthook\@undefined
3654   \def\do##1{\global\let##1\@undefined}%
3655   \@preamblecmds
3656   \global\let\do\noexpand}
3657 \ifx\@begindocumenthook\@undefined
3658   \def\@begindocumenthook{}
3659 \fi
3660 \@onlypreamble\@begindocumenthook
3661 \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`.

```

3662 \def\AtEndOfPackage#1{\g@addto@macro\@endofldf{#1}}
3663 \@onlypreamble\AtEndOfPackage
3664 \def\@endofldf{}
3665 \@onlypreamble\@endofldf
3666 \let\bbl@afterlang\@empty
3667 \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.

```

3668 \ifx@if@files\@undefined
3669 \expandafter\let\csname if@files\expandafter\endcsname
3670 \csname iffalse\endcsname
3671 \fi

```

Mimick \LaTeX 's commands to define control sequences.

```

3672 \def\newcommand{\@star@or@long\new@command}
3673 \def\new@command#1{%
3674 \testopt{\@newcommand#1}0}
3675 \def\@newcommand#1[#2]{%
3676 \@ifnextchar [{\@xargdef#1[#2]}%
3677 {\@argdef#1[#2]}}
3678 \long\def\@argdef#1[#2]#3{%
3679 \@yargdef#1\@ne{#2}{#3}}
3680 \long\def\@xargdef#1[#2][#3]#4{%
3681 \expandafter\def\expandafter#1\expandafter{%
3682 \expandafter\@protected@testopt\expandafter #1%
3683 \csname\string#1\expandafter\endcsname{#3}}%
3684 \expandafter\@yargdef \csname\string#1\endcsname
3685 \tw@{#2}{#4}}
3686 \long\def\@yargdef#1#2#3{%
3687 \@tempcnta#3\relax
3688 \advance \@tempcnta \@ne
3689 \let\@hash@\relax
3690 \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
3691 \@tempcntb #2%
3692 \@whilenum \@tempcntb < \@tempcnta
3693 \do{%
3694 \edef\reserved@a{\reserved@a\@hash@the\@tempcntb}%
3695 \advance \@tempcntb \@ne}%
3696 \let\@hash@##%
3697 \@ngrelex\expandafter\def\expandafter#1\reserved@a}
3698 \def\providecommand{\@star@or@long\provide@command}
3699 \def\provide@command#1{%
3700 \begingroup
3701 \escapechar\m@ne\xdef\@gtempa{\string#1}%
3702 \endgroup
3703 \expandafter\@ifundefined\@gtempa
3704 {\def\reserved@a{\new@command#1}}%
3705 {\let\reserved@a\relax
3706 \def\reserved@a{\new@command\reserved@a}}%
3707 \reserved@a}%
3708 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
3709 \def\declare@robustcommand#1{%
3710 \edef\reserved@a{\string#1}%
3711 \def\reserved@b{#1}%
3712 \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
3713 \edef#1{%
3714 \ifx\reserved@a\reserved@b

```

```

3715         \noexpand\x@protect
3716         \noexpand#1%
3717     \fi
3718     \noexpand\protect
3719     \expandafter\noexpand\csname\bbl@stripslash#1 \endcsname
3720 }%
3721 \expandafter\newcommand\csname\bbl@stripslash#1 \endcsname
3722 }
3723 \def\x@protect#1{%
3724     \ifx\protect\@typeset@protect\else
3725         \@x@protect#1%
3726     \fi
3727 }
3728 \def\@x@protect#1\fi#2#3{%
3729     \fi\protect#1%
3730 }

```

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`.

```

3731 \def\bbl@tempa{\csname newif\endcsname\ifin@}
3732 \ifx\in@\@undefined
3733     \def\in@#1#2{%
3734         \def\in@@##1#1##2##3\in@{%
3735             \ifx\in@##2\in@false\else\in@true\fi}%
3736         \in@@#2#1\in@\in@@}
3737 \else
3738     \let\bbl@tempa\@empty
3739 \fi
3740 \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).

```

3741 \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.

```

3742 \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 $\LaTeX 2_{\epsilon}$ versions; just enough to make things work in plain \TeX environments.

```

3743 \ifx\@tempcnta\@undefined
3744     \csname newcount\endcsname\@tempcnta\relax
3745 \fi
3746 \ifx\@tempcntb\@undefined
3747     \csname newcount\endcsname\@tempcntb\relax
3748 \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`).

```

3749 \ifx\bye\@undefined
3750     \advance\count10 by -2\relax

```



```

3751 \fi
3752 \ifx\@ifnextchar\@undefined
3753   \def\@ifnextchar#1#2#3{%
3754     \let\reserved@d=#1%
3755     \def\reserved@a{#2}\def\reserved@b{#3}%
3756     \futurelet\@let@token\@ifnch}
3757   \def\@ifnch{%
3758     \ifx\@let@token\@sptoken
3759       \let\reserved@c\@xifnch
3760     \else
3761       \ifx\@let@token\reserved@d
3762         \let\reserved@c\reserved@a
3763       \else
3764         \let\reserved@c\reserved@b
3765       \fi
3766     \fi
3767     \reserved@c}
3768   \def\:{\let\@sptoken= } \: % this makes \@sptoken a space token
3769   \def\:{\@xifnch} \expandafter\def\:{\futurelet\@let@token\@ifnch}
3770 \fi
3771 \def\@testopt#1#2{%
3772   \@ifnextchar[#{1}{#1[#2]}}
3773 \def\@protected@testopt#1{%
3774   \ifx\protect\@typeset@protect
3775     \expandafter\@testopt
3776   \else
3777     \@x@protect#1%
3778   \fi}
3779 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
3780   #2\relax}\fi}
3781 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
3782   \else\expandafter\@gobble\fi{#1}}

```

15.4 Encoding related macros

Code from `ltoutenc.dtx`, adapted for use in the plain \TeX environment.

```

3783 \def\DeclareTextCommand{%
3784   \@dec@text@cmd\providecommand
3785 }
3786 \def\ProvideTextCommand{%
3787   \@dec@text@cmd\providecommand
3788 }
3789 \def\DeclareTextSymbol#1#2#3{%
3790   \@dec@text@cmd\chardef#1{#2}#3\relax
3791 }
3792 \def\@dec@text@cmd#1#2#3{%
3793   \expandafter\def\expandafter#2%
3794     \expandafter{%
3795       \csname#3-cmd\expandafter\endcsname
3796       \expandafter#2%
3797       \csname#3\string#2\endcsname
3798     }%
3799 %   \let\@ifdefinable\rc@ifdefinable
3800   \expandafter#1\csname#3\string#2\endcsname
3801 }
3802 \def\@current@cmd#1{%
3803   \ifx\protect\@typeset@protect\else
3804     \noexpand#1\expandafter\@gobble

```

```

3805 \fi
3806 }
3807 \def\@changed@cmd#1#2{%
3808 \ifx\protect\@typeset@protect
3809 \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
3810 \expandafter\ifx\csname ?\string#1\endcsname\relax
3811 \expandafter\def\csname ?\string#1\endcsname{%
3812 \@changed@x@err{#1}%
3813 }%
3814 \fi
3815 \global\expandafter\let
3816 \csname\cf@encoding\string#1\expandafter\endcsname
3817 \csname ?\string#1\endcsname
3818 \fi
3819 \csname\cf@encoding\string#1%
3820 \expandafter\endcsname
3821 \else
3822 \noexpand#1%
3823 \fi
3824 }
3825 \def\@changed@x@err#1{%
3826 \errhelp{Your command will be ignored, type <return> to proceed}%
3827 \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
3828 \def\DeclareTextCommandDefault#1{%
3829 \DeclareTextCommand#1?%
3830 }
3831 \def\ProvideTextCommandDefault#1{%
3832 \ProvideTextCommand#1?%
3833 }
3834 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
3835 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
3836 \def\DeclareTextAccent#1#2#3{%
3837 \DeclareTextCommand#1{#2}[1]{\accent#3 #1}
3838 }
3839 \def\DeclareTextCompositeCommand#1#2#3#4{%
3840 \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
3841 \edef\reserved@b{\string#1}%
3842 \edef\reserved@c{%
3843 \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
3844 \ifx\reserved@b\reserved@c
3845 \expandafter\expandafter\expandafter\ifx
3846 \expandafter\@car\reserved@a\relax\relax\@nil
3847 \@text@composite
3848 \else
3849 \edef\reserved@b##1{%
3850 \def\expandafter\noexpand
3851 \csname#2\string#1\endcsname####1{%
3852 \noexpand\@text@composite
3853 \expandafter\noexpand\csname#2\string#1\endcsname
3854 ####1\noexpand\@empty\noexpand\@text@composite
3855 {##1}%
3856 }%
3857 }%
3858 \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
3859 \fi
3860 \expandafter\def\csname\expandafter\string\csname
3861 #2\endcsname\string#1-\string#3\endcsname{#4}
3862 \else
3863 \errhelp{Your command will be ignored, type <return> to proceed}%

```

```

3864 \errmessage{\string\DeclareTextCompositeCommand\space used on
3865 inappropriate command \protect#1}
3866 \fi
3867 }
3868 \def\@text@composite#1#2#3\@text@composite{%
3869 \expandafter\@text@composite@x
3870 \csname\string#1-\string#2\endcsname
3871 }
3872 \def\@text@composite@x#1#2{%
3873 \ifx#1\relax
3874 #2%
3875 \else
3876 #1%
3877 \fi
3878 }
3879 %
3880 \def\@strip@args#1:#2-#3\@strip@args{#2}
3881 \def\DeclareTextComposite#1#2#3#4{%
3882 \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
3883 \bgroup
3884 \lccode\@=#4%
3885 \lowercase{%
3886 \egroup
3887 \reserved@a \@
3888 }%
3889 }
3890 %
3891 \def\UseTextSymbol#1#2{%
3892 % \let\@curr@enc\cf@encoding
3893 % \@use@text@encoding{#1}%
3894 #2%
3895 % \@use@text@encoding\@curr@enc
3896 }
3897 \def\UseTextAccent#1#2#3{%
3898 % \let\@curr@enc\cf@encoding
3899 % \@use@text@encoding{#1}%
3900 % #2{\@use@text@encoding\@curr@enc\selectfont#3}%
3901 % \@use@text@encoding\@curr@enc
3902 }
3903 \def\@use@text@encoding#1{%
3904 % \edef\font@encoding{#1}%
3905 % \xdef\font@name{%
3906 % \csname\curr@fontshape/\font@size\endcsname
3907 % }%
3908 % \pickup@font
3909 % \font@name
3910 % \@@enc@update
3911 }
3912 \def\DeclareTextSymbolDefault#1#2{%
3913 \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
3914 }
3915 \def\DeclareTextAccentDefault#1#2{%
3916 \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
3917 }
3918 \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.

```

3919 \DeclareTextAccent{"}{OT1}{127}

```

```

3920 \DeclareTextAccent{\'}{OT1}{19}
3921 \DeclareTextAccent{\^}{OT1}{94}
3922 \DeclareTextAccent{\`}{OT1}{18}
3923 \DeclareTextAccent{\~}{OT1}{126}

```

The following control sequences are used in `babel.def` but are not defined for plain \TeX .

```

3924 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
3925 \DeclareTextSymbol{\textquotedblright}{OT1}{`\"}
3926 \DeclareTextSymbol{\textquoteleft}{OT1}{`\'}
3927 \DeclareTextSymbol{\textquoteright}{OT1}{`\'}
3928 \DeclareTextSymbol{\i}{OT1}{16}
3929 \DeclareTextSymbol{\ss}{OT1}{25}

```

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

```

3930 \ifx\scriptsize@undefined
3931   \let\scriptsize\sevenrm
3932 \fi

```

15.5 Babel options

The file `babel.def` expects some definitions made in the \LaTeX style file. So we must provide them 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).

```

3933 \let\bbbl@opt@shorthands\@nnil
3934 \def\bbbl@ifshorthand#1#2#3{#2}%
3935 \ifx\babeloptionstrings@undefined
3936   \let\bbbl@opt@strings\@nnil
3937 \else
3938   \let\bbbl@opt@strings\babeloptionstrings
3939 \fi
3940 \def\bbbl@tempa{normal}
3941 \ifx\babeloptionmath\bbbl@tempa
3942   \def\bbbl@mathnormal{\noexpand\textormath}
3943 \fi
3944 \def\BabelStringsDefault{generic}
3945 \ifx\BabelModifiers@undefined\let\BabelModifiers\relax\fi
3946 \let\bbbl@afterlang\relax
3947 \let\bbbl@language@opts\@empty
3948 \ifx\@uclclist\@undefined\let\@uclclist\@empty\fi
3949 \def\AfterBabelLanguage#1#2{}
3950 \</plain>

```

16 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.