

Babel

Version 3.11
2017/06/26

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). New features related to font selection, bidi writing and the like will be added incrementally.

Contents

I	User guide	4
1	The user interface	4
1.1	Monolingual documents	4
1.2	Multilingual documents	5
1.3	Modifiers	5
1.4	Notes	5
1.5	Plain	6
1.6	Basic language selectors	6
1.7	Auxiliary language selectors	7
1.8	More on selection	8
1.9	Getting the current language name	9
1.10	Selecting scripts	9
1.11	Shorthands	10
1.12	Package options	13
1.13	The base option	15
1.14	Creating a language	16
1.15	Hooks	17
1.16	Hyphenation tools	18
1.17	Language attributes	20
1.18	Languages supported by babel	20
1.19	Tips, workarounds, know issues and notes	21
1.20	Future work	22
1.21	Tentative and experimental code	23
2	Loading languages with language.dat	24
2.1	Format	24
3	The interface between the core of babel and the language definition files	25
3.1	Basic macros	26
3.2	Skeleton	28
3.3	Support for active characters	29
3.4	Support for saving macro definitions	29
3.5	Support for extending macros	29
3.6	Macros common to a number of languages	30
3.7	Encoding-dependent strings	30
4	Changes	34
4.1	Changes in babel version 3.9	34
4.2	Changes in babel version 3.7	34
4.3	Changes in babel version 3.6	36
4.4	Changes in babel version 3.5	37
II	The code	37
5	Identification and loading of required files	37
6	Tools	38
6.1	Multiple languages	41

7	The Package File (<code>\LaTeX</code>, <code>babel.sty</code>)	43
7.1	base	43
7.2	key=value options and other general option	44
7.3	Conditional loading of shorthands	45
7.4	Language options	46
8	The kernel of Babel (<code>babel.def</code>, common)	49
8.1	Tools	50
8.2	Hooks	52
8.3	Setting up language files	54
8.4	Shorthands	56
8.5	Language attributes	65
8.6	Support for saving macro definitions	68
8.7	Short tags	69
8.8	Hyphens	69
8.9	Multiencoding strings	71
8.10	Macros common to a number of languages	77
8.11	Making glyphs available	77
	8.11.1 Quotation marks	77
	8.11.2 Letters	79
	8.11.3 Shorthands for quotation marks	80
	8.11.4 Umlauts and tremas	81
9	The kernel of Babel (<code>babel.def</code>, only <code>\LaTeX</code>)	82
9.1	The redefinition of the style commands	82
9.2	Creating languages	83
9.3	Cross referencing macros	86
9.4	Marks	90
9.5	Preventing clashes with other packages	91
	9.5.1 <code>ifthen</code>	91
	9.5.2 <code>varioref</code>	92
	9.5.3 <code>hhline</code>	92
	9.5.4 <code>hyperref</code>	93
	9.5.5 <code>fancyhdr</code>	93
9.6	Encoding and fonts	93
9.7	Basic bidi support	95
9.8	Local Language Configuration	97
10	Multiple languages (<code>switch.def</code>)	98
10.1	Selecting the language	99
10.2	Errors	107
11	Loading hyphenation patterns	108
12	Tentative font handling with <code>fontspec</code>	113
13	Hooks for <code>XeTeX</code> and <code>LuaTeX</code>	114
13.1	<code>XeTeX</code>	114
13.2	<code>LuaTeX</code>	115
14	The ‘<code>nil</code>’ language	121

15 Support for Plain T_EX (plain.def)	122
15.1 Not renaming hyphen.tex	122
15.2 Emulating some L ^A T _E X features	123
15.3 General tools	124
15.4 Encoding related macros	127
15.5 Babel options	130
16 Tentative font handling	131
17 Hooks for XeT_EX and LuaT_EX	132
17.1 XeT _E X	132
17.2 LuaT _E X	133
18 Conclusion	139
19 Acknowledgements	139

Part I

User guide

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. 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 `french` 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 L^AT_EX, the preamble of the document:

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

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

You can also set the main language explicitly:

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

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

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

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

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 Notes

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

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

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

```
! 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.

1.5 Plain

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

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

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

1.6 Basic language selectors

This section describes the commands to be used in the document to switch the language in multilingual document. 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.

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

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

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

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

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

\foreignlanguage {<language>}{<text>}

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

1.7 Auxiliary language selectors

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

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

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

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

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

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

This environment was originally intended for intermixing left-to-right typesetting with right-to-left typesetting in engines not supporting a change in the writing direction inside a line. However, by default it never complied with the documented behaviour and it was 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).

1.8 More on selection

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

New 3.9i In multilingual documents with many language switches the commands above can be cumbersome. With this tool shorter names can be defined. It adds nothing really new – it is just syntactical sugar. It defines `\text{\langle tag1 \rangle{\langle text \rangle}}` to be `\foreignlanguage{\langle language1 \rangle}{\langle text \rangle}`, and `\begin{\langle tag1 \rangle}` to be `\begin{otherlanguage*}{\langle language1 \rangle}`, and so on. Note `\langle tag1 \rangle` is also allowed, but remember to set it locally inside a group. So, with

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

yo can write

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

and

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

Something like `\babeltag{finnish = finnish}` is legitimate – it defines `\textfinnish` and `\finnish`.

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

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

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

Of course, T_EX 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. 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}
```

⁵With it encoded string may not work as expected.

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

1.9 Getting the current language name

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

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

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

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

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

1.10 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.⁶

The same applies for bidi texts – writing direction is intrinsic to each writing system and therefore it is best set by the language (which could be a dummy one). 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.⁷

`\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.

⁶The 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.

⁷But still defined for backwards compatibility.

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.

1.11 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 in the keyboards (and sometimes not even as separated or precomposed Unicode characters). As to the point 2, now `pdfTeX` provides `\knbcode`. Tools of point 3 can be still very useful in general.

There are three levels of shorthands: *user*, *language*, and *system* (by order of precedence). Version 3.9 introduces the *language user* level on top of the user level, as described below. In most cases, you will use only shorthands provided by languages.

Please, 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⟩,⟨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⟩`). 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 discretionary (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{"}{/}`. *Please note* the substitute character must *not* have been declared before as shorthand (in such case, `\aliasshorthands` is ignored). The following example shows how to replace a shorthand by another

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

However, 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, ^ calls `\active@char~` or `\normal@char~`). Furthermore, if you change the system value of ^ with `\defineshorthand` nothing happens.

\languageshorthands {<language>}

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).⁸ 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 {<shorthand>}

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

⁸Actually, 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.

For your records, here is a list of shorthands, but you must double check them, as they may change:⁹

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.¹⁰

1.12 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= <char><char>... | 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

⁹Thanks to Enrico Gregorio

¹⁰This declaration serves to nothing, but it is preserved for backward compatibility.

expanded by L^AT_EX 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 L^AT_EX 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= `<file>`

Load `<file>.cfg` instead of the default config file `bblopts.cfg` (the file is loaded even with `noconfigs`).

main= `<language>`

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= `<language>`

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.¹¹

¹¹You can use alternatively the package `silence`.

strings= generic | unicode | encoded | $\langle label \rangle$ | $\langle font\ encoding \rangle$

Selects the encoding of strings in languages supporting this feature. Predefined labels are generic (for traditional 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.¹² 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;¹³

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.¹⁴

1.13 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!).

For 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:

¹²Turned off in plain.

¹³Duplicated options count as several ones.

¹⁴Providing 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.


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

1.14 Creating a language

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

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

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

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

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

captions= *language-tag*

Currently, the [*options*] may be used to load captions from an ini file, but note these files are still tentative and might change. For example:

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

It just loads the strings, no more. Encoding, font, fontspec language and script, writing direction, etc., are not touched at all. Unicode engines load the UTF-8 variants, while 8-bit engines load the LICR (ie, with macros like \’ or \ss) ones.

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.

NOTE (1) Setting `\today` is not so easy, and requires more work (some tools are on the way). (2) If you need shorthands, you can use `\useshorthands` and `\defineshorthand` as described above. (3) Captions and `\today` are “ensured” with `\babelensure` (because this will be the default in ini-based languages).

1.15 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 `\useshorthands*` to add a hook for the event `afterextras`).

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

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

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

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

defaultcommands Used (locally) in `\StartBabelCommands`.

encodedcommands (input, font encodings) Used (locally) in `\StartBabelCommands`.

Both `xetex` and `luatex` make sure the encoded text is read correctly.

stopcommands Used to reset the the above, if necessary.

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

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

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

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

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

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

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

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

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

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

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

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

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

\BabelContentsFiles **New 3.9a** This macro contains a list of “toc” types which require 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).

1.16 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 L^AT_EX: (1) the character used is that set for the current font, while in L^AT_EX it is hardwired to - (a typical value); (2) the hyphen to be used in fonts with a negative `\hyphenchar` is -, like in L^AT_EX, but it can be changed to another value by redefining `\babelnulhyphen`; (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*,¹⁵ adds or replaces patterns for the languages given or, without the optional argument, for *all* languages. If a pattern for a certain combination already exists, it gets replaced by the new one.

It can be used only in the preamble, and patterns are added when the language is first selected, thus taking into account changes of `\lccodes`’s done in

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

`\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.

1.17 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).

1.18 Languages supported by babel

In the following table most of the languages supported by babel are listed, together with the names of the options which you can load babel with for each language. Note this list is open and the current options may be different.

Afrikaans	afrikaans
Bahasa	bahasa, indonesian, indon, bahasai, bahasam, malay, melayu
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
Interlingua	interlingua
Irish Gaelic	irish
Italian	italian
Latin	latin

Lower Sorbian lowersorbian
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 L^AT_EX.

1.19 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), L^AT_EX 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 `\usesshorthands` to activate ' and `\defineshorthand`, 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.

mkpattern Generates hyphenation patterns.

tracklang Tracks which languages have been requested.

1.20 Future work

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

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.

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.

internals.

More 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.” may be referred to as either “ítem 3.º” or “3.º ítem”, and so on. Even more interesting is right-to-left, vertical and bidi typesetting. 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).

1.21 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, which solve the two main cases, are provided by babel ($\geq 3.9g$) with a partial solution (only xetex and luatex, for obvious reasons), but use them at your own risk, as they might be removed in the future.

- `\babelFSstore{<babel-language>}` sets the current three basic families (rm, sf, tt) as the default for the language given. In most cases, this macro will be enough.
- `\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). Use it only if you select some fonts in the document with `\fontspec`.

So, for example:

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

Note you can set any feature required for the language – not only Language, but also Script and even raw features. This makes those macros a bit more verbose, but also more powerful.

Bidi writing is taking its *first steps*. Here is a simple example:

```
\documentclass{article}

\usepackage[english]{babel}
\babelprovide{arabic} % declare a new empty language

\usepackage{fontenc}
\setmainfont[Script=Arabic,Language=Arabic]{Traditional Arabic}
\babelFSstore[Arabic]{arabic}

\begin{document}
English \foreignlanguage{arabic}{Arabic} English
\end{document}
```


First steps means exactly that. For example, in luatex digits and short Latin texts must be marked up explicitly in RL mode. On the other hand, xetex poses quite different challenges. The bidi mechanism is activated when an RL script is passed as the new optional argument of `\babelFSstore`. See the code section for `\foreignlanguage*` (a new starred version of `\foreignlanguage`).

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

Unfortunately, the new model is intrinsically incompatible with the previous one, which means you can experience some problems with polyglossia. If using the latter, you must load the patterns with babel as shown in the following example:

```
\usepackage[base,french,dutch,spanish,english]{babel}
\usepackage{polyglossia}
\setmainlanguage{french}
\setotherlanguages{dutch,spanish,english}
```

Be aware this is, very likely, a temporary solution.

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.

The file can contain empty lines and comments, as well as lines which start with an equals (=) sign. Such a line will instruct L^AT_EX 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
```

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.

```
=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.
- The language definition files must define five macros, used to activate and deactivate the language-specific definitions. These macros are `\<lang>hyphenmins`, `\captions<lang>`, `\date<lang>`, `\extras<lang>` and `\noextras<lang>` (the last two may be left empty); where `<lang>` is either the name of the language definition file or the name of the \LaTeX option that is to be used. These macros and their functions are discussed below. You must define all or none for a language (or a dialect); defining, say, `\date<lang>` but not `\captions<lang>` does not raise an error but can lead to unexpected results.

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

- 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 `plain.tex` version 3.x. For older versions of `plain.tex` and `lplain.tex` a substitute definition is used. Here “language” is used in the \TeX sense of set of hyphenation patterns.

`\adddialect` The macro `\adddialect` can be used when two languages can (or must) use the

²²But not removed, for backward compatibility.

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 `\language0`. Here “language” is used in the T_EX sense of set of hyphenation patterns.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

`\substitutefontfamily` (Deprecated.) This command takes three arguments, a font encoding and two font

family names. It creates a font description file for the first font in the given encoding. This .fd file will instruct L^AT_EX 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. ?? (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@attribute{<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.

<code>\initiate@active@char</code>	The internal macro <code>\initiate@active@char</code> is used in language definition files to instruct L ^A T _E X 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.
<code>\bbl@activate</code> <code>\bbl@deactivate</code>	The command <code>\bbl@activate</code> is used to change the way an active character expands. <code>\bbl@activate</code> ‘switches on’ the active behaviour of the character. <code>\bbl@deactivate</code> lets the active character expand to its former (mostly) non-active self.
<code>\declare@shorthand</code>	The macro <code>\declare@shorthand</code> 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 <i>not</i> raise an error if the shorthand character has not been “initiated”.)
<code>\bbl@add@special</code> <code>\bbl@remove@special</code>	The T _E Xbook states: “Plain T _E X includes a macro called <code>\dospecials</code> that is essentially a set macro, representing the set of all characters that have a special category code.” [?, p. 380] It is used to set text ‘verbatim’. To make this work if more characters get a special category code, you have to add this character to the macro <code>\dospecials</code> . L ^A T _E X adds another macro called <code>\@sanitize</code> representing the same character set, but without the curly braces. The macros <code>\bbl@add@special⟨char⟩</code> and <code>\bbl@remove@special⟨char⟩</code> add and remove the character <code>⟨char⟩</code> 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.

<code>\babel@save</code>	To save the current meaning of any control sequence, the macro <code>\babel@save</code> is provided. It takes one argument, <code>⟨curname⟩</code> , the control sequence for which the meaning has to be saved.
<code>\babel@savevariable</code>	A second macro is provided to save the current value of a variable. In this context, anything that is allowed after the <code>\the</code> primitive is considered to be a variable. The macro takes one argument, the <code>⟨variable⟩</code> . The effect of the preceding macros is to append a piece of code to the current definition of <code>\originalTeX</code> . When <code>\originalTeX</code> 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

<code>\addto</code>	The macro <code>\addto{⟨control sequence⟩}{⟨T_EX code⟩}</code> can be used to extend the definition of a macro. The macro need not be defined (ie, it can be undefined or <code>\relax</code>). This macro can, for instance, be used in adding instructions to a macro like <code>\extrasenglish</code> . 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
---------------------	--

²³This mechanism was introduced by Bernd Raichle.

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

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

3.7 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 translated to the internal representation used by the engine, typically `utf8`, which is the only value supported currently (default is no translations). Note `charset` is applied by `luatex` and `xetex` when reading the file, not when the macro or string is used in the document.

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

Blocks without a selector are read always if the key `strings` has been used. They provide fallback values, and therefore must be the last blocks; they should be provided always if possible and all strings should be defined somehow inside it; they can be the only blocks (mainly LGC scripts using the LICR). Blocks without a selector can be activated explicitly with `strings=generic` (no block is taken into account except those). With `strings=encoded`, strings in those blocks are set as default (internally, `?`). With `strings=encoded` strings are protected, but they are correctly expanded in `\MakeUppercase` and the like. If there is no key `strings`, string definitions are ignored, but `\SetCases` are still honoured (in an encoded way). The `<category>` is either `captions`, `date` or `extras`. You must stick to these three categories, even if no error is raised when using other name.²⁴ It may be empty, too, but in such a case using `\SetString` is an error (but not `\SetCase`).

```
\StartBabelCommands{language}{captions}
  [unicode, fontenc=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=EU1 EU2, charset=utf8]
\SetString\monthiname{Jänner}

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

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

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

²⁴In future releases further categories may be added.


```

\StartBabelCommands{german,austrian}{date}
  \SetString\monthiiname{Februar}
  \SetString\monthiiname{M\{"a}rz}
  \SetString\monthivname{April}
  \SetString\monthvname{Mai}
  \SetString\monthviname{Juni}
  \SetString\monthviiname{Juli}
  \SetString\monthviiiname{August}
  \SetString\monthixname{September}
  \SetString\monthxname{Oktober}
  \SetString\monthxiname{November}
  \SetString\monthxiiname{Dezenber}
  \SetString\today{\number\day.\sim%
    \csname month\romannumeral\month name\endcsname\space
    \number\year}

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

\EndBabelCommands

```

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

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

The starred version just forces strings to take a value – if not set as package option, then the default for the engine is used. This is not done by default to prevent backward incompatibilities, but if you are creating a new language this version is better. It's up to the maintainers of the current languages to decide if using it is appropriate.²⁵

\EndBabelCommands Marks the end of the series of blocks.

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

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

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

Adds $\langle macro-name \rangle$ to the current category, and defines globally $\langle lang-macro-name \rangle$ to $\langle code \rangle$ (after applying the transformation corresponding to the current charset or defined with the hook `stringprocess`). Use this command to define strings, without including any “logic” if possible, which should be a separated macro. See the example above for the date.

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

²⁵This replaces in 3.9g a short-lived $\backslash\text{UseStrings}$ which has been removed because it did not work.

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

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

`#1` is replaced by the roman numeral.

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

Sets globally code to be executed at `\MakeUppercase` and `\MakeLowercase`. The code would be typically things like `\let\BB\bb` and `\uccode` or `\lccode` (although for the reasons explained above, changes in lc/uc codes may not work). A `<map-list>` is a series of macros using the internal format of `\@uclclist` (eg, `\bb\BB\cc\CC`). The mandatory arguments take precedence over the optional one. This command, unlike `\SetString`, is executed always (even without strings), and it is intended for minor readjustments only.

For example, as T1 is the default case mapping in L^AT_EX, we could set for Turkish:

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

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

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

\EndBabelCommands
```

(Note the mapping for OT1 is not complete.)

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

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

There are three helper macros to be used inside `\SetHyphenMap`:

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

lccode, which is also increased (MM stands for *many-to-many*).

- `\BabelLowerM0{⟨uccode-from⟩}{⟨uccode-to⟩}{⟨step⟩}{⟨lccode⟩}` loops through the given uppercase codes, using the step, and assigns them the lccode, which is fixed (M0 stands for *many-to-one*).

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

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

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

4 Changes

4.1 Changes in babel version 3.9

Most of changes in version 3.9 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`. 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 Greek has been enhanced. Code from the `kdgree` package (suggested by the author) was added and `\greeknumeral` has been added.
- Support for typesetting Basque is now available thanks to Juan Aguirregabiria.
- Support for typesetting Serbian with Latin script is now available thanks to Dejan Muhamedagić and Jankovic Slobodan.
- 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.
- Support for typesetting Bulgarian is now available thanks to Georgi Boshnakov.
- Support for typesetting Latin is now available, thanks to Claudio Beccari and Krzysztof Konrad Żelechowski.
- Support for typesetting North Sami is now available, thanks to Regnor Jernsletten.
- The options `canadian`, `canadien` and `acadien` have been added for Canadian English and French use.
- 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* (“Polutoniko” or multi-accented) Greek way of typesetting texts. These attributes will possibly find wider use in future releases.
- 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.

4.3 Changes in babel version 3.6

In babel version 3.6 a number of bugs that were found in version 3.5 are fixed. Also a number of changes and additions have occurred:

- A new environment `otherlanguage*` is introduced. it only switches the ‘specials’, but leaves the ‘captions’ untouched.
- The shorthands are no longer fully expandable. Some problems could only be solved by peeking at the token following an active character. The advantage is that ‘`{}`’a works as expected for languages that have the ‘ active.
- Support for typesetting french texts is much enhanced; the file `francais.ldf` is now replaced by `french.ldf` which is maintained by Daniel Flipo.
- Support for typesetting the russian language is again available. The language definition file was originally developed by Olga Lapko from CyrTUG. The fonts needed to typeset the russian language are now part of the babel distribution. The support is not yet up to the level which is needed according to Olga, but this is a start.
- Support for typesetting greek texts is now also available. What is offered in this release is a first attempt; it will be enhanced later on by Yannis Haralambous.
- in babel 3.6j some hooks have been added for the development of support for Hebrew typesetting.
- Support for typesetting texts in Afrikaans (a variant of Dutch, spoken in South Africa) has been added to `dutch.ldf`.
- Support for typesetting Welsh texts is now available.
- A new command `\aliasshorthand` is introduced. It seems that in Poland various conventions are used to type the necessary Polish letters. It is now possible to use the character `/` as a shorthand character instead of the character `"`, by issuing the command `\aliasshorthand{"}{/}`.
- The shorthand mechanism now deals correctly with characters that are already active.
- Shorthand characters are made active at `\begin{document}`, not earlier. This is to prevent problems with other packages.
- A *preambleonly* command `\substitutefontfamily` has been added to create `.fd` files on the fly when the font families of the Latin text differ from the families used for the Cyrillic or Greek parts of the text.
- Three new commands `\LdfInit`, `\ldf@quit` and `\ldf@finish` are introduced that perform a number of standard tasks.
- In babel 3.6k the language Ukrainian has been added and the support for Russian typesetting has been adapted to the package ‘cyrillic’ to be released with the December 1998 release of L^AT_EX 2_ε.

4.4 Changes in babel version 3.5

In babel version 3.5 a lot of changes have been made when compared with the previous release. Here is a list of the most important ones:

- the selection of the language is delayed until `\begin{document}`, which means you must add appropriate `\selectlanguage` commands if you include `\hyphenation` lists in the preamble of your document.
- babel now has a language environment and a new command `\foreignlanguage`;
- the way active characters are dealt with is completely changed. They are called ‘shorthands’; one can have three levels of shorthands: on the user level, the language level, and on ‘system level’. A consequence of the new way of handling active characters is that they are now written to auxiliary files ‘verbatim’;
- A language change now also writes information in the `.aux` file, as the change might also affect typesetting the table of contents. The consequence is that an `.aux` file generated by a `LaTeXformat` with babel preloaded gives errors when read with a `LaTeXformat` without babel; but I think this probably doesn’t occur;
- babel is now compatible with the `inputenc` and `fontenc` packages;
- the language definition files now have a new extension, `ldf`;
- the syntax of the file `language.dat` is extended to be compatible with the `french` package by Bernard Gaulle;
- each language definition file looks for a configuration file which has the same name, but the extension `.cfg`. It can contain any valid `LaTeX` code.

NB

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 it 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 L^AT_EX package, which set options and load language styles.
plain.def defines some L^AT_EX 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=3.11>>
2 <<date=2017/06/26>>
```

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 L^AT_EX is executed twice, but we need them when defining options and babel.def cannot be load until options have been defined. This does not hurt, but should be fixed somehow.

```
3 <<*Basic macros>> ≡
4 \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@csarg#1#2{\expandafter#1\csname bbl@#2\endcsname}%
10 \def\bbl@loop#1#2#3{\bbl@loop#1{#3}#2,\@nnil,}
11 \def\bbl@loopx#1#2{\expandafter\bbl@loop\expandafter#1\expandafter{#2}}
12 \def\bbl@loop#1#2#3,{%
13   \ifx\@nnil#3\relax\else
14     \def#1{#3}#2\bbl@afterfi\bbl@loop#1{#2}%
15   \fi}
16 \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.

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

\bbl@afterelse Because the code that is used in the handling of active characters may need to look ahead, we take extra care to ‘throw’ it over the \else and \fi parts of an

\if-statement²⁶. These macros will break if another \if... \fi statement appears in one of the arguments and it is not enclosed in braces.

```
23 \long\def\bbl@afterelse#1\else#2\fi{\fi#1}
24 \long\def\bbl@afterfi#1\fi{\fi#1}
```

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.

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

To check if a macro is defined, we create a new macro, which does the same as \@ifundefined. However, in an ϵ -tex engine, it is redefined more robust

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

A tool from url, by Donald Arseneau, which test if a string is empty or space.

```
56 \def\bbl@ifblank#1{%
57   \bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
58 \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

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

macro thus named, not an empty argument, which is what you get with <key>= and no value).

```

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

```

70 \def\bb@vforeach#1#2{%
71   \def\bb@forcmd##1{#2}%
72   \bb@fornext#1,\@nil,}
73 \def\bb@fornext#1,{%
74   \ifx\@nil#1\relax\else
75     \bb@ifblank{#1}{\bb@trim\bb@forcmd{#1}}%
76     \expandafter\bb@fornext
77   \fi}
78 \def\bb@foreach#1{\expandafter\bb@vforeach\expandafter{#1}}
79 \def\bb@replace#1#2#3{% in #1 -> repl #2 by #3
80   \toks@{}%
81   \def\bb@replace@aux##1#2##2#2{%
82     \ifx\bb@nil##2%
83       \toks@\expandafter{\the\toks@##1}%
84     \else
85       \toks@\expandafter{\the\toks@##1#3}%
86       \bb@afterfi
87       \bb@replace@aux##2#2%
88     \fi}%
89   \expandafter\bb@replace@aux#1#2\bb@nil#2%
90   \edef#1{\the\toks@}}

```

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.

```

91 \def\bb@exp#1{%
92   \begingroup
93   \let\\\noexpand
94   \def\<##1>{\expandafter\noexpand\csname##1\endcsname}%
95   \edef\bb@exp@aux{\endgroup#1}%
96   \bb@exp@aux}

```

Two more tools. \bb@samestring first expand its arguments and then compare their expansion (sanitized, so that the catcodes do not matter). \bb@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.

```

97 \def\bb@ifsamestring#1#2{%

```

```

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

```

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

```

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

```

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

```

134 <<*Load macros for plain if not LaTeX>> \equiv
135 \ifx\AtBeginDocument\@undefined
136 \input plain.def\relax
137 \fi
138 <</Load macros for plain if not LaTeX>>

```

6.1 Multiple languages

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

and hyphen.cfg; the latter may seem redundant, but remember babel doesn't require loading switch.def in the format.

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

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

Identify each file that is produced from this source file.

```
160 <*driver&!user>
161 \ProvidesFile{babel.drv}[<<date>> <<version>>]
162 </driver&!user>
163 <*driver & user>
164 \ProvidesFile{user.drv}[<<date>> <<version>>]
165 </driver & user>
```

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 (L^AT_EX, `babel.sty`)

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

Apart from all the language options below we also have a few options that influence the 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 L^AT_EX forgets about the first loading. After `switch.def` has been loaded (above) and `\AfterBabelLanguage` defined, exits.

```

166 (*package)
167 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
168 \ProvidesPackage{babel}[\langle date \rangle \langle version \rangle The Babel package]
169 \ifpackagewith{babel}{debug}
170   {\let\bbl@debug\@firstofone}
171   {\let\bbl@debug\@gobble}
172 \input switch.def\relax
173 \langle Load patterns in luatex \rangle
174 \langle Basic macros \rangle
175 \def\AfterBabelLanguage#1{%
176   \global\expandafter\bbl@add\csname#1.ldf-h\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.

```

177 \ifx\bbl@languages\undefined\else
178   \begingroup
179     \catcode'\^^I=12
180     \@ifpackagewith{babel}{showlanguages}{%
181       \begingroup
182         \def\bbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}%
183         \wlog{<*languages>}%
184         \bbl@languages
185         \wlog{</languages>}%
186       \endgroup}{%
187     \endgroup
188     \def\bbl@elt#1#2#3#4{%
189       \ifnum#2=\z@
190         \gdef\bbl@nulllanguage{#1}%
191         \def\bbl@elt##1##2##3##4{}}%
192     \fi}%
193 \bbl@languages

```

```
194 \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.

```
195 \@ifpackagewith{babel}{base}{%
196   \ifx\directlua\undefined
197     \DeclareOption*{\bbl@patterns{\CurrentOption}}%
198   \else
199     \DeclareOption*{\bbl@patterns@lua{\CurrentOption}}%
200   \fi
201   \DeclareOption{base}{}%
202   \DeclareOption{showlanguages}{}%
203   \ProcessOptions
204   \global\expandafter\let\csname opt@babel.sty\endcsname\relax
205   \global\expandafter\let\csname ver@babel.sty\endcsname\relax
206   \global\let\@ifl@ter@@\@ifl@ter
207   \def\@ifl@ter#1#2#3#4#5{\global\let\@ifl@ter\@ifl@ter@@}%
208   \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).

```
209 \bbl@csarg\let{tempa\expandafter}\csname opt@babel.sty\endcsname
210 \def\bbl@tempb#1.#2{%
211   #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%
212 \def\bbl@tempd#1.#2\@nnil{%
213   \ifx\@empty#2%
214     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
215   \else
216     \in@{=}{#1}\ifin@
217     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
218   \else
219     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
220     \bbl@csarg\edef{mod@#1}{\bbl@tempb#2}%
221   \fi
222   \fi}
223 \let\bbl@tempc\@empty
224 \bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty\@nnil}
225 \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.

```
226 \DeclareOption{KeepShorthandsActive}{}
227 \DeclareOption{activeacute}{}
228 \DeclareOption{activegrave}{}
229 \DeclareOption{debug}{}
230 \DeclareOption{noconfigs}{}
231 \DeclareOption{showlanguages}{}%
```

```

232 \DeclareOption{silent}{}
233 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
234 <<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.

```

235 \let\bbl@opt@shorthands\@nnil
236 \let\bbl@opt@config\@nnil
237 \let\bbl@opt@main\@nnil
238 \let\bbl@opt@headfoot\@nnil

```

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

```

239 \def\bbl@tempa#1=#2\bbl@tempa{%
240   \bbl@csarg\ifx{opt@#1}\@nnil
241     \bbl@csarg\edef{opt@#1}{#2}%
242   \else
243     \bbl@error{%
244       Bad option ‘#1=#2’. Either you have misspelled the\\%
245       key or there is a previous setting of ‘#1’}%
246     Valid keys are ‘shorthands’, ‘config’, ‘strings’, ‘main’,\\%
247     ‘headfoot’, ‘safe’, ‘math’}
248   \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.

```

249 \let\bbl@language@opts\@empty
250 \DeclareOption*{%
251   \@expandtwoargs\in@{\string=}{\CurrentOption}%
252   \ifin@
253     \expandafter\bbl@tempa\CurrentOption\bbl@tempa
254   \else
255     \bbl@add@list\bbl@language@opts{\CurrentOption}%
256   \fi}

```

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

```

257 \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=...`

```

258 \def\bbl@sh@string#1{%
259   \ifx#1\@empty\else
260     \ifx#1t\string~%
261     \else\ifx#1c\string,%
262     \else\string#1%
263   \fi\fi

```

```

264 \expandafter\bb@sh@string
265 \fi}
266 \ifx\bb@opt@shorthands\@nnil
267 \def\bb@ifshorthand#1#2#3{#2}%
268 \else\ifx\bb@opt@shorthands\@empty
269 \def\bb@ifshorthand#1#2#3{#3}%
270 \else

```

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

```

271 \def\bb@ifshorthand#1{%
272 \@expandtwoargs\in{\string#1}{\bb@opt@shorthands}%
273 \ifin@
274 \expandafter\@firstoftwo
275 \else
276 \expandafter\@secondoftwo
277 \fi}

```

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

```

278 \edef\bb@opt@shorthands{%
279 \expandafter\bb@sh@string\bb@opt@shorthands\@empty}%

```

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

```

280 \bb@ifshorthand{'}%
281 {\PassOptionsToPackage{activeacute}{babel}}{}
282 \bb@ifshorthand{'}%
283 {\PassOptionsToPackage{activegrave}{babel}}{}
284 \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.

```

285 \ifx\bb@opt@headfoot\@nnil\else
286 \g@addto@macro\@resetactivechars{%
287 \set@typeset@protect
288 \expandafter\select@language@x\expandafter{\bb@opt@headfoot}%
289 \let\protect\noexpand}
290 \fi

```

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

```

291 \ifx\bb@opt@safe\@undefined
292 \def\bb@opt@safe{BR}
293 \fi
294 \ifx\bb@opt@main\@nnil\else
295 \edef\bb@language@opts{%
296 \ifx\bb@language@opts\@empty\else\bb@language@opts,\fi
297 \bb@opt@main}
298 \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).

```

299 \let\bbl@afterlang\relax
300 \let\BabelModifiers\relax
301 \let\bbl@loaded\@empty
302 \def\bbl@load@language#1{%
303   \InputIfFileExists{#1.ldf}%
304   {\edef\bbl@loaded{\CurrentOption
305     \ifx\bbl@loaded\@empty\else,\bbl@loaded\fi}%
306     \expandafter\let\expandafter\bbl@afterlang
307       \csname\CurrentOption.ldf-h@k\endcsname
308     \expandafter\let\expandafter\BabelModifiers
309       \csname bbl@mod@\CurrentOption\endcsname}%
310   {\bbl@error{%
311     Unknown option ‘\CurrentOption’. Either you misspelled it\\%
312     or the language definition file \CurrentOption.ldf was not found}}%
313   Valid options are: shorthands=, KeepShorthandsActive,\\%
314   activeacute, activegrave, noconfigs, safe=, main=, math=\\%
315   headfoot=, strings=, config=, hyphenmap=, or a language name.}}}
```

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

```

316 \DeclareOption{acadian}{\bbl@load@language{frenchb}}
317 \DeclareOption{afrikaans}{\bbl@load@language{dutch}}
318 \DeclareOption{brazil}{\bbl@load@language{portuges}}
319 \DeclareOption{brazilian}{\bbl@load@language{portuges}}
320 \DeclareOption{canadien}{\bbl@load@language{frenchb}}
321 \DeclareOption{français}{\bbl@load@language{frenchb}}
322 \DeclareOption{french}{\bbl@load@language{frenchb}}%
323 \DeclareOption{hebrew}{%
324   \input{rlbabel.def}%
325   \bbl@load@language{hebrew}}
326 \DeclareOption{hungarian}{\bbl@load@language{magyar}}
327 \DeclareOption{lowersorbian}{\bbl@load@language{lsorbian}}
328 \DeclareOption{nynorsk}{\bbl@load@language{norsk}}
329 \DeclareOption{polutonikogreek}{%
330   \bbl@load@language{greek}%
331   \languageattribute{greek}{polutoniko}}
332 \DeclareOption{portuguese}{\bbl@load@language{portuges}}
333 \DeclareOption{russian}{\bbl@load@language{russianb}}
334 \DeclareOption{ukrainian}{\bbl@load@language{ukraineb}}
335 \DeclareOption{uppersorbian}{\bbl@load@language{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.

```

336 \ifx\bbl@opt@config\@nnil
337   \@ifpackagewith{babel}{noconfigs}}}%
338   {\InputIfFileExists{bblopts.cfg}%
339     {\typeout{*****^J%
340       * Local config file bblopts.cfg used^^J%
341       *}}}%
342     {}}}%
343 \else
```



```

344 \InputIfFileExists{\bbl@opt@config.cfg}%
345 {\typeout{*****^J%
346          * Local config file \bbl@opt@config.cfg used^J%
347          *}}%
348 {\bbl@error{%
349   Local config file '\bbl@opt@config.cfg' not found}{%
350   Perhaps you misspelled it.}}%
351 \fi

```

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

```

352 \bbl@for\bbl@tempa\bbl@language@opts{%
353   \bbl@ifunset{ds@\bbl@tempa}%
354   {\edef\bbl@tempb{%
355     \noexpand\DeclareOption
356     {\bbl@tempa}%
357     {\noexpand\bbl@load@language{\bbl@tempa}}}%
358   \bbl@tempb}%
359   \@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.

```

360 \bbl@foreach\@classoptionslist{%
361   \bbl@ifunset{ds@#1}%
362   {\IfFileExists{#1.ldf}%
363    {\DeclareOption{#1}{\bbl@load@language{#1}}}%
364    {}}%
365   {}}

```

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

```

366 \ifx\bbl@opt@main\@nnil\else
367   \expandafter
368   \let\expandafter\bbl@loadmain\csname ds@\bbl@opt@main\endcsname
369   \DeclareOption{\bbl@opt@main}{}
370 \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):

```

371 \def\AfterBabelLanguage#1{%
372   \bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang{}}
373   \DeclareOption*{}
374   \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`.

```

375 \ifx\bbbl@opt@main\@nnil
376 \edef\bbbl@tempa{\@classoptionslist,\bbbl@language@opts}
377 \let\bbbl@tempc\@empty
378 \bbbl@for\bbbl@tempb\bbbl@tempa{%
379   \@expandtwoargs\in@{,\bbbl@tempb,}{,\bbbl@loaded,}%
380   \ifin@edef\bbbl@tempc{\bbbl@tempb}\fi}
381 \def\bbbl@tempa#1,#2\@nnil{\def\bbbl@tempb{#1}}
382 \expandafter\bbbl@tempa\bbbl@loaded,\@nnil
383 \ifx\bbbl@tempb\bbbl@tempc\else
384   \bbbl@warning{%
385     Last declared language option is '\bbbl@tempc',\\%
386     but the last processed one was '\bbbl@tempb'.\\%
387     The main language cannot be set as both a global\\%
388     and a package option. Use 'main=\bbbl@tempc' as\\%
389     option. Reported}%
390 \fi
391 \else
392 \DeclareOption{\bbbl@opt@main}{\bbbl@loadmain}
393 \ExecuteOptions{\bbbl@opt@main}
394 \DeclareOption*{}
395 \ProcessOptions*
396 \fi
397 \def\AfterBabelLanguage{%
398   \bbbl@error
399   {Too late for \string\AfterBabelLanguage}%
400   {Languages have been loaded, so I can do nothing}}

```

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

```

401 \ifx\bbbl@main@language\@undefined
402   \bbbl@error{%
403     You haven't specified a language option}{%
404     You need to specify a language, either as a global option\\%
405     or as an optional argument to the \string\usepackage\space
406     command;\\%
407     You shouldn't try to proceed from here, type x to quit.}
408 \fi
409 \</package>

```

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 L^AT_EX-specific stuff.

Because plain T_EX users might want to use some of the features of the babel system too, care has to be taken that plain T_EX can process the files. For this reason the current format will have to be checked in a number of places. Some of

the code below is common to plain T_EX and L^AT_EX, some of it is for the L^AT_EX case only.

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

8.1 Tools

```

410 <core>
411 <<Make sure ProvidesFile is defined>>
412 \ProvidesFile{babel.def}[[<date>] <<version>> Babel common definitions]
413 <<Load macros for plain if not LaTeX>>
414 \ifx\bbbl@ifshorthand\@undefined
415   \def\bbbl@ifshorthand#1#2#3{#2}%
416   \def\bbbl@opt@safe{BR}
417   \def\AfterBabelLanguage#1#2{}
418   \let\bbbl@afterlang\relax
419   \let\bbbl@language@opts\@empty
420 \fi
421 \input switch.def\relax
422 \ifx\bbbl@languages\@undefined
423   \ifx\directlua\@undefined
424     \openin1 = language.def
425     \ifeof1
426       \closein1
427       \message{I couldn't find the file language.def}
428     \else
429       \closein1
430       \begingroup
431 \def\addlanguage#1#2#3#4#5{%
432   \expandafter\ifx\csname lang@#1\endcsname\relax\else
433     \global\expandafter\let\csname l@#1\expandafter\endcsname
434       \csname lang@#1\endcsname
435   \fi}%
436 \def\uselanguage#1{%
437 \input language.def
438   \endgroup
439 \fi
440 \fi
441 \chardef\l@english\z@
442 \fi
443 <<Load patterns in luatex>>
444 <<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.

```

445 \def\addto#1#2{%
446   \ifx#1\@undefined
447     \def#1{#2}%
448   \else
449     \ifx#1\relax
450       \def#1{#2}%
451     \else
452       {\toks@\expandafter{#1#2}%
453        \xdef#1{\the\toks@}}%
454     \fi
455   \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.

```

456 \def\bbl@withactive#1#2{%
457   \begingroup
458   \lccode'\~='#2\relax
459   \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 L^AT_EX 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@...`

```

460 \def\bbl@redefine#1{%
461   \edef\bbl@tempa{\bbl@stripslash#1}%
462   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
463   \expandafter\def\csname\bbl@tempa\endcsname}

```

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

```

464 \@onlypreamble\bbl@redefine

```

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

```

465 \def\bbl@redefine@long#1{%
466   \edef\bbl@tempa{\bbl@stripslash#1}%
467   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
468   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
469 \@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_`.

```

470 \def\bbl@redefineroobust#1{%
471   \edef\bbl@tempa{\bbl@stripslash#1}%
472   \bbl@ifunset{\bbl@tempa\space}%
473   {\expandafter\let\csname org@\bbl@tempa\endcsname#1%
474    \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
475   {\bbl@exp{\let\<org@\bbl@tempa>\<\bbl@tempa\space>}}}%
476   \@namedef{\bbl@tempa\space}}

```

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

```
477 \onlypreamble\babel@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 vety little to catch errors, but it is intended for developpers, after all. \babel@usehooks is the commands used by babel to execute hooks defined for an event.

```
478 \def\AddBabelHook#1#2{%
479   \babel@ifunset{babel@hk@#1}{\EnableBabelHook{#1}}{}%
480   \def\babel@tempa##1,#2=##2,##3\@empty{\def\babel@tempb{##2}}%
481   \expandafter\babel@tempa\babel@evargs,#2=,\@empty
482   \babel@ifunset{babel@ev@#1@#2}%
483     {\babel@csarg\babel@add{ev@#2}{\babel@elt{#1}}}%
484     \babel@csarg\newcommand}%
485     {\babel@csarg\let{ev@#1@#2}\relax
486     \babel@csarg\newcommand}%
487     {ev@#1@#2}{\babel@tempb}}
488 \def\EnableBabelHook#1{\babel@csarg\let{hk@#1}\@firstofone}
489 \def\DisableBabelHook#1{\babel@csarg\let{hk@#1}\@gobble}
490 \def\babel@usehooks#1#2{%
491   \def\babel@elt##1{%
492     \@nameuse{babel@hk@##1}{\@nameuse{babel@ev@##1@#1}#2}}%
493   \@nameuse{babel@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).

```
494 \def\babel@evargs{,% don't delete the comma
495   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
496   adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
497   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
498   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 \babel@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 \babel@e@<language> contains

\babel@ensure{<include>}{<exclude>}{<fontenc>}, which in in turn loops over the macros names in \babel@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.

```
499 \newcommand\babelensure[2][{}% TODO - revise test files
500   \AddBabelHook{babel-ensure}{afterextras}{%
501     \ifcase\babel@select@type
502       \@nameuse{babel@e@\language\name}%
```

```

503 \fi}%
504 \begingroup
505 \let\bbl@ens@include\@empty
506 \let\bbl@ens@exclude\@empty
507 \def\bbl@ens@fontenc{\relax}%
508 \def\bbl@tempb##1{%
509 \ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}%
510 \edef\bbl@tempa{\bbl@tempb##1\@empty}%
511 \def\bbl@tempb##1=##2\@@{\@namedef\bbl@ens@##1}{##2}}%
512 \bbl@foreach\bbl@tempa{\bbl@tempb##1\@@}%
513 \def\bbl@tempc{\bbl@ensure}%
514 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
515 \expandafter{\bbl@ens@include}}%
516 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
517 \expandafter{\bbl@ens@exclude}}%
518 \toks@\expandafter{\bbl@tempc}%
519 \bbl@exp{%
520 \endgroup
521 \def<bbl@e@#2>{\the\toks@{\bbl@ens@fontenc}}}%
522 \def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
523 \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
524 \ifx##1\@empty\else
525 \in{##1}{#2}%
526 \ifin\else
527 \bbl@ifunset\bbl@ensure@\language}%
528 {\bbl@exp{%
529 \\\DeclareRobustCommand<bbl@ensure@\language>[1]{%
530 \\\foreignlanguage{\language}%
531 {\ifx\relax#3\else
532 \\\fontencoding{#3}\selectfont
533 \fi
534 #####1}}}%
535 }%
536 \toks@\expandafter{##1}%
537 \edef##1{%
538 \bbl@csarg\noexpand{ensure@\language}%
539 {\the\toks@}}%
540 \fi
541 \expandafter\bbl@tempb
542 \fi}%
543 \expandafter\bbl@tempb\bbl@captionslist\today\@empty
544 \def\bbl@tempa##1{% elt for include list
545 \ifx##1\@empty\else
546 \bbl@csarg\in{ensure@\language\expandafter}\expandafter{##1}%
547 \ifin\else
548 \bbl@tempb##1\@empty
549 \fi
550 \expandafter\bbl@tempa
551 \fi}%
552 \bbl@tempa#1\@empty}
553 \def\bbl@captionslist{%
554 \prefacename\refname\abstractname\bibname\chaptername\appendixname
555 \contentsname\listfigurename\listtablename\indexname\figurename
556 \tablename\partname\enclname\ccname\headtoname\pagename\seename
557 \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`.

```
558 \def\bbl@ldfinit{%
559   \let\bbl@screset\@empty
560   \let\BabelStrings\bbl@opt@string
561   \let\BabelOptions\@empty
562   \let\BabelLanguages\relax
563   \ifx\originalTeX\@undefined
564     \let\originalTeX\@empty
565   \else
566     \originalTeX
567   \fi}
568 \def\LdfInit#1#2{%
569   \chardef\atcatcode=\catcode'\@
570   \catcode'\@=11\relax
571   \chardef\eqcatcode=\catcode'\=
572   \catcode'\==12\relax
573   \expandafter\if\expandafter\@backslashchar
574     \expandafter\@car\string#2\@nil
575   \ifx#2\@undefined\else
576     \ldf@quit{#1}%
577   \fi
578 \else
579   \expandafter\ifx\cscname#2\endcscname\relax\else
580     \ldf@quit{#1}%
581   \fi
582 \fi
583 \bbl@ldfinit}
```

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

```
584 \def\ldf@quit#1{%
585   \expandafter\main@language\expandafter{#1}%
586   \catcode'\@=\atcatcode \let\atcatcode\relax
```

```

587 \catcode'\==\eqcatcode \let\eqcatcode\relax
588 \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.

```

589 \def\bbl@afterldf#1{%
590 \bbl@afterlang
591 \let\bbl@afterlang\relax
592 \let\BabelModifiers\relax
593 \let\bbl@screset\relax}%
594 \def\ldf@finish#1{%
595 \loadlocalcfg{#1}%
596 \bbl@afterldf{#1}%
597 \expandafter\main@language\expandafter{#1}%
598 \catcode'\@=\atcatcode \let\atcatcode\relax
599 \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 .

```

600 \@onlypreamble\LdfInit
601 \@onlypreamble\ldf@quit
602 \@onlypreamble\ldf@finish

```

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

```

603 \def\main@language#1{%
604 \def\bbl@main@language{#1}%
605 \let\language\name\bbl@main@language
606 \bbl@patterns{\language}}

```

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

```

607 \AtBeginDocument{%
608 \expandafter\selectlanguage\expandafter{\bbl@main@language}}

```

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

```

609 \def\select@language@x#1{%
610 \ifcase\bbl@select@type
611 \bbl@ifsamestring\language\name{#1}{\select@language{#1}}%
612 \else
613 \select@language{#1}%
614 \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.

```

615 \def\bb@add@special#1{% 1:a macro like \", \?, etc.
616   \bb@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
617   \bb@ifunset{@sanitize}{ }\bb@add@sanitize{\@makeother#1}}%
618   \ifx\nfss@catcodes\undefined\else % TODO - same for above
619     \begingroup
620       \catcode'#1\active
621       \nfss@catcodes
622       \ifnum\catcode'#1=\active
623         \endgroup
624         \bb@add\nfss@catcodes{\@makeother#1}%
625       \else
626         \endgroup
627     \fi
628   \fi}

```

`\bb@remove@special` The companion of the former macro is `\bb@remove@special`. It removes a character from the set macros `\dospecials` and `\@sanitize`, but it is not used at all in the babel core.

```

629 \def\bb@remove@special#1{%
630   \begingroup
631   \def\x##1##2{\ifnum'#1=##2\noexpand\@empty
632     \else\noexpand##1\noexpand##2\fi}%
633   \def\do{\x\do}%
634   \def\@makeother{\x\@makeother}%
635   \edef\x{\endgroup
636     \def\noexpand\dospecials{\dospecials}%
637     \expandafter\ifx\csname @sanitize\endcsname\relax\else
638       \def\noexpand\@sanitize{\@sanitize}%
639     \fi}%
640   \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 `\bb@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 `\bb@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).

```

641 \def\bbl@active@def#1#2#3#4{%
642   \@namedef{#3#1}{%
643     \expandafter\ifx\csname#2@sh@#1\endcsname\relax
644       \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
645     \else
646       \bbl@afterfi\csname#2@sh@#1\endcsname
647     \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.

```

648   \long\@namedef{#3@arg#1}##1{%
649     \expandafter\ifx\csname#2@sh@#1\string##1\endcsname\relax
650       \bbl@afterelse\csname#4#1\endcsname##1%
651     \else
652       \bbl@afterfi\csname#2@sh@#1\string##1\endcsname
653     \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.

```

654 \def\initiate@active@char#1{%
655   \bbl@ifunset{active@char\string#1}%
656   {\bbl@withactive
657     {\expandafter\@initiate@active@char\expandafter}#1\string#1}%
658   {}}

```

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).

```

659 \def\@initiate@active@char#1#2#3{%
660   \bbl@csarg\edef{oricat@#2}{\catcode'#2=\the\catcode'#2\relax}%
661   \ifx#1\@undefined
662     \bbl@csarg\edef{oridef@#2}{\let\noexpand#1\noexpand\@undefined}%
663   \else
664     \bbl@csarg\let{oridef@#2}#1%
665     \bbl@csarg\edef{oridef@#2}{%
666       \let\noexpand#1%
667       \expandafter\noexpand\csname bbl@oridef@#2\endcsname}%
668   \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).

```

669   \ifx#1#3\relax
670     \expandafter\let\csname normal@char#2\endcsname#3%
671   \else
672     \bbl@info{Making #2 an active character}%
673     \ifnum\mathcode'#2="8000
674       \@namedef{normal@char#2}{%

```

```

675      \textormath{#3}{\csname bbl@oridef@@#2\endcsname}}%
676      \else
677      \@namedef{normal@char#2}{#3}%
678      \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).

```

679      \bbl@restoreactive{#2}%
680      \AtBeginDocument{%
681      \catcode'#2\active
682      \if@filesw
683      \immediate\write\@mainaux{\catcode'\string#2\active}%
684      \fi}%
685      \expandafter\bbl@add@special\csname#2\endcsname
686      \catcode'#2\active
687      \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⟩`).

```

688      \let\bbl@tempa\@firstoftwo
689      \if\string^#2%
690      \def\bbl@tempa{\noexpand\textormath}%
691      \else
692      \ifx\bbl@mathnormal\@undefined\else
693      \let\bbl@tempa\bbl@mathnormal
694      \fi
695      \fi
696      \expandafter\edef\csname active@char#2\endcsname{%
697      \bbl@tempa
698      {\noexpand\if@safe@actives
699      \noexpand\expandafter
700      \expandafter\noexpand\csname normal@char#2\endcsname
701      \noexpand\else
702      \noexpand\expandafter
703      \expandafter\noexpand\csname bbl@doactive#2\endcsname
704      \noexpand\fi}%
705      {\expandafter\noexpand\csname normal@char#2\endcsname}}%
706      \bbl@csarg\edef{doactive#2}{%
707      \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!).

```

708 \bbl@csarg\edef{active@#2}{%
709   \noexpand\active@prefix\noexpand#1%
710   \expandafter\noexpand\csname active@char#2\endcsname}%
711 \bbl@csarg\edef{normal@#2}{%
712   \noexpand\active@prefix\noexpand#1%
713   \expandafter\noexpand\csname normal@char#2\endcsname}%
714 \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.

```

715 \bbl@active@def#2\user@group{user@active}{language@active}%
716 \bbl@active@def#2\language@group{language@active}{system@active}%
717 \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.

```

718 \expandafter\edef\csname\user@group @sh#2@@\endcsname
719   {\expandafter\noexpand\csname normal@char#2\endcsname}%
720 \expandafter\edef\csname\user@group @sh#2@\string\protect@\endcsname
721   {\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.

```

722 \if\string'#2%
723   \let\prim@s\bbl@prim@s
724   \let\active@math@prime#1%
725 \fi
726 \bbl@usehooks{initiateactive}{{#1}{#2}{#3}}

```

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

```

727 <<(*More package options)>> ≡
728 \DeclareOption{math=active}{}
729 \DeclareOption{math=normal}{{\def\bbl@mathnormal{\noexpand\textormath}}}
730 <</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*.

```

731 \ifpackagewith{babel}{KeepShorthandsActive}%
732   {\let\bbl@restoreactive@gobble}%
733   {\def\bbl@restoreactive#1{%
734     \bbl@exp{%
735       \\AfterBabelLanguage\\CurrentOption
736       {\catcode'#1=\the\catcode'#1\relax}%
737       \\AtEndOfPackage

```

```

738      {\catcode'#1=\the\catcode'#1\relax}}}%
739      \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.

```

740 \def\bbl@sh@select#1#2{%
741   \expandafter\ifx\csname#1@sh@#2@sel\endcsname\relax
742     \bbl@afterelse\bbl@scndcs
743   \else
744     \bbl@afterfi\csname#1@sh@#2@sel\endcsname
745   \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`.

```

746 \def\active@prefix#1{%
747   \ifx\protect\@typeset@protect
748   \else

```

When `\protect` is set to `\@unexpandable@protect` we make sure that the active character is *als 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).

```

749     \ifx\protect\@unexpandable@protect
750       \noexpand#1%
751     \else
752       \protect#1%
753     \fi
754     \expandafter\@gobble
755   \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>`.

```

756 \newif\if@safe@actives
757 \@safe@activesfalse

```

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

```

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

```

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

```

759 \def\bbl@activate#1{%

```

```

760 \bbl@withactive{\expandafter\let\expandafter}#1%
761 \csname bbl@active@\string#1\endcsname}
762 \def\bbl@deactivate#1{%
763 \bbl@withactive{\expandafter\let\expandafter}#1%
764 \csname bbl@normal@\string#1\endcsname}

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

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

\declare@shorthand The command \declare@shorthand is used to declare a shorthand on a certain
level. It takes three arguments:

1. a name for the collection of shorthands, i.e. 'system', or 'dutch';
2. the character (sequence) that makes up the shorthand, i.e. ~ or "a;
3. the code to be executed when the shorthand is encountered.

767 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil}
768 \def\@decl@short#1#2#3\@nil#4{%
769 \def\bbl@tempa{#3}%
770 \ifx\bbl@tempa\@empty
771 \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs
772 \bbl@ifunset{#1@sh@\string#2@}{}%
773 {\def\bbl@tempa{#4}%
774 \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa
775 \else
776 \bbl@info
777 {Redefining #1 shorthand \string#2\\%
778 in language \CurrentOption}%
779 \fi}%
780 \@namedef{#1@sh@\string#2@}{#4}%
781 \else
782 \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs
783 \bbl@ifunset{#1@sh@\string#2@\string#3@}{}%
784 {\def\bbl@tempa{#4}%
785 \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
786 \else
787 \bbl@info
788 {Redefining #1 shorthand \string#2\string#3\\%
789 in language \CurrentOption}%
790 \fi}%
791 \@namedef{#1@sh@\string#2@\string#3@}{#4}%
792 \fi}

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

793 \def\textormath{%
794 \ifmmode
795 \expandafter\@secondoftwo
796 \else
797 \expandafter\@firstoftwo
798 \fi}

```

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

```
799 \def\user@group{user}
800 \def\language@group{english}
801 \def\system@group{system}
```

`\usesshorthands` This is the user level command to tell L^AT_EX 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.

```
802 \def\usesshorthands{%
803   \ifstar\bb@usesesh@s{\bb@usesesh@x{}}
804 \def\bb@usesesh@s#1{%
805   \bb@usesesh@x
806     {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bb@activate{#1}}}%
807   {#1}}
808 \def\bb@usesesh@x#1#2{%
809   \bb@ifshorthand{#2}%
810     {\def\user@group{user}%
811       \initiate@active@char{#2}%
812       #1%
813       \bb@activate{#2}}%
814   {\bb@error
815     {Cannot declare a shorthand turned off (\string#2)}
816     {Sorry, but you cannot use shorthands which have been\\
817       turned off in the package options}}}
```

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

```
818 \def\user@language@group{user@\language@group}
819 \def\bb@set@user@generic#1#2{%
820   \bb@ifunset{user@generic@active#1}%
821     {\bb@active@def#1\user@language@group{user@active}{user@generic@active}%
822       \bb@active@def#1\user@group{user@generic@active}{language@active}%
823       \expandafter\edef\csname#2@sh@#1@\endcsname{%
824         \expandafter\noexpand\csname normal@char#1\endcsname}%
825       \expandafter\edef\csname#2@sh@#1@\string\protect\endcsname{%
826         \expandafter\noexpand\csname user@active#1\endcsname}}%
827   \@empty}
828 \newcommand\defineshorthand[3][user]{%
829   \edef\bb@tempa{\zap@space#1 \@empty}%
830   \bb@for\bb@tempb\bb@tempa{%
831     \if*\expandafter\@car\bb@tempb\@nil
832       \edef\bb@tempb{user\expandafter\@gobble\bb@tempb}%
833     \@expandtwoargs
834     \bb@set@user@generic{\expandafter\string\@car#2\@nil}\bb@tempb
```

```

835 \fi
836 \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.

```

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

```

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

```

838 \def\aliasshorthand#1#2{%
839 \bbl@ifshorthand{#2}%
840 {\expandafter\ifx\csname active@char\string#2\endcsname\relax
841 \ifx\document\@notprerr
842 \@notshorthand{#2}%
843 \else
844 \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`.

```

845 \expandafter\let\csname active@char\string#2\expandafter\endcsname
846 \csname active@char\string#1\endcsname
847 \expandafter\let\csname normal@char\string#2\expandafter\endcsname
848 \csname normal@char\string#1\endcsname
849 \bbl@activate{#2}%
850 \fi
851 \fi}%
852 {\bbl@error
853 {Cannot declare a shorthand turned off (\string#2)}
854 {Sorry, but you cannot use shorthands which have been\\%
855 turned off in the package options}}}

```

`\@notshorthand`

```

856 \def\@notshorthand#1{%
857 \bbl@error{%
858 The character '\string #1' should be made a shorthand character;\\%
859 add the command \string\usesshorthands\string{#1\string} to
860 the preamble.\\%
861 I will ignore your instruction}%
862 {You may proceed, but expect unexpected results}}

```

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

`\shorthandoff`

```

863 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
864 \DeclareRobustCommand*\shorthandoff{%
865 \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
866 \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.

```

867 \def\bb@switch@sh#1#2{%
868   \ifx#2\@nnil\else
869     \bb@ifunset{bb@active@\string#2}%
870     {\bb@error
871       {I cannot switch ‘\string#2’ on or off--not a shorthand}%
872       {This character is not a shorthand. Maybe you made\\%
873         a typing mistake? I will ignore your instruction}}%
874     {\ifcase#1%
875       \catcode'#212\relax
876       \or
877       \catcode'#2\active
878       \or
879       \csname bb@oricat@\string#2\endcsname
880       \csname bb@oridef@\string#2\endcsname
881       \fi}%
882     \bb@afterfi\bb@switch@sh#1%
883   \fi}

```

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

```

884 \def\babelshorthand{\active@prefix\babelshorthand\bb@putsh}
885 \def\bb@putsh#1{%
886   \bb@ifunset{bb@active@\string#1}%
887   {\bb@putsh@i#1\@empty\@nnil}%
888   {\csname bb@active@\string#1\endcsname}}
889 \def\bb@putsh@i#1#2\@nnil{%
890   \csname\language @sh@\string#1@%
891     \ifx\@empty#2\else\string#2@\fi\endcsname}
892 \ifx\bb@opt@shorthands\@nnil\else
893   \let\bb@s@initiate@active@char\initiate@active@char
894   \def\initiate@active@char#1{%
895     \bb@ifshorthand{#1}{\bb@s@initiate@active@char{#1}}{}}
896   \let\bb@s@switch@sh\bb@switch@sh
897   \def\bb@switch@sh#1#2{%
898     \ifx#2\@nnil\else
899       \bb@afterfi
900       \bb@ifshorthand{#2}{\bb@s@switch@sh#1{#2}}{\bb@switch@sh#1}%
901     \fi}
902   \let\bb@s@activate\bb@activate
903   \def\bb@activate#1{%
904     \bb@ifshorthand{#1}{\bb@s@activate{#1}}{}}
905   \let\bb@s@deactivate\bb@deactivate
906   \def\bb@deactivate#1{%
907     \bb@ifshorthand{#1}{\bb@s@deactivate{#1}}{}}
908 \fi

```

`\bb@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.

```

909 \def\bb@prim@s{%
910   \prime\futurelet\@let@token\bb@pr@m@s}

```

```

911 \def\bbl@if@primes#1#2{%
912   \ifx#1\@let@token
913     \expandafter\@firstoftwo
914   \else\ifx#2\@let@token
915     \bbl@afterelse\expandafter\@firstoftwo
916   \else
917     \bbl@afterfi\expandafter\@secondoftwo
918   \fi\fi}
919 \begingroup
920   \catcode'\^=7 \catcode'\*=\active \lccode'\*='^
921   \catcode'\'=12 \catcode'\"=\active \lccode'\"='\'
922   \lowercase{%
923     \gdef\bbl@pr@ms{%
924       \bbl@if@primes" '%
925       \pr@@@s
926       {\bbl@if@primes*^ \pr@@@t\egroup}}
927 \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).

```

928 \initiate@active@char{~}
929 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
930 \bbl@activate{~}

```

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

```

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

```

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

```

933 \ifx\f@encoding\undefined
934   \def\f@encoding{OT1}
935 \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.

```

936 \newcommand\languageattribute[2]{%
937   \def\bbl@tempc{#1}%
938   \bbl@fixname\bbl@tempc
939   \bbl@iflanguage\bbl@tempc{%
940     \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.

```
941 \ifx\bbl@known@attrs\@undefined
942 \in@false
943 \else
```

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

```
944 \@expandtwoargs\in@{,\bbl@tempc-##1,}{,\bbl@known@attrs,}%
945 \fi
```

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

```
946 \ifin@
947 \bbl@warning{%
948 You have more than once selected the attribute '##1'\%
949 for language #1}%
950 \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.

```
951 \bbl@exp{%
952 \\bbl@add@list\\bbl@known@attrs{\bbl@tempc-##1}}%
953 \edef\bbl@tempa{\bbl@tempc-##1}%
954 \expandafter\bbl@ifknown@trib\expandafter{\bbl@tempa}\bbl@attributes%
955 {\csname\bbl@tempc @attr@##1\endcsname}%
956 {\@attrerr{\bbl@tempc}{##1}}%
957 \fi}}
```

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

```
958 \@onlypreamble\languageattribute
```

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

```
959 \newcommand*{\@attrerr}[2]{%
960 \bbl@error
961 {The attribute #2 is unknown for language #1.}%
962 {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}`.

```
963 \def\bbl@declare@ttribute#1#2#3{%
964 \@expandtwoargs\in@{,#2,}{,\BabelModifiers,}%
965 \ifin@
966 \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
967 \fi
968 \bbl@add@list\bbl@attributes{#1-#2}%
969 \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.

```
970 \def\bb@ifattributeset#1#2#3#4{%
```

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

```
971 \ifx\bb@known@attrs\@undefined
```

```
972 \in@false
```

```
973 \else
```

The we need to check the list of known attributes.

```
974 \@expandtwoargs\in@{,#1-#2,}{,\bb@known@attrs,}%
```

```
975 \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`'.

```
976 \ifin@
```

```
977 \bb@afterelse#3%
```

```
978 \else
```

```
979 \bb@afterfi#4%
```

```
980 \fi
```

```
981 }
```

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

```
982 \def\bb@ifknown@trib#1#2{%
```

We first assume the attribute is unknown.

```
983 \let\bb@tempa\@secondoftwo
```

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

```
984 \bb@loopx\bb@tempb{#2}{%
```

```
985 \expandafter\in@\expandafter{\expandafter,\bb@tempb,}{,#1,}%
```

```
986 \ifin@
```

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

```
987 \let\bb@tempa\@firstoftwo
```

```
988 \else
```

```
989 \fi}%
```

Finally we execute `\bb@tempa`.

```
990 \bb@tempa
```

```
991 }
```

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

```
992 \def\bb@clear@tribs{%
```

```
993 \ifx\bb@attributes\@undefined\else
```

```
994 \bb@loopx\bb@tempa{\bb@attributes}{%
```

```
995 \expandafter\bb@clear@trib\bb@tempa.
```

```
996 }%
```

```
997 \let\bb@attributes\@undefined
```

```

998 \fi}
999 \def\bbl@clear@ttrib#1-#2.{%
1000 \expandafter\let\csname#1@attr@#2\endcsname\@undefined}
1001 \AtBeginDocument{\bbl@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` `\def\babel@beginsave{\babel@savecnt\z@}`

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

```

1003 \newcount\babel@savecnt
1004 \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.

```

1005 \def\babel@save#1{%
1006 \expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax
1007 \toks@{\expandafter{\originalTeX\let#1=}%
1008 \bbl@exp{%
1009 \def\\originalTeX{\the\toks@<babel@number\babel@savecnt>\relax}}%
1010 \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.

```

1011 \def\babel@savevariable#1{%
1012 \toks@{\expandafter{\originalTeX #1=}%
1013 \bbl@exp{\def\\originalTeX{\the\toks@the#1\relax}}}

```

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

```

1014 \def\bbl@frenchspacing{%
1015 \ifnum\the\sfcodes'\.=\@m
1016 \let\bbl@nonfrenchspacing\relax
1017 \else
1018 \frenchspacing
1019 \let\bbl@nonfrenchspacing\nonfrenchspacing
1020 \fi}
1021 \let\bbl@nonfrenchspacing\nonfrenchspacing

```

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

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.

```

1022 \def\babeltags#1{%
1023   \edef\bbl@tempa{\zap@space#1 \@empty}%
1024   \def\bbl@tempb##1=##2\@{%
1025     \edef\bbl@tempc{%
1026       \noexpand\newcommand
1027       \expandafter\noexpand\csname ##1\endcsname{%
1028         \noexpand\protect
1029         \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
1030       \noexpand\newcommand
1031       \expandafter\noexpand\csname text##1\endcsname{%
1032         \noexpand\foreignlanguage{##2}}
1033     \bbl@tempc}%
1034   \bbl@for\bbl@tempa\bbl@tempa{%
1035     \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.

```

1036 \@onlypreamble\babelhyphenation
1037 \AtEndOfPackage{%
1038   \newcommand\babelhyphenation[2][\@empty]{%
1039     \ifx\bbl@hyphenation@relax
1040       \let\bbl@hyphenation@\@empty
1041     \fi
1042     \ifx\bbl@hyphlist\@empty\else
1043       \bbl@warning{%
1044         You must not intermingle \string\selectlanguage\space and\\%
1045         \string\babelhyphenation\space or some exceptions will not\\%
1046         be taken into account. Reported}%
1047     \fi
1048     \ifx\@empty#1%
1049       \protected@edef\bbl@hyphenation@{\bbl@hyphenation@\space#2}%
1050     \else
1051       \bbl@vforeach{#1}{%
1052         \def\bbl@tempa{##1}%
1053         \bbl@fixname\bbl@tempa
1054         \bbl@iflanguage\bbl@tempa{%
1055           \bbl@csarg\protected@edef{hyphenation@\bbl@tempa}{%
1056             \bbl@ifunset{\bbl@hyphenation@\bbl@tempa}%
1057             \@empty
1058             {\csname bbl@hyphenation@\bbl@tempa\endcsname\space}%
1059             #2}}}%
1060       \fi}}

```

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

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

```

1061 \def\bb@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}
1062 \def\bb@t@one{T1}
1063 \def\allowhyphens{\ifx\cf@encoding\bb@t@one\else\bb@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`.

```

1064 \newcommand\babelnullhyphen{\char\hyphenchar\font}
1065 \def\babelhyphen{\active@prefix\babelhyphen\bb@hyphen}
1066 \def\bb@hyphen{%
1067   \@ifstar{\bb@hyphen@i @}{\bb@hyphen@i \@empty}}
1068 \def\bb@hyphen@i#1#2{%
1069   \bb@iifunset{\bb@hy@#1#2\@empty}%
1070   {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%
1071   {\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 discretionaty 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.

```

1072 \def\bb@usehyphen#1{%
1073   \leavevmode
1074   \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
1075   \nobreak\hskip\z@skip}
1076 \def\bb@@usehyphen#1{%
1077   \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

The following macro inserts the hyphen char.

```

1078 \def\bb@hyphenchar{%
1079   \ifnum\hyphenchar\font=\m@ne
1080     \babelnullhyphen
1081   \else
1082     \char\hyphenchar\font
1083   \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 `\bb@hy@nobreak` is redundant.

```

1084 \def\bb@hy@soft{\bb@usehyphen{\discretionary{\bb@hyphenchar}{}}{}}
1085 \def\bb@hy@@soft{\bb@@usehyphen{\discretionary{\bb@hyphenchar}{}}{}}
1086 \def\bb@hy@hard{\bb@usehyphen\bb@hyphenchar}
1087 \def\bb@hy@@hard{\bb@@usehyphen\bb@hyphenchar}
1088 \def\bb@hy@nobreak{\bb@usehyphen{\mbox{\bb@hyphenchar}}{}}
1089 \def\bb@hy@@nobreak{\mbox{\bb@hyphenchar}}
1090 \def\bb@hy@repeat{%
1091   \bb@usehyphen{%
1092     \discretionary{\bb@hyphenchar}{\bb@hyphenchar}{\bb@hyphenchar}}}
1093 \def\bb@hy@@repeat{%
1094   \bb@@usehyphen{%

```

glue node.

```

1095 \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}
1096 \def\bbl@hy@empty{\hskip\z@skip}
1097 \def\bbl@hy@empty{\discretionary{}{}{}}

```

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

```

1098 \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.

```

1099 \def\bbl@tglobal#1{\global\let#1#1}
1100 \def\bbl@recatcode#1{%
1101   \@tempcnta="7F
1102   \def\bbl@tempa{%
1103     \ifnum\@tempcnta>"FF\else
1104       \catcode\@tempcnta=#1\relax
1105       \advance\@tempcnta@ne
1106       \expandafter\bbl@tempa
1107     \fi}%
1108   \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).

```

1109 \@ifpackagewith{babel}{nocase}%
1110 {\let\bbl@patchuclc\relax}%
1111 {\def\bbl@patchuclc{%
1112   \global\let\bbl@patchuclc\relax
1113   \@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}%
1114   \gdef\bbl@uclc##1{%
1115     \let\bbl@encoded\bbl@encoded@uclc
1116     \bbl@ifunset{\language @bbl@uclc}% and resumes it
1117     {##1}%
1118     {\let\bbl@tempa##1\relax % Used by LANG@bbl@uclc
1119      \csname\language @bbl@uclc\endcsname}%
1120     {\bbl@tolower\@empty}{\bbl@toupper\@empty}}%
1121   \gdef\bbl@tolower{\csname\language @bbl@lc\endcsname}%
1122   \gdef\bbl@toupper{\csname\language @bbl@uc\endcsname}}

```



```

1123 <<*More package options>> ≡
1124 \DeclareOption{nocase}{}
1125 <</More package options>>

```

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

```

1126 <<*More package options>> ≡
1127 \let\bbl@opt@strings\@nnil % accept strings=value
1128 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
1129 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
1130 \def\BabelStringsDefault{generic}
1131 <</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.

```

1132 \@onlypreamble\StartBabelCommands
1133 \def\StartBabelCommands{%
1134   \begingroup
1135   \bbl@reecatcode{11}%
1136   <<Macros local to BabelCommands>>
1137   \def\bbl@provstring##1##2{%
1138     \providecommand##1{##2}%
1139     \bbl@togglebal##1}%
1140   \global\let\bbl@scafter\@empty
1141   \let\StartBabelCommands\bbl@startcmds
1142   \ifx\BabelLanguages\relax
1143     \let\BabelLanguages\CurrentOption
1144   \fi
1145   \begingroup
1146   \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
1147   \StartBabelCommands}
1148 \def\bbl@startcmds{%
1149   \ifx\bbl@screset\@nnil\else
1150     \bbl@usehooks{stopcommands}{}%
1151   \fi
1152   \endgroup
1153   \begingroup
1154   \@ifstar
1155     {\ifx\bbl@opt@strings\@nnil
1156       \let\bbl@opt@strings\BabelStringsDefault
1157     \fi
1158     \bbl@startcmds@i}%
1159   \bbl@startcmds@i}
1160 \def\bbl@startcmds@i#1#2{%
1161   \edef\bbl@L{\zap@space#1 \@empty}%
1162   \edef\bbl@G{\zap@space#2 \@empty}%
1163   \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.

```

1164 \newcommand\bbbl@startcmds@ii[1][\@empty]{%
1165   \let\SetString\@gobbletwo
1166   \let\bbbl@stringdef\@gobbletwo
1167   \let\AfterBabelCommands\@gobble
1168   \ifx\@empty#1%
1169     \def\bbbl@sc@label{generic}%
1170     \def\bbbl@encstring##1##2{%
1171       \ProvideTextCommandDefault##1{##2}%
1172       \bbbl@toglobal##1%
1173       \expandafter\bbbl@toglobal\csname\string?\string##1\endcsname}%
1174     \let\bbbl@sctest\in@true
1175   \else
1176     \let\bbbl@sc@charset\space % <- zapped below
1177     \let\bbbl@sc@fontenc\space % <- " "
1178     \def\bbbl@tempa##1=##2\@nil{%
1179       \bbbl@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
1180     \bbbl@vforeach{label=#1}{\bbbl@tempa##1\@nil}%
1181     \def\bbbl@tempa##1 ##2{% space -> comma
1182       ##1%
1183       \ifx\@empty##2\else\ifx,##1,\else,\fi\bbbl@afterfi\bbbl@tempa##2\fi}%
1184     \edef\bbbl@sc@fontenc{\expandafter\bbbl@tempa\bbbl@sc@fontenc\@empty}%
1185     \edef\bbbl@sc@label{\expandafter\zap@space\bbbl@sc@label\@empty}%
1186     \edef\bbbl@sc@charset{\expandafter\zap@space\bbbl@sc@charset\@empty}%
1187     \def\bbbl@encstring##1##2{%
1188       \bbbl@foreach\bbbl@sc@fontenc{%
1189         \bbbl@ifunset{T@####1}%
1190         }%
1191         {\ProvideTextCommand##1{####1}{##2}%
1192         \bbbl@toglobal##1%
1193         \expandafter
1194         \bbbl@toglobal\csname####1\string##1\endcsname}}}%
1195     \def\bbbl@sctest{%
1196       \@expandtwoargs
1197       \in@{\bbbl@opt@strings,}{\bbbl@sc@label,\bbbl@sc@fontenc,}}%
1198   \fi
1199   \ifx\bbbl@opt@strings\@nnil % ie, no strings key -> defaults
1200   \else\ifx\bbbl@opt@strings\relax % ie, strings=encoded
1201     \let\AfterBabelCommands\bbbl@aftercmds
1202     \let\SetString\bbbl@setstring
1203     \let\bbbl@stringdef\bbbl@encstring
1204   \else % ie, strings=value
1205     \bbbl@sctest
1206   \ifin@
1207     \let\AfterBabelCommands\bbbl@aftercmds
1208     \let\SetString\bbbl@setstring
1209     \let\bbbl@stringdef\bbbl@provstring
1210   \fi\fi\fi
1211   \bbbl@scswitch
1212   \ifx\bbbl@G\@empty
1213     \def\SetString##1##2{%

```

```

1214 \bbl@error{Missing group for string \string##1}%
1215 {You must assign strings to some category, typically\\%
1216 captions or extras, but you set none}}%
1217 \fi
1218 \ifx\@empty#1%
1219 \bbl@usehooks{defaultcommands}{}%
1220 \else
1221 \@expandtwoargs
1222 \bbl@usehooks{encodedcommands}{{\bbl@sc@charset}{\bbl@sc@fontenc}}%
1223 \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) .

```

1224 \def\bbl@forlang#1#2{%
1225 \bbl@for#1\bbl@L{%
1226 \expandtwoargs\in{,#1,}{,\BabelLanguages,}%
1227 \ifin#2\relax\fi}}
1228 \def\bbl@scswitch{%
1229 \bbl@forlang\bbl@tempa{%
1230 \ifx\bbl@G\@empty\else
1231 \ifx\SetString\@gobbletwo\else
1232 \edef\bbl@GL{\bbl@G\bbl@tempa}%
1233 \expandtwoargs\in{,\bbl@GL,}{,\bbl@screset,}%
1234 \ifin\else
1235 \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
1236 \xdef\bbl@screset{\bbl@screset,\bbl@GL}%
1237 \fi
1238 \fi
1239 \fi}}
1240 \AtEndOfPackage{%
1241 \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{#2}}}%
1242 \let\bbl@scswitch\relax}
1243 \onlypreamble\EndBabelCommands
1244 \def\EndBabelCommands{%
1245 \bbl@usehooks{stopcommands}{}%
1246 \endgroup
1247 \endgroup
1248 \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.

```

1249 \def\bbl@setstring#1#2{%
1250   \bbl@forlang\bbl@tempa{%
1251     \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
1252     \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
1253     {\global\expandafter % TODO - con \bbl@exp ?
1254       \bbl@add\csname\bbl@G\bbl@tempa\expandafter\endcsname\expandafter
1255       {\expandafter\bbl@scset\expandafter#1\csname\bbl@LC\endcsname}}}%
1256     }%
1257   \def\BabelString{#2}%
1258   \bbl@usehooks{stringprocess}{}%
1259   \expandafter\bbl@stringdef
1260   \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`.

```

1261 \ifx\bbl@opt@strings\relax
1262   \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
1263   \bbl@patchuclc
1264   \let\bbl@encoded\relax
1265   \def\bbl@encoded@uclc#1{%
1266     \@inmathwarn#1%
1267     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
1268       \expandafter\ifx\csname ?\string#1\endcsname\relax
1269         \TextSymbolUnavailable#1%
1270       \else
1271         \csname ?\string#1\endcsname
1272       \fi
1273     \else
1274       \csname\cf@encoding\string#1\endcsname
1275     \fi}
1276 \else
1277   \def\bbl@scset#1#2{\def#1{#2}}
1278 \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.

```

1279 <<*Macros local to BabelCommands>> ≡
1280 \def\SetStringLoop##1##2{%
1281   \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
1282   \count@z@
1283   \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
1284     \advance\count@\@ne
1285     \toks@\expandafter{\bbl@tempa}%
1286     \bbl@exp{%
1287       \\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
1288       \count@=\the\count@\relax}}}%
1289 <</Macros local to BabelCommands>>

```

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

```

1290 \def\bbl@aftercmds#1{%
1291   \toks@\expandafter{\bbl@scafter#1}%
1292   \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.

```

1293 <<*Macros local to BabelCommands>> ≡
1294 \newcommand\SetCase[3][]{%
1295   \bbl@patchuclc
1296   \bbl@forlang\bbl@tempa{%
1297     \expandafter\bbl@encstring
1298     \csname\bbl@tempa @bbl@uclc\endcsname{\bbl@tempa##1}%
1299     \expandafter\bbl@encstring
1300     \csname\bbl@tempa @bbl@uc\endcsname{##2}%
1301     \expandafter\bbl@encstring
1302     \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
1303 <</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.

```

1304 <<*Macros local to BabelCommands>> ≡
1305 \newcommand\SetHyphenMap[1]{%
1306   \bbl@forlang\bbl@tempa{%
1307     \expandafter\bbl@stringdef
1308     \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}%
1309 <</Macros local to BabelCommands>>

```

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

```

1310 \newcommand\BabelLower[2]{% one to one.
1311   \ifnum\lccode#1=#2\else
1312     \babel@savevariable{\lccode#1}%
1313     \lccode#1=#2\relax
1314   \fi}
1315 \newcommand\BabelLowerMM[4]{% many-to-many
1316   \@tempcnta=#1\relax
1317   \@tempcntb=#4\relax
1318   \def\bbl@tempa{%
1319     \ifnum\@tempcnta>#2\else
1320       \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
1321       \advance\@tempcnta#3\relax
1322       \advance\@tempcntb#3\relax
1323       \expandafter\bbl@tempa
1324     \fi}%
1325   \bbl@tempa}
1326 \newcommand\BabelLowerM0[4]{% many-to-one
1327   \@tempcnta=#1\relax
1328   \def\bbl@tempa{%
1329     \ifnum\@tempcnta>#2\else
1330       \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
1331       \advance\@tempcnta#3
1332       \expandafter\bbl@tempa
1333     \fi}%
1334   \bbl@tempa}

```

The following package options control the behaviour of hyphenation mapping.

```

1335 <<*More package options>> ≡
1336 \DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
1337 \DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\@ne}

```

```

1338 \DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
1339 \DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@@}
1340 \DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}
1341 <</More package options>>

```

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

```

1342 \AtEndOfPackage{%
1343   \ifx\bbl@opt@hyphenmap\undefined
1344     \@expandtwoargs\in@{,}{\bbl@language@opts}%
1345     \chardef\bbl@opt@hyphenmap\ifin@4\else\@ne\fi
1346   \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.

```

1347 \def\set@low@box#1{\setbox\tw@\hbox{,}\setbox\z@\hbox{#1}%
1348   \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@%
1349   \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.

```

1350 \def\save@sf@q#1{\leavevmode
1351   \begingroup
1352   \edef\@SF{\spacefactor\the\spacefactor}#1\@SF
1353   \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 `Tlenc.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.

```

1354 \ProvideTextCommand{\quotedblbase}{OT1}{%
1355   \save@sf@q{\set@low@box{\textquotedblright\}%
1356     \box\z@\kern-.04em\bbl@allowhyphens}}

```

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

```

1357 \ProvideTextCommandDefault{\quotedblbase}{%
1358   \UseTextSymbol{OT1}{\quotedblbase}}

```

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

```

1359 \ProvideTextCommand{\quotesinglbase}{OT1}{%
1360   \save@sf@q{\set@low@box{\textquoteright\}%
1361     \box\z@\kern-.04em\bbl@allowhyphens}}

```

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

```

1362 \ProvideTextCommandDefault{\quotesinglbase}{%
1363   \UseTextSymbol{OT1}{\quotesinglbase}}

```

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

```
\guillemotright 1364 \ProvideTextCommand{\guillemotleft}{OT1}{%
1365   \ifmmode
1366     \ll
1367   \else
1368     \save@sf@q{\nobreak
1369       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bbl@allowhyphens}%
1370   \fi}
1371 \ProvideTextCommand{\guillemotright}{OT1}{%
1372   \ifmmode
1373     \gg
1374   \else
1375     \save@sf@q{\nobreak
1376       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bbl@allowhyphens}%
1377   \fi}
```

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

```
1378 \ProvideTextCommandDefault{\guillemotleft}{%
1379   \UseTextSymbol{OT1}{\guillemotleft}}
1380 \ProvideTextCommandDefault{\guillemotright}{%
1381   \UseTextSymbol{OT1}{\guillemotright}}
```

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

```
\guilsinglright 1382 \ProvideTextCommand{\guilsinglleft}{OT1}{%
1383   \ifmmode
1384     <%
1385   \else
1386     \save@sf@q{\nobreak
1387       \raise.2ex\hbox{$\scriptscriptstyle<$}\bbl@allowhyphens}%
1388   \fi}
1389 \ProvideTextCommand{\guilsinglright}{OT1}{%
1390   \ifmmode
1391     >%
1392   \else
1393     \save@sf@q{\nobreak
1394       \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
1395   \fi}
```

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

```
1396 \ProvideTextCommandDefault{\guilsinglleft}{%
1397   \UseTextSymbol{OT1}{\guilsinglleft}}
1398 \ProvideTextCommandDefault{\guilsinglright}{%
1399   \UseTextSymbol{OT1}{\guilsinglright}}
```

8.11.2 Letters

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

```
1400 \DeclareTextCommand{\ij}{OT1}{%
1401   i\kern-0.02em\bbl@allowhyphens j}
1402 \DeclareTextCommand{\IJ}{OT1}{%
1403   I\kern-0.02em\bbl@allowhyphens J}
1404 \DeclareTextCommand{\ij}{T1}{\char188}
1405 \DeclareTextCommand{\IJ}{T1}{\char156}
```

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

```
1406 \ProvideTextCommandDefault{\ij}{%
1407   \UseTextSymbol{OT1}{\ij}}
1408 \ProvideTextCommandDefault{\IJ}{%
1409   \UseTextSymbol{OT1}{\IJ}}
```

\dj The croatian language needs the letters \dj and \DJ; they are available in the T1 encoding, but not in the OT1 encoding by default.

Some code to construct these glyphs for the OT1 encoding was made available to me by Stipcevic Mario, (stipcevic@olimp.irb.hr).

```
1410 \def\crrtic@{\hrule height0.1ex width0.3em}
1411 \def\crttic@{\hrule height0.1ex width0.33em}
1412 \def\ddj@{%
1413   \setbox0\hbox{d}\dimen@=\ht0
1414   \advance\dimen@lex
1415   \dimen@.45\dimen@
1416   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1417   \advance\dimen@ii.5ex
1418   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
1419 \def\DDJ@{%
1420   \setbox0\hbox{D}\dimen@=.55\ht0
1421   \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
1422   \advance\dimen@ii.15ex % correction for the dash position
1423   \advance\dimen@ii-.15\fontdimen7\font % correction for cmTT font
1424   \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
1425   \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
1426 %
1427 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
1428 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}
```

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

```
1429 \ProvideTextCommandDefault{\dj}{%
1430   \UseTextSymbol{OT1}{\dj}}
1431 \ProvideTextCommandDefault{\DJ}{%
1432   \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.

```
1433 \DeclareTextCommand{\SS}{OT1}{SS}
1434 \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.

\glq The ‘german’ single quotes.

```
\grq 1435 \ProvideTextCommand{\glq}{OT1}{%
1436   \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}
1437 \ProvideTextCommand{\glq}{T1}{%
1438   \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}
1439 \ProvideTextCommandDefault{\glq}{\UseTextSymbol{OT1}{\glq}}
```


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

```

1440 \ProvideTextCommand{\grq}{T1}{%
1441   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
1442 \ProvideTextCommand{\grq}{OT1}{%
1443   \save@sf@q{\kern-.0125em%
1444   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
1445   \kern.07em\relax}}
1446 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}

```

`\glqq` The ‘german’ double quotes.

```

\grqq 1447 \ProvideTextCommand{\glqq}{OT1}{%
1448   \textormath{\quotedblbase}{\mbox{\quotedblbase}}}
1449 \ProvideTextCommand{\glqq}{T1}{%
1450   \textormath{\quotedblbase}{\mbox{\quotedblbase}}}
1451 \ProvideTextCommandDefault{\glqq}{\UseTextSymbol{OT1}\glqq}

```

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

```

1452 \ProvideTextCommand{\grqq}{T1}{%
1453   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
1454 \ProvideTextCommand{\grqq}{OT1}{%
1455   \save@sf@q{\kern-.07em%
1456   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
1457   \kern.07em\relax}}
1458 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}

```

`\flq` The ‘french’ single guillemets.

```

\frq 1459 \ProvideTextCommand{\flq}{OT1}{%
1460   \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
1461 \ProvideTextCommand{\flq}{T1}{%
1462   \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
1463 \ProvideTextCommandDefault{\flq}{\UseTextSymbol{OT1}\flq}

1464 \ProvideTextCommand{\frq}{OT1}{%
1465   \textormath{\guilsinglright}{\mbox{\guilsinglright}}}
1466 \ProvideTextCommand{\frq}{T1}{%
1467   \textormath{\guilsinglright}{\mbox{\guilsinglright}}}
1468 \ProvideTextCommandDefault{\frq}{\UseTextSymbol{OT1}\frq}

```

`\flqq` The ‘french’ double guillemets.

```

\frqq 1469 \ProvideTextCommand{\flqq}{OT1}{%
1470   \textormath{\guillemotleft}{\mbox{\guillemotleft}}}
1471 \ProvideTextCommand{\flqq}{T1}{%
1472   \textormath{\guillemotleft}{\mbox{\guillemotleft}}}
1473 \ProvideTextCommandDefault{\flqq}{\UseTextSymbol{OT1}\flqq}

1474 \ProvideTextCommand{\frqq}{OT1}{%
1475   \textormath{\guillemotright}{\mbox{\guillemotright}}}
1476 \ProvideTextCommand{\frqq}{T1}{%
1477   \textormath{\guillemotright}{\mbox{\guillemotright}}}
1478 \ProvideTextCommandDefault{\frqq}{\UseTextSymbol{OT1}\frqq}

```

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).

```
1479 \def\umlauthigh{%
1480   \def\bbl@umlauta##1{\leavevmode\bgroup%
1481     \expandafter\accent\csname f@encoding dqpos\endcsname
1482     ##1\bbl@allowhyphens\egroup}%
1483   \let\bbl@umlaute\bbl@umlauta}
1484 \def\umlautlow{%
1485   \def\bbl@umlauta{\protect\lower@umlaut}}
1486 \def\umlautelow{%
1487   \def\bbl@umlaute{\protect\lower@umlaut}}
1488 \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.

```
1489 \expandafter\ifx\csname U@D\endcsname\relax
1490   \csname newdimen\endcsname U@D
1491 \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.

```
1492 \def\lower@umlaut#1{%
1493   \leavevmode\bgroup
1494   \U@D lex%
1495   {\setbox\z@\hbox{%
1496     \expandafter\char\csname f@encoding dqpos\endcsname}%
1497     \dimen@ -.45ex\advance\dimen@\ht\z@
1498     \ifdim lex<\dimen@ \fontdimen5\font\dimen@ \fi}%
1499   \expandafter\accent\csname f@encoding dqpos\endcsname
1500   \fontdimen5\font\U@D #1%
1501   \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).

```

1502 \AtBeginDocument{%
1503   \DeclareTextCompositeCommand{"}{OT1}{a}{\bbl@umlauta{a}}%
1504   \DeclareTextCompositeCommand{"}{OT1}{e}{\bbl@umlaute{e}}%
1505   \DeclareTextCompositeCommand{"}{OT1}{i}{\bbl@umlaute{i}}%
1506   \DeclareTextCompositeCommand{"}{OT1}{\i}{\bbl@umlaute{\i}}%
1507   \DeclareTextCompositeCommand{"}{OT1}{o}{\bbl@umlauta{o}}%
1508   \DeclareTextCompositeCommand{"}{OT1}{u}{\bbl@umlauta{u}}%
1509   \DeclareTextCompositeCommand{"}{OT1}{A}{\bbl@umlauta{A}}%
1510   \DeclareTextCompositeCommand{"}{OT1}{E}{\bbl@umlaute{E}}%
1511   \DeclareTextCompositeCommand{"}{OT1}{I}{\bbl@umlaute{I}}%
1512   \DeclareTextCompositeCommand{"}{OT1}{O}{\bbl@umlauta{O}}%
1513   \DeclareTextCompositeCommand{"}{OT1}{U}{\bbl@umlauta{U}}%
1514 }

```

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

```

1515 \ifx\l@english\undefined
1516   \chardef\l@english\z@
1517 \fi
1518 \main@language{english}

```

Now we load definition files for engines.

```

1519 \ifcase\bbl@engine\or
1520   \input luababel.def
1521 \or
1522   \input xebabel.def
1523 \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`.

```

1524 {\def\format{plain}
1525 \ifx\fmtname\format
1526 \else
1527   \def\format{LaTeX2e}
1528   \ifx\fmtname\format
1529   \else
1530     \aftergroup\endinput
1531   \fi
1532 \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.

```

1533 \newcommand\babelprovide[2][]{%
1534   \let\bbl@savelangname\languagename
1535   \def\languagename{#2}%
1536   \let\bbl@KVP@captions\@nil
1537   \let\bbl@KVP@main\@nil
1538   \let\bbl@KVP@hyphenrules\@nil
1539   \bbl@forkv{#1}{\bbl@csarg\def{KVP@##1}{##2}}%  TODO - error handling
1540   \bbl@ifunset{date#2}%
1541     {\bbl@provide@new{#2}}%
1542     {\bbl@ifblank{#1}%
1543       {\bbl@error
1544         {If you want to modify ‘#2’ you must tell how in\\%
1545         the optional argument. Currently there are two\\%
1546         options: captions=lang-tag, hyphenrules=lang-list}%
1547         {Use this macro as documented}}}%
1548       {\bbl@provide@renew{#2}}}%
1549   \babelensure{#2}%
1550   \let\languagename\bbl@savelangname}

```

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

```

1551 \def\bbl@provide@new#1{%
1552   \bbl@provide@hyphens{#1}%
1553   \@namedef{date#1}{}% marks lang exists - required by \StartBabelCommands
1554   \@namedef{extras#1}{}%
1555   \@namedef{noextras#1}{}%
1556   \StartBabelCommands*{#1}{captions}%
1557   \ifx\bbl@KVP@captions\@nil
1558     \def\bbl@tempb##1{%                               elt for \bbl@captionslist
1559       \ifx##1\@empty\else
1560         \bbl@exp{%
1561           \\SetString\\##1{%
1562             \\bbl@nocaption{\bbl@stripslash##1}{\<#1\bbl@stripslash##1>}}}%
1563           \expandafter\bbl@tempb
1564         \fi}%
1565     \expandafter\bbl@tempb\bbl@captionslist\@empty
1566   \else
1567     \bbl@read@ini{\bbl@KVP@captions}% Here all letters cat = 11
1568     \bbl@after@ini
1569     \bbl@savestrings
1570   \fi
1571   \StartBabelCommands*{#1}{date}%
1572   \bbl@exp{%
1573     \\SetString\\today{\\bbl@nocaption{today}{\<#1today>}}}%
1574   \EndBabelCommands
1575   \expandafter\gdef\csname#1hyphenmins\endcsname{23}%
1576   \ifx\bbl@KVP@main\@nil\else
1577     \expandafter\main@language\expandafter{#1}%
1578   \fi}
1579 \def\bbl@provide@renew#1{%
1580   \bbl@provide@hyphens{#1}%
1581   \ifx\bbl@KVP@captions\@nil\else
1582     \StartBabelCommands*{#1}{captions}%
1583     \bbl@read@ini{\bbl@KVP@captions}% Here all letters cat = 11

```

```

1584     \bbl@after@ini
1585     \bbl@savestrings
1586     \EndBabelCommands
1587     \fi}

```

The hyphenrules option is handled with an auxiliary macro.

```

1588 \def\bbl@provide@hyphens#1{%
1589     \let\bbl@tempa\relax
1590     \ifx\bbl@KVP@hyphenrules\@nil\else
1591         \bbl@replace\bbl@KVP@hyphenrules{ }{,}%
1592         \bbl@foreach\bbl@KVP@hyphenrules{%
1593             \ifx\bbl@tempa\relax % if not yet found
1594                 \bbl@ifsamestring{##1}{+}%
1595                 {\bbl@exp{\\addlanguage\<l@##1>}}}%
1596                 {}%
1597             \bbl@ifunset{l@##1}%
1598             {}%
1599             {\bbl@exp{\let\bbl@tempa\<l@##1>}}%
1600         \fi}%
1601     \fi
1602     \ifx\bbl@tempa\relax %           if no option or no language found
1603         \bbl@ifunset{l@#1}%         no hyphenrules found - fallback
1604         {\bbl@exp{\\adddialect\<l@#1>\language}}%
1605         {}%
1606     \else
1607         \bbl@exp{\\adddialect\<l@#1>\bbl@tempa}% found in opt list
1608     \fi}

```

The reader of ini files. There are 3 possible cases: a section name (in the form [...]), a comment (starting with ;) and a key/value pair. *TODO - Work in progress.*

```

1609 \def\bbl@read@ini#1{%
1610     \openin1=babel-#1.ini
1611     \ifeof1
1612         \bbl@error
1613         {There is no ini file for the requested language\\
1614          (#1). Perhaps you misspelled it or your installation\\
1615           is not complete.}%
1616         {Fix the name or reinstall babel.}%
1617     \else
1618         \let\bbl@section\@empty
1619         \let\bbl@savestrings\@empty
1620         \loop
1621             \endlinechar\m@ne
1622             \read1 to \bbl@line
1623             \endlinechar'\^^M
1624             \if T\ifeof1F\fi T\relax % Trick, because inside \loop
1625             \ifx\bbl@line\@empty\else
1626                 \expandafter\bbl@iniline\bbl@line\bbl@iniline
1627             \fi
1628         \repeat
1629     \fi}
1630 \def\bbl@iniline#1\bbl@iniline{%
1631     \@ifnextchar[\bbl@inisec{\@ifnextchar\bbl@iniskip\bbl@inikv}#1\@@} %]
1632 \def\bbl@iniskip#1\@@{%           if starts with ;
1633 \def\bbl@inisec[#1]#2\@@{%       if starts with opening bracket
1634     \def\bbl@section{#1}%

```

```

1635 \bbl@debug{\message{[[ #1 ]]^J}}
1636 \def\bbl@inikv#1=#2\@{% otherwise, key=value
1637 \bbl@trim@def\bbl@tempa{#1}%
1638 \bbl@trim\toks@{#2}%
1639 \bbl@ifunset\bbl@ini@\bbl@section}%
1640 {}%
1641 {\bbl@exp{%
1642 \<bbl@ini@\bbl@section>\bbl@tempa=\the\toks@\nil}}}
1643 \def\bbl@after@ini{%
1644 % make sure Script and Language takes some value
1645 \bbl@exp{\bbl@ifblank{\@nameuse\bbl@lotf@language}}}%
1646 {\bbl@csarg\gdef\lotf@language{\dflt}}}%
1647 \bbl@exp{\bbl@ifblank{\@nameuse\bbl@sotf@language}}}%
1648 {\bbl@csarg\gdef\sotf@language{DFLT}}}%

```

Key-value pairs are treated differently depending on the section in the ini file. The following macros are the readers for captions (with Unicode) or captions.licr (for 8-bit) and for identification. But first, an auxiliary macro.

```

1649 \def\bbl@exportkey#1#2#3#4{% 1:bbl id,2:ini name,3:ini key,4:ini val
1650 \bbl@ifsamestring{#2}{#3}%
1651 {\bbl@csarg\gdef{#1@\language}{#4}}%
1652 {}%
1653 \ifcase\bbl@engine
1654 \bbl@csarg\def{ini@captions.licr}#1=#2\@nil{% TODO - cypaste pattern
1655 \bbl@ifblank{#2}%
1656 {\bbl@exp{%
1657 \toks@{\bbl@nocaption{#1}\<\language#1name>}}}%
1658 {\toks@{#2}}}%
1659 \bbl@exp{%
1660 \bbl@add\bbl@savestrings{% NOTE - with date will be global
1661 \SetString\<#1name>\the\toks@}}}
1662 \else
1663 \def\bbl@ini@captions#1=#2\@nil{%
1664 \bbl@ifblank{#2}%
1665 {\bbl@exp{%
1666 \toks@{\bbl@nocaption{#1}\<\language#1name>}}}%
1667 {\toks@{#2}}}%
1668 \bbl@exp{%
1669 \bbl@add\bbl@savestrings{%
1670 \SetString\<#1name>\the\toks@}}}
1671 \fi
1672 \def\bbl@ini@identification#1=#2\@nil{% TODO - not only with captions
1673 \bbl@exportkey{lname}{name.english}{#1}{#2}%
1674 \bbl@exportkey{lbcpr}{tag.bcp47}{#1}{#2}%
1675 \bbl@exportkey{lotf}{tag.opentype}{#1}{#2}%
1676 \bbl@exportkey{sname}{script.name}{#1}{#2}%
1677 \bbl@exportkey{sbcpr}{script.tag.bcp47}{#1}{#2}%
1678 \bbl@exportkey{sotf}{script.tag.opentype}{#1}{#2}%

```

Dates will require some macros for the basic formatting. They may be redefined by language, so “semi-public” names (camel case) are used.

```

1679 \newcommand\BabelDateSpace{\nobreakspace}
1680 \newcommand\BabelDateDot{.\@}
1681 \newcommand\BabelDated[1]{\number#1}
1682 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
1683 \newcommand\BabelDateM[1]{\number#1}

```

```

1684 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}}
1685 \newcommand\BabelDateMMM[1]{\%
1686 \csname month\romannumeral\month name\endcsname}}%
1687 \newcommand\BabelDatey[1]{\number#1}}%
1688 \newcommand\BabelDateyy[1]{\%
1689 \ifnum#1<10 0\number#1 %
1690 \else\ifnum#1<100 \number#1 %
1691 \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
1692 \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
1693 \else\ERRORd % TODO - real error
1694 \fi\fi\fi\fi}}
1695 \newcommand\BabelDateyyyy[1]{\number#1}}
1696 \def\bbl@TG@date{%
1697 \bbl@replace\bbl@toreplace{[ ]}{\BabelDateSpace{}}}%
1698 \bbl@replace\bbl@toreplace{[. ]}{\BabelDateDot{}}}%
1699 \bbl@replace\bbl@toreplace{[d]}{\BabelDated{###3}}}%
1700 \bbl@replace\bbl@toreplace{[dd]}{\BabelDatedd{###3}}}%
1701 \bbl@replace\bbl@toreplace{[M]}{\BabelDateM{###2}}}%
1702 \bbl@replace\bbl@toreplace{[MM]}{\BabelDateMM{###2}}}%
1703 \bbl@replace\bbl@toreplace{[MMM]}{\BabelDateMMM{###2}}}%
1704 \bbl@replace\bbl@toreplace{[y]}{\BabelDatey{###1}}}%
1705 \bbl@replace\bbl@toreplace{[yy]}{\BabelDateyy{###1}}}%
1706 \bbl@replace\bbl@toreplace{[yyyy]}{\BabelDateyyyy{###1}}}%

```

9.3 Cross referencing macros

The L^AT_EX book states:

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

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

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

The only way to accomplish this in most cases is to use the trick described in the T_EXbook [?] (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.

```

1707 %\bbl@redefine\newlabel#1#2{%
1708 % \@safe@activestrue\org@newlabel{#1}{#2}\@safe@activesfalse}

```

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

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

```

1709 <<{*More package options}>> ≡
1710 \DeclareOption{safe=none}{\let\bbl@opt@safe\empty}
1711 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}

```

```

1712 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
1713 <</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.

```

1714 \ifx\bbl@opt@safe\empty\else
1715   \def\newl@bel#1#2#3{%
1716     {\@safe@activestrue
1717       \bbl@ifunset{#1@#2}%
1718       \relax
1719       {\gdef\@multiplelabels{%
1720         \@latex@warning@no@line{There were multiply-defined labels}}%
1721         \@latex@warning@no@line{Label ‘#2’ multiply defined}}%
1722       \global\@namedef{#1@#2}{#3}}}%

```

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

```

1723 \CheckCommand*\@testdef[3]{%
1724   \def\reserved@a{#3}%
1725   \expandafter\ifx\csname#1@#2\endcsname\reserved@a
1726   \else
1727     \@tempswatrue
1728   \fi}

```

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

```

1729 \def\@testdef#1#2#3{%
1730   \@safe@activestrue

```

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

```

1731   \expandafter\let\expandafter\bbl@tempa\csname #1@#2\endcsname

```

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

```

1732   \def\bbl@tempb{#3}%
1733   \@safe@activesfalse

```

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

```

1734   \ifx\bbl@tempa\relax
1735   \else
1736     \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
1737   \fi

```

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

```

1738   \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%

```

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

```

1739   \ifx\bbl@tempa\bbl@tempb
1740   \else
1741     \@tempswatrue
1742   \fi}
1743 \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.

```

1744 \@expandtwoargs\in@{R}\bbl@opt@safe
1745 \ifin@
1746   \bbl@redefineroobust\ref#1{%
1747     \@safe@activetrue\org@ref{#1}\@safe@activesfalse}
1748   \bbl@redefineroobust\pageref#1{%
1749     \@safe@activetrue\org@pageref{#1}\@safe@activesfalse}
1750 \else
1751   \let\org@ref\ref
1752   \let\org@pageref\pageref
1753 \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.

```

1754 \@expandtwoargs\in@{B}\bbl@opt@safe
1755 \ifin@
1756   \bbl@redefine\@citex[#1]#2{%
1757     \@safe@activetrue\edef\@tempa{#2}\@safe@activesfalse
1758     \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.

```

1759 \AtBeginDocument{%
1760   \@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.)

```

1761   \def\@citex[#1][#2]#3{%
1762     \@safe@activetrue\edef\@tempa{#3}\@safe@activesfalse
1763     \org@citex[#1][#2]{\@tempa}}%
1764   }{}}

```

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

```

1765 \AtBeginDocument{%
1766   \@ifpackageloaded{cite}{%
1767     \def\@citex[#1]#2{%
1768       \@safe@activetrue\org@citex[#1][#2]\@safe@activesfalse}%
1769     }{}}

```

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

```

1770 \bbl@redefine\nocite#1{%
1771   \@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@activestrue` is in effect. This switch needs to be reset inside the `\hbox` which contains the citation label. In order to determine during `.aux` file processing which definition of `\bibcite` is needed we define `\bibcite` in such a way that it redefines itself with the proper definition.

```
1772 \bbl@redefine\bibcite{%
```

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

```
1773 \bbl@cite@choice
```

```
1774 \bibcite}
```

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

```
1775 \def\bbl@bibcite#1#2{%
```

```
1776 \org@bibcite{#1}{\@safe@activesfalse#2}}
```

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

```
1777 \def\bbl@cite@choice{%
```

First we give `\bibcite` its default definition.

```
1778 \global\let\bibcite\bbl@bibcite
```

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

```
1779 \@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{}%
```

For `cite` we do the same.

```
1780 \@ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{}%
```

Make sure this only happens once.

```
1781 \global\let\bbl@cite@choice\relax}
```

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

```
1782 \AtBeginDocument{\bbl@cite@choice}
```

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

```
1783 \bbl@redefine\@bibitem#1{%
```

```
1784 \@safe@activestrue\org@@bibitem{#1}\@safe@activesfalse}
```

```
1785 \else
```

```
1786 \let\org@nocite\nocite
```

```
1787 \let\org@@citex\@citex
```

```
1788 \let\org@bibcite\bibcite
```

```
1789 \let\org@@bibitem\@bibitem
```

```
1790 \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@activestrue` is in effect.

```
1791 \bbl@redefine\markright#1{%
1792   \bbl@ifblank{#1}%
1793   {\org@markright{}}%
1794   {\toks@{#1}%
1795    \bbl@exp{%
1796     \\org@markright{\\protect\\foreignlanguage{\language}\language}%
1797     {\\protect\\bbl@restore@actives\the\toks@}}}}
```

`\markboth` The definition of `\markboth` is equivalent to that of `\markright`, except that we
`\@mkboth` 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`.

```
1798 \ifx\@mkboth\markboth
1799   \def\bbl@tempc{\let\@mkboth\markboth}
1800 \else
1801   \def\bbl@tempc{}
1802 \fi
```

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

```
1803 \bbl@redefine\markboth#1#2{%
1804   \protected@edef\bbl@tempb##1{%
1805     \protect\foreignlanguage{\language}\protect\bbl@restore@actives##1}%
1806   \bbl@ifblank{#1}%
1807   {\toks@{}}%
1808   {\toks@\expandafter{\bbl@tempb{#1}}}%
1809   \bbl@ifblank{#2}%
1810   {\@temptokena{}}%
1811   {\@temptokena\expandafter{\bbl@tempb{#2}}}%
1812   \bbl@exp{\\org@markboth{\the\toks@}{\the\@temptokena}}}
```

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

```
1813 \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.

```
1814 \@expandtwoargs\in@{R}\bbl@opt@safe
1815 \ifin@
1816 \AtBeginDocument{%
1817   \@ifpackageloaded{ifthen}{%
```

Then we can redefine `\ifthenelse`:

```
1818   \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.

```
1819   \let\bbl@temp@pref\pageref
1820   \let\pageref\org@pageref
1821   \let\bbl@temp@ref\ref
1822   \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.

```
1823   \@safe@activetrue
1824   \org@ifthenelse{#1}%
1825   {\let\pageref\bbl@temp@pref
1826    \let\ref\bbl@temp@ref
1827    \@safe@activesfalse
1828    #2}%
1829   {\let\pageref\bbl@temp@pref
1830    \let\ref\bbl@temp@ref
1831    \@safe@activesfalse
1832    #3}%
1833   }%
1834   }{}%
1835 }
```

9.5.2 varioref

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

```
1836 \AtBeginDocument{%
1837   \@ifpackageloaded{varioref}{%
1838     \bbl@redefine\@vpageref#1[#2]#3{%
1839       \@safe@activetrue
1840       \org@@vpageref{#1}[#2]#3}%
1841       \@safe@activesfalse}%
```

The same needs to happen for `\vrefpagenum`.

```
1842      \bbl@redefine\vrefpagenum#1#2{%
1843      \@safe@activetrue
1844      \org@vrefpagenum{#1}{#2}%
1845      \@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.

```
1846      \expandafter\def\csname Ref \endcsname#1{%
1847      \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
1848      }{}%
1849  }
1850 \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.

```
1851 \AtEndOfPackage{%
1852   \AtBeginDocument{%
1853     \@ifpackageloaded{hhline}%
```

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

```
1854     {\expandafter\ifx\csname normal@char\string\endcsname\relax
1855     \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.

```
1856       \makeatletter
1857       \def\@currname{hhline}\input{hhline.sty}\makeatother
1858       \fi}%
1859   {}}}
```

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.

```
1860 \AtBeginDocument{%
1861   \ifx\pdfstringdefDisableCommands\undefined\else
1862     \pdfstringdefDisableCommands{\languageshorthands{system}}%
1863   \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.

```
1864 \DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
1865   \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.

```
1866 \def\substitutefontfamily#1#2#3{%
1867   \lowercase{\immediate\openout15=#1#2.fd\relax}%
1868   \immediate\write15{%
1869     \string\ProvidesFile{#1#2.fd}%
1870     [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
1871     \space generated font description file]^J
1872     \string\DeclareFontFamily{#1}{#2}{}}^J
1873     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{}}^J
1874     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{}}^J
1875     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{}}^J
1876     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{}}^J
1877     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{}}^J
1878     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{}}^J
1879     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{}}^J
1880     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{}}^J
1881   }%
1882   \closeout15
1883 }
```

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

```
1884 \@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\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`

```
1885 \newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU,}
1886 \let\org@TeX\TeX
1887 \let\org@LaTeX\LaTeX
1888 \let\ensureascii\@firstofone
1889 \AtBeginDocument{%
1890   \in@false
1891   \bbl@foreach\BabelNonASCII{% is there a non-ascii enc?
1892     \ifin@else
```

```

1893     \lowercase{\@expandtwoargs\in@{,#1enc.def,},{,\@filelist,}}%
1894     \fi}%
1895 \ifin@ % if a non-ascii has been loaded
1896     \def\ensureascii#1{{\fontencoding{OT1}\selectfont#1}}%
1897     \DeclareTextCommandDefault{\TeX}{\org@TeX}%
1898     \DeclareTextCommandDefault{\LaTeX}{\org@LaTeX}%
1899     \def\bbl@tempb#1\@@{\uppercase{\bbl@tempc#1}ENC.DEF\@empty\@@}%
1900     \def\bbl@tempc#1ENC.DEF#2\@@{\%
1901         \ifx\@empty#2\else
1902             \bbl@ifunset{T@#1}%
1903             {}%
1904             {\@expandtwoargs\in@{,#1,},{,\BabelNonASCII,}%
1905             \ifin@
1906                 \DeclareTextCommand{\TeX}{#1}{\ensureascii{\org@TeX}}%
1907                 \DeclareTextCommand{\LaTeX}{#1}{\ensureascii{\org@LaTeX}}%
1908             \else
1909                 \def\ensureascii##1{{\fontencoding{#1}\selectfont##1}}%
1910             \fi}%
1911     \fi}%
1912     \bbl@foreach\@filelist{\bbl@tempb#1\@@}% TODO - \@@ de mas??
1913     \@expandtwoargs\in@{,\cf@encoding,},{,\BabelNonASCII,}%
1914     \ifin@else
1915         \edef\ensureascii#1{%
1916             \noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}}%
1917     \fi
1918 \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.

```

1919 \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.

```

1920 \AtBeginDocument{%
1921     \@ifpackageloaded{fontspec}%
1922     {\xdef\latinencoding{%
1923         \ifx\UTFencname\@undefined
1924             EU\ifcase\bbl@engine\or2\or1\fi
1925         \else
1926             \UTFencname
1927         \fi}}%
1928     {\gdef\latinencoding{OT1}%
1929     \ifx\cf@encoding\bbl@t@one
1930         \xdef\latinencoding{\bbl@t@one}%
1931     \else
1932         \@ifl@aded{def}{tlenc}{\xdef\latinencoding{\bbl@t@one}}{}%
1933     \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.

```
1934 \DeclareRobustCommand{\latintext}{%
1935   \fontencoding{\latinencoding}\selectfont
1936   \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.

```
1937 \ifx\@undefined\DeclareTextFontCommand
1938   \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
1939 \else
1940   \DeclareTextFontCommand{\textlatin}{\latintext}
1941 \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 the font engine. However, very little is done at the paragraph level.
- LuaTeX could provide the most complete solution, as we can manipulate almost freely the node list, the generated lines, and so on, but unlike XeTeX 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 LuaTeX-ja shows, vertical typesetting is possible.

```
1942 \def\bbl@trtscripts{%
1943   ,Arabic,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
1944   Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaeen,%
1945   Manichaeen,Meroitic Cursive,Meroitic,Old North Arabian,%
1946   Nabataean,N'Ko,Orkhon,Palmyrene,Inscriptional Pahlavi,%
1947   Psalter Pahlavi,Phoenician,Inscriptional Parthian,Samaritan,%
1948   Old South Arabian,Syriac,Thaana,}%
1949 \def\bbl@ensuredir{%
1950   \@expandtwoargs
1951   \in@\csname bbl@script@\language\endcsname{\bbl@trtscripts}%
1952   \ifin@
1953     \bbl@setdirs\@ne
1954   \else
1955     \bbl@setdirs\z@
1956   \fi}
1957 \def\bbl@setdirs#1{% TODO - math
1958   \ifcase\bbl@select@type % TODO - strictly, not the right test
1959     \bbl@pagedir{#1}%
1960     \bbl@bodydir{#1}%
1961     \bbl@pardir{#1}%
1962   \fi
1963   \bbl@textdir{#1}}
1964 \ifcase\bbl@engine
1965 \or
1966   \AddBabelHook{babel-bidi}{afterextras}{\bbl@ensuredir}
```



```

1967 \DisableBabelHook{babel-bidi}
1968 \def\bbl@getluadir#1{%
1969   \directlua{
1970     if tex.#1dir == 'TLT' then
1971       tex.sprint('0')
1972     elseif tex.#1dir == 'TRT' then
1973       tex.sprint('1')
1974     end}}
1975 \def\bbl@setdir#1#2#3{% 1=text/par.. 2=\textdir.. 3=0 lr/1 rl
1976   \ifcase#3\relax
1977     \ifcase\bbl@getluadir{#1}\relax\else
1978       #2 TLT\relax
1979     \fi
1980   \else
1981     \ifcase\bbl@getluadir{#1}\relax
1982       #2 TRT\relax
1983     \fi
1984   \fi}
1985 \def\bbl@textdir{\bbl@setdir{text}\textdir}% TODO - ?\linedir
1986 \def\bbl@pardir{\bbl@setdir{par}\pardir}
1987 \def\bbl@bodydir{\bbl@setdir{body}\bodydir}
1988 \def\bbl@pagedir{\bbl@setdir{page}\pagedir}
1989 \def\bbl@dirparastext{\pardir\the\textdir\relax}%   %%%
1990 \or
1991 \AddBabelHook{babel-bidi}{afterextras}{\bbl@ensuredir}
1992 \DisableBabelHook{babel-bidi}
1993 \newcount\bbl@dirlevel
1994 \chardef\bbl@thetextdir\z@
1995 \chardef\bbl@thepardir\z@
1996 \def\bbl@textdir#1{%
1997   \ifcase#1\relax
1998     \chardef\bbl@thetextdir\z@
1999     \bbl@textdir@i\beginL\endL
2000   \else
2001     \chardef\bbl@thetextdir@ne
2002     \bbl@textdir@i\beginR\endR
2003   \fi}
2004 \def\bbl@textdir@i#1#2{%
2005   \ifhmode
2006     \ifnum\currentgrouplevel>\z@
2007       \ifnum\currentgrouplevel=\bbl@dirlevel
2008         \bbl@error{Multiple bidi settings inside a group}%
2009         {I'll insert a new group, but expect wrong results.}%
2010         \bgroup\aftergroup#2\aftergroup\egroup
2011       \else
2012         \ifcase\currentgrouptype\or % 0 bottom
2013           \aftergroup#2% 1 simple {}
2014         \or
2015           \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
2016         \or
2017           \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
2018         \or\or\or % vbox vtop align
2019         \or
2020           \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
2021         \or\or\or\or\or\or % output math disc insert vcent mathchoice
2022         \or

```

```

2023         \aftergroup#2% 14 \begingroup
2024     \else
2025         \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
2026     \fi
2027 \fi
2028     \bbl@dirlevel\currentgrouplevel
2029 \fi
2030     #1%
2031 \fi}
2032 \def\bbl@pardir#1{\chardef\bbl@thepardir#1\relax}
2033 \let\bbl@bodydir\@gobble
2034 \let\bbl@pagedir\@gobble
2035 \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.

```

2036 \def\bbl@xebidipar{%
2037     \let\bbl@xebidipar\relax
2038     \TeXeTstate\@ne
2039     \def\bbl@xeverypar{%
2040         \ifcase\bbl@thepardir\else
2041             {\setbox\z@\lastbox\beginR\box\z@}%
2042         \fi
2043         \ifcase\bbl@thetextdir\else\beginR\fi}%
2044     \let\bbl@severypar\everypar
2045     \newtoks\everypar
2046     \everypar=\bbl@severypar
2047     \bbl@severypar{\bbl@xeverypar\the\everypar}}
2048 \fi

```

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`.

```

2049 \ifx\loadlocalcfg\@undefined
2050     \@ifpackagewith{babel}{noconfigs}%
2051     {\let\loadlocalcfg\@gobble}%
2052     {\def\loadlocalcfg#1{%
2053         \InputIfFileExists{#1.cfg}%
2054         {\typeout{*****^J%
2055                 * Local config file #1.cfg used^J%
2056                 *}}%
2057         \@empty}}
2058 \fi

```

Just to be compatible with L^AT_EX 2.09 we add a few more lines of code:

```

2059 \ifx\@unexpandable@protect\@undefined
2060     \def\@unexpandable@protect{\noexpand\protect\noexpand}
2061     \long\def\protected@write#1#2#3{%
2062         \begingroup

```

```

2063 \let\thepage\relax
2064 #2%
2065 \let\protect\@unexpandable@protect
2066 \edef\reserved@a{\write#1{#3}}%
2067 \reserved@a
2068 \endgroup
2069 \if@nobreak\ifvmode\nobreak\fi\fi}
2070 \fi
2071 \</core>

```

10 Multiple languages (switch.def)

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

```

2072 <*kernel>
2073 <<Make sure ProvidesFile is defined>>
2074 \ProvidesFile{switch.def}[\<<date>> \<<version>> Babel switching mechanism]
2075 <<Load macros for plain if not LaTeX>>
2076 <<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.

```

2077 \def\bbl@version{\<<version>>}
2078 \def\bbl@date{\<<date>>}
2079 \def\adddialect#1#2{%
2080   \global\chardef#1#2\relax
2081   \bbl@usehooks{adddialect}{\#1}{\#2}}%
2082   \wlog{\string#1 = a dialect from \string\language#2}

```

`\bbl@iflanguage` executes code only if the language `l@` exists. Otherwise raises and error.

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

```

2083 \def\bbl@fixname#1{%
2084   \begingroup
2085     \def\bbl@tempe{l@}%
2086     \edef\bbl@tempd{\noexpand\ifundefined{\noexpand\bbl@tempe#1}}%
2087     \bbl@tempd
2088       {\lowercase\expandafter{\bbl@tempd}}%
2089       {\uppercase\expandafter{\bbl@tempd}}%
2090       \@empty
2091       {\edef\bbl@tempd{\def\noexpand#1{\#1}}%
2092         \uppercase\expandafter{\bbl@tempd}}}%
2093       {\edef\bbl@tempd{\def\noexpand#1{\#1}}%
2094         \lowercase\expandafter{\bbl@tempd}}}%
2095     \@empty
2096     \edef\bbl@tempd{\endgroup\def\noexpand#1{\#1}}%
2097   \bbl@tempd}

```

```

2098 \def\bbl@iflanguage#1{%
2099   \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.

```

2100 \def\iflanguage#1{%
2101   \bbl@iflanguage{#1}{%
2102     \ifnum\csname l@#1\endcsname=\language
2103       \expandafter\@firstoftwo
2104     \else
2105       \expandafter\@secondoftwo
2106     \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.

```

2107 \let\bbl@select@type\z@
2108 \edef\selectlanguage{%
2109   \noexpand\protect
2110   \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`.

```

2111 \ifx\@undefined\protect\let\protect\relax\fi

```

As L^AT_EX 2.09 writes to files *expanded* whereas L^AT_EX 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.

```

2112 \ifx\documentclass\@undefined
2113   \def\xstring{\string\string\string}

```

```

2114 \else
2115   \let\xstring\string
2116 \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 T_EX'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.

```

2117 \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
`\bbl@pop@language` function can be simple:

```

2118 \def\bbl@push@language{%
2119   \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.

```

2120 \def\bbl@pop@lang#1+#2-#3{%
2121   \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 T_EX 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.

```

2122 \def\bbl@pop@language{%
2123   \expandafter\bbl@pop@lang\bbl@language@stack-\bbl@language@stack
2124   \expandafter\bbl@set@language\expandafter{\language}}

```

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.

```

2125 \expandafter\def\csname selectlanguage \endcsname#1{%
2126   \ifnum\bbl@hymapsel=\@ccclv\let\bbl@hymapsel\tw@\fi
2127   \bbl@push@language
2128   \aftergroup\bbl@pop@language
2129   \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` or `\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.

```

2130 \def\BabelContentsFiles{toc,lof,lot}
2131 \def\bbl@set@language#1{%
2132   \edef\language{%
2133     \ifnum\escapechar=\expandafter'\string#1\@empty
2134     \else\string#1\@empty\fi}%
2135   \select@language{\language}%
2136   \expandafter\ifx\csname date\language\endcsname\relax\else
2137     \if@filesw
2138       \protected@write\auxout{{\string\select@language{\language}}}%
2139       \bbl@foreach\BabelContentsFiles{%
2140         \addtocontents{##1}{\xstring\select@language{\language}}}%
2141       \bbl@usehooks{write}}}%
2142   \fi
2143 \fi}
2144 \def\select@language#1{%
2145   \ifnum\bbl@hymapsel=\@ccclv\chardef\bbl@hymapsel4\relax\fi
2146   \edef\language{#1}%
2147   \bbl@fixname\language
2148   \bbl@iflanguage\language{%
2149     \expandafter\ifx\csname date\language\endcsname\relax
2150       \bbl@error
2151       {Unknown language '#1'. Either you have\\%
2152        misspelled its name, it has not been installed,\\%
2153        or you requested it in a previous run. Fix its name,\\%
2154        install it or just rerun the file, respectively}%
2155       {You may proceed, but expect wrong results}%
2156   \else
2157     \let\bbl@select@type\z@
2158     \expandafter\bbl@switch\expandafter{\language}%
2159   \fi}}

```

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

```

2160 \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.

```

2161 \def\bbl@switch#1{%
2162   \originalTeX
2163   \expandafter\def\expandafter\originalTeX\expandafter{%
2164     \csname noextras#1\endcsname
2165     \let\originalTeX\@empty
2166     \babel@beginsave}%
2167   \bbl@usehooks{afterreset}{}%
2168   \languageshorthands{none}%
2169   \ifcase\bbl@select@type
2170     \csname captions#1\endcsname\relax
2171     \csname date#1\endcsname\relax
2172   \fi
2173   \bbl@usehooks{beforeextras}{}%
2174   \csname extras#1\endcsname\relax
2175   \bbl@usehooks{afterextras}{}%
2176   \ifcase\bbl@opt@hyphenmap\or
2177     \def\BabelLower##1##2{\lccode##1=##2\relax}%
2178     \ifnum\bbl@hymapsel>4\else
2179       \csname\language\name @bbl@hyphenmap\endcsname
2180     \fi
2181     \chardef\bbl@opt@hyphenmap\z@
2182   \else
2183     \ifnum\bbl@hymapsel>\bbl@opt@hyphenmap\else
2184       \csname\language\name @bbl@hyphenmap\endcsname
2185     \fi
2186   \fi
2187   \global\let\bbl@hymapsel\@ccclv
2188   \bbl@patterns{#1}%
2189   \babel@savevariable\lefthyphenmin
2190   \babel@savevariable\righthyphenmin
2191   \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2192     \set@hyphenmins\tw@\thr@@\relax
2193   \else
2194     \expandafter\expandafter\expandafter\set@hyphenmins
2195     \csname #1hyphenmins\endcsname\relax
2196   \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.

```

2197 \long\def\otherlanguage#1{%
2198   \ifnum\bbl@hymapsel=\@ccclv\let\bbl@hymapsel\thr@@\fi

```

```

2199 \csname selectlanguage \endcsname{#1}%
2200 \ignorespaces}

```

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

```

2201 \long\def\endotherlanguage{%
2202 \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`.

```

2203 \expandafter\def\csname otherlanguage*\endcsname#1{%
2204 \ifnum\bbl@hymapset=\@ccclv\chardef\bbl@hymapset4\relax\fi
2205 \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”.

```

2206 \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.

```

2207 \let\bbl@beforeforeign\@empty
2208 \edef\foreignlanguage{%
2209 \noexpand\protect
2210 \expandafter\noexpand\csname foreignlanguage \endcsname}
2211 \expandafter\def\csname foreignlanguage \endcsname{%
2212 \@ifstar\bbl@foreign@s\bbl@foreign@x}
2213 \def\bbl@foreign@x#1#2{%
2214 \begingroup
2215 \let\BabelText\@firstofone
2216 \bbl@beforeforeign

```



```

2217 \foreign@language{#1}%
2218 \bbl@usehooks{foreign}{}%
2219 \BabelText{#2}% Now in horizontal mode!
2220 \endgroup}
2221 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \@setpar, ?\@@par
2222 \begingroup
2223 {\par}%
2224 \let\BabelText\@firstofone
2225 \foreign@language{#1}%
2226 \bbl@usehooks{foreign*}{}%
2227 \bbl@dirparastext
2228 \BabelText{#2}% Still in vertical mode!
2229 {\par}%
2230 \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`.

```

2231 \def\foreign@language#1{%
2232 \edef\languageName{#1}%
2233 \bbl@fixname\languageName
2234 \bbl@iflanguage\languageName{%
2235 \expandafter\ifx\csname date\languageName\endcsname\relax
2236 \bbl@warning
2237 {Unknown language ‘#1’. Either you have\\%
2238 misspelled its name, it has not been installed,\\%
2239 or you requested it in a previous run. Fix its name,\\%
2240 install it or just rerun the file, respectively.\\%
2241 I’ll proceed, but expect wrong results.\\%
2242 Reported}%
2243 \fi
2244 \let\bbl@select@type\@ne
2245 \expandafter\bbl@switch\expandafter{\languageName}}

```

`\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.

```

2246 \let\bbl@hyphlist\@empty
2247 \let\bbl@hyphenation@\relax
2248 \let\bbl@pttnlist\@empty
2249 \let\bbl@patterns@\relax
2250 \let\bbl@hymapsel=\@cclv
2251 \def\bbl@patterns#1{%
2252 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
2253 \csname l@#1\endcsname
2254 \edef\bbl@tempa{#1}%
2255 \else
2256 \csname l@#1:\f@encoding\endcsname
2257 \edef\bbl@tempa{#1:\f@encoding}%

```

```

2258 \fi
2259 \@expandtwoargs\bbl@usehooks{patterns}{\#1}{\bbl@tempa}}%
2260 \@ifundefined{bbl@hyphenation@}{\% Can be \relax!
2261 \begingroup
2262 \@expandtwoargs\in@{,\number\language,}{,\bbl@hyphlist}%
2263 \ifin@else
2264 \@expandtwoargs\bbl@usehooks{hyphenation}{\#1}{\bbl@tempa}}%
2265 \hyphenation{%
2266 \bbl@hyphenation@
2267 \@ifundefined{bbl@hyphenation@#1}%
2268 \empty
2269 {\space\csname bbl@hyphenation@#1\endcsname}}%
2270 \xdef\bbl@hyphlist{\bbl@hyphlist\number\language,}%
2271 \fi
2272 \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 `otherlanguage*`.

```

2273 \def\hyphenrules#1{%
2274 \edef\language{\#1}%
2275 \bbl@fixname\language
2276 \bbl@iflanguage\language{%
2277 \expandafter\bbl@patterns\expandafter{\language}%
2278 \languageshorthands{none}%
2279 \expandafter\ifx\csname\language hyphenmins\endcsname\relax
2280 \set@hyphenmins\tw@\thr@@\relax
2281 \else
2282 \expandafter\expandafter\expandafter\set@hyphenmins
2283 \csname\language hyphenmins\endcsname\relax
2284 \fi}}
2285 \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.

```

2286 \def\providehyphenmins#1#2{%
2287 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2288 \namedef{#1hyphenmins}{#2}%
2289 \fi}

```

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

```

2290 \def\set@hyphenmins#1#2{\lefthyphenmin#1\relax\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.

```

2291 \ifx\ProvidesFile\@undefined

```

```

2292 \def\ProvidesLanguage#1[#2 #3 #4]{%
2293   \wlog{Language: #1 #4 #3 <#2>}%
2294 }
2295 \else
2296 \def\ProvidesLanguage#1{%
2297   \begingroup
2298   \catcode'\ 10 %
2299   \@makeother\/%
2300   \@ifnextchar[%]
2301     {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}}
2302 \def\@provideslanguage#1[#2]{%
2303   \wlog{Language: #1 #2}%
2304   \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
2305   \endgroup}
2306 \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`

```

2307 \def\LdfInit{%
2308   \chardef\atcatcode=\catcode'\@
2309   \catcode'\@=11\relax
2310   \input babel.def\relax
2311   \catcode'\@=\atcatcode \let\atcatcode\relax
2312   \LdfInit}

```

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

```

2313 \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.

```

2314 \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’:

```

2315 \providecommand\setlocale{%
2316   \bbl@error
2317   {Not yet available}%
2318   {Find an armchair, sit down and wait}}
2319 \let\uselocale\setlocale
2320 \let\locale\setlocale
2321 \let\selectlocale\setlocale
2322 \let\textlocale\setlocale
2323 \let\textlanguage\setlocale
2324 \let\languagetext\setlocale

```

10.2 Errors

`\@nolanerr` The babel package will signal an error when a documents tries to select a
`\@nopatterns` 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 L^AT_EX 2_ε, so we can safely use its error handling interface. Otherwise we'll have to 'keep it simple'.

```

2325 \edef\bbl@nulllanguage{\string\language=0}
2326 \ifx\PackageError\@undefined
2327   \def\bbl@error#1#2{%
2328     \begingroup
2329       \newlinechar='^^J
2330       \def\{\^^J(babel) }%
2331       \errhelp{#2}\errmessage{\#1}%
2332     \endgroup}
2333   \def\bbl@warning#1{%
2334     \begingroup
2335       \newlinechar='^^J
2336       \def\{\^^J(babel) }%
2337       \message{\#1}%
2338     \endgroup}
2339   \def\bbl@info#1{%
2340     \begingroup
2341       \newlinechar='^^J
2342       \def\{\^^J}%
2343       \wlog{#1}%
2344     \endgroup}
2345 \else
2346   \def\bbl@error#1#2{%
2347     \begingroup
2348       \def\{\MessageBreak}%
2349       \PackageError{babel}{#1}{#2}%
2350     \endgroup}
2351   \def\bbl@warning#1{%
2352     \begingroup
2353       \def\{\MessageBreak}%
2354       \PackageWarning{babel}{#1}%
2355     \endgroup}
2356   \def\bbl@info#1{%
2357     \begingroup
2358       \def\{\MessageBreak}%
2359       \PackageInfo{babel}{#1}%
2360     \endgroup}
2361 \fi
2362 \@ifpackagewith{babel}{silent}
2363   {\let\bbl@info@gobble
2364    \let\bbl@warning@gobble}
2365 {}
2366 \def\bbl@nocaption#1#2{% 1: text to be printed 2: caption macro \langXname
2367   \gdef#2{\textbf{?#1?}}%
2368   #2%
2369   \bbl@warning{%
2370     \string#2 not set. Please, define\\%
2371     it in the preamble with something like:\\%
2372     \string\renewcommand\string#2{..}\\%

```

```

2373     Reported}}
2374 \def\nolanerr#1{%
2375   \bbl@error
2376   {You haven't defined the language #1\space yet}%
2377   {Your command will be ignored, type <return> to proceed}}
2378 \def\nopatterns#1{%
2379   \bbl@warning
2380   {No hyphenation patterns were preloaded for\\%
2381     the language '#1' into the format.\\%
2382     Please, configure your TeX system to add them and\\%
2383     rebuild the format. Now I will use the patterns\\%
2384     preloaded for \bbl@nulllanguage\space instead}}
2385 \let\bbl@usehooks@gobbletwo
2386 </kernel>

```

11 Loading hyphenation patterns

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

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

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

To make sure that `LATEX 2.09` executes the `\@begindocumenthook` we would want to alter `\begin{document}`, but as this done too often already, we add the new code at the front of `\@preamblecmds`. But we can only do that after it has been defined, so we add this piece of code to `\dump`.

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

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

```

2387 (*patterns)
2388 <<Make sure ProvidesFile is defined>>
2389 \ProvidesFile{hyphen.cfg}[<<date>> <<version>>] Babel hyphens]
2390 \xdef\bbl@format{\jobname}
2391 \ifx\AtBeginDocument\@undefined
2392   \def\@empty{}
2393   \let\orig@dump\dump
2394   \def\dump{%
2395     \ifx\@ztryfc\@undefined
2396     \else
2397       \toks0=\expandafter{\@preamblecmds}%
2398       \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
2399       \def\@begindocumenthook{}%
2400     \fi
2401     \let\dump\orig@dump\let\orig@dump\@undefined\dump}
2402 \fi
2403 <<Define core switching macros>>
2404 \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.

```

2405 \def\process@line#1#2 #3 #4 {%
2406   \ifx=#1%
2407     \process@synonym{#2}%
2408   \else
2409     \process@language{#1#2}{#3}{#4}%
2410   \fi
2411   \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.

```

2412 \toks@{}
2413 \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.

```

2414 \def\process@synonym#1{%
2415   \ifnum\last@language=\m@ne
2416     \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
2417   \else
2418     \expandafter\chardef\csname l@#1\endcsname\last@language
2419     \wlog{\string\l@#1=\string\language\the\last@language}%
2420     \expandafter\let\csname #1hyphenmins\endcsname\expandafter\endcsname
2421     \csname\language\hyphenmins\endcsname

```

```

2422 \let\bbl@elt\relax
2423 \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}}{}}}%
2424 \fi}

```

`\process@language` The macro `\process@language` is used to process a non-empty line from the ‘configuration file’. It has three arguments, each delimited by white space. The first argument is the ‘name’ of a language; the second is the name of the file that contains the patterns. The optional third argument is the name of a file containing hyphenation exceptions.

The first thing to do is call `\addlanguage` to allocate a pattern register and to make that register ‘active’. Then the ‘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`. T_EX does not keep track of these assignments. Therefore we try to detect such assignments and store them in the `\(lang)hyphenmins` macro. When no assignments were made we provide a default setting.

Some pattern files contain changes to the `\lccode` 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.

```

2425 \def\process@language#1#2#3{%
2426 \expandafter\addlanguage\csname l@#1\endcsname
2427 \expandafter\language\csname l@#1\endcsname
2428 \edef\languagename{#1}%
2429 \bbl@hook@everylanguage{#1}%
2430 \bbl@get@enc#1::\@@@
2431 \begingroup
2432 \lefthyphenmin\m@ne
2433 \bbl@hook@loadpatterns{#2}%
2434 \ifnum\lefthyphenmin=\m@ne
2435 \else
2436 \expandafter\xdef\csname #1hyphenmins\endcsname{%
2437 \the\lefthyphenmin\the\righthyphenmin}%
2438 \fi
2439 \endgroup
2440 \def\bbl@tempa{#3}%
2441 \ifx\bbl@tempa\empty\else

```

```

2442 \bbl@hook@loadexceptions{#3}%
2443 \fi
2444 \let\bbl@elt\relax
2445 \edef\bbl@languages{%
2446 \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
2447 \ifnum\the\language=\z@
2448 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
2449 \set@hyphenmins\tw@\thr@@\relax
2450 \else
2451 \expandafter\expandafter\expandafter\set@hyphenmins
2452 \csname #1hyphenmins\endcsname
2453 \fi
2454 \the\toks@
2455 \toks@{}}%
2456 \fi}

```

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

```

2457 \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.

```

2458 \def\bbl@hook@everylanguage#1{}
2459 \def\bbl@hook@loadpatterns#1{\input #1\relax}
2460 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
2461 \let\bbl@hook@loadkernel\bbl@hook@loadpatterns
2462 \begingroup
2463 \def\AddBabelHook#1#2{%
2464 \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
2465 \def\next{\toks1}%
2466 \else
2467 \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname###1}%
2468 \fi
2469 \next}
2470 \ifx\directlua\@undefined
2471 \ifx\XeTeXinputencoding\@undefined\else
2472 \input xebabel.def
2473 \fi
2474 \else
2475 \input luababel.def
2476 \fi
2477 \openin1 = babel-\bbl@format.cfg
2478 \ifeof1
2479 \else
2480 \input babel-\bbl@format.cfg\relax
2481 \fi
2482 \closein1
2483 \endgroup
2484 \bbl@hook@loadkernel{switch.def}

```

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

```

2485 \openin1 = language.dat

```

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


```

2486 \def\language{english}%
2487 \ifeof1
2488   \message{I couldn't find the file language.dat,\space
2489           I will try the file hyphen.tex}
2490   \input hyphen.tex\relax
2491   \chardef\l@english\z@
2492 \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`.

```

2493 \last@language\m@ne

```

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

```

2494 \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.

```

2495   \endlinechar\m@ne
2496   \read1 to \bbl@line
2497   \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.

```

2498   \if T\ifeof1F\fi T\relax
2499   \ifx\bbl@line@empty\else
2500     \edef\bbl@line{\bbl@line\space\space\space}%
2501     \expandafter\process@line\bbl@line\relax
2502   \fi
2503 \repeat

```

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

```

2504 \begingroup
2505   \def\bbl@elt#1#2#3#4{%
2506     \global\language=#2\relax
2507     \gdef\language{#1}%
2508     \def\bbl@elt##1##2##3##4{}}%
2509   \bbl@languages
2510 \endgroup
2511 \fi

```

and close the configuration file.

```

2512 \closein1

```

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

```

2513 \if/\the\toks@/\else
2514   \errhelp{language.dat loads no language, only synonyms}
2515   \errmessage{Orphan language synonym}
2516 \fi

```

```

2517 \advance\last@language\@ne
2518 \edef\bbl@tempa{%
2519   \everyjob{%
2520     \the\everyjob
2521     \ifx\typeout\@undefined
2522       \immediate\writel6%
2523     \else
2524       \noexpand\typeout
2525     \fi
2526     {\the\toks8 \the\last@language\space language(s) loaded.}}
2527 \advance\last@language\@ne
2528 \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.

```

2529 \let\bbl@line\@undefined
2530 \let\process@line\@undefined
2531 \let\process@synonym\@undefined
2532 \let\process@language\@undefined
2533 \let\bbl@get@enc\@undefined
2534 \let\bbl@hyph@enc\@undefined
2535 \let\bbl@tempa\@undefined
2536 \let\bbl@hook@loadkernel\@undefined
2537 \let\bbl@hook@everylanguage\@undefined
2538 \let\bbl@hook@loadpatterns\@undefined
2539 \let\bbl@hook@loadexceptions\@undefined
2540 \let\patterns\@undefined

```

Here the code for iniT_EX ends.

12 Tentative font handling with fontspec

A general solution is far from trivial:

- \addfontfeature only sets it for the current family and it's not very efficient, and
- \defaultfontfeatures requires to redefine the font (and the options are not "orthogonal").

```

2541 <<(*Font selection)>> ≡
2542 \newcommand\babelFSstore[2][]{%
2543   \bbl@ifblank{#1}%
2544   {\bbl@csarg\def{script@#2}{Latin}}%
2545   {\bbl@csarg\def{script@#2}{#1}}%
2546   \@expandtwoargs % TODO should go to the ini loaders
2547   \in@{\csname bbl@script@#2\endcsname}{\bbl@trtscripts}%
2548   \ifin@
2549     \let\bbl@beforeforeign\leavevmode
2550     \EnableBabelHook{babel-bidi}%
2551     \@nameuse{bbl@xebidipar}%
2552   \fi
2553   \bbl@foreach{#2}{%
2554     \bbl@FSstore{##1}{rm}\rmdefault\bbl@save@rmdefault
2555     \bbl@FSstore{##1}{sf}\sfdefault\bbl@save@sfdefault

```

```

2556 \bbl@FSstore{##1}{tt}\ttdefault\bbl@save@ttdefault}}
2557 \def\bbl@FSstore#1#2#3#4{%
2558 \bbl@csarg\edef{#2default#1}{#3}%
2559 \expandafter\addto\csname extras#1\endcsname{%
2560 \let#4#3%
2561 \ifx#3\f@family
2562 \edef#3{\csname bbl@#2default#1\endcsname}%
2563 \fontfamily{#3}\selectfont
2564 \else
2565 \edef#3{\csname bbl@#2default#1\endcsname}%
2566 \fi}%
2567 \expandafter\addto\csname noextras#1\endcsname{%
2568 \ifx#3\f@family
2569 \fontfamily{#4}\selectfont
2570 \fi
2571 \let#3#4}}
2572 \let\bbl@langfeatures\@empty
2573 \def\babelFSfeatures{% make sure \fontspec is redefined once
2574 \let\bbl@ori@fontspec\fontspec
2575 \renewcommand\fontspec[1][{}]{%
2576 \bbl@ori@fontspec[\bbl@langfeatures##1]}
2577 \let\babelFSfeatures\bbl@FSfeatures
2578 \babelFSfeatures}
2579 \def\bbl@FSfeatures#1#2{%
2580 \expandafter\addto\csname extras#1\endcsname{%
2581 \babel@save\bbl@langfeatures
2582 \edef\bbl@langfeatures{#2,}}}%
2583 <</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.

L^AT_EX 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 L^AT_EX. Anyway, for consistency Lua_TE_X also resets the catcodes.

```

2584 <<{*Restore Unicode catcodes before loading patterns}>> ≡
2585 \begingroup
2586 % Reset chars "80-"C0 to category "other", no case mapping:
2587 \catcode'\@=11 \count@=128
2588 \loop\ifnum\count@<192
2589 \global\uccode\count@=0 \global\lccode\count@=0
2590 \global\catcode\count@=12 \global\sffcode\count@=1000
2591 \advance\count@ by 1 \repeat
2592 % Other:
2593 \def\0 ##1 {%
2594 \global\uccode"##1=0 \global\lccode"##1=0
2595 \global\catcode"##1=12 \global\sffcode"##1=1000 }%
2596 % Letter:
2597 \def\L ##1 ##2 ##3 {\global\catcode"##1=11
2598 \global\uccode"##1="##2

```

```

2599      \global\lccode"##1="##3
2600      % Uppercase letters have sfcode=999:
2601      \ifnum"##1="##3 \else \global\sfcode"##1=999 \fi }%
2602      % Letter without case mappings:
2603      \def\l ##1 {\L ##1 ##1 ##1 }%
2604      \l 00AA
2605      \L 00B5 039C 00B5
2606      \l 00BA
2607      \O 00D7
2608      \l 00DF
2609      \O 00F7
2610      \L 00FF 0178 00FF
2611      \endgroup
2612      \input #1\relax
2613      <</Restore Unicode catcodes before loading patterns>>

```

Now, the code.

```

2614 <*\xetex>
2615 \def\BabelStringsDefault{unicode}
2616 \let\xebbl@stop\relax
2617 \AddBabelHook{xetex}{encodedcommands}{%
2618   \def\bbl@tempa{#1}%
2619   \ifx\bbl@tempa\@empty
2620     \XeTeXinputencoding"bytes"%
2621   \else
2622     \XeTeXinputencoding"#1"%
2623   \fi
2624   \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
2625 \AddBabelHook{xetex}{stopcommands}{%
2626   \xebbl@stop
2627   \let\xebbl@stop\relax}
2628 \AddBabelHook{xetex}{loadkernel}{%
2629 <<Restore Unicode catcodes before loading patterns>>}
2630 <<Font selection>>
2631 </xetex>

```

13.2 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 has been preloaded into the format. This is not optimal, but it shouldn't happen very often – with luatex patterns are best loaded when the document is typeset, and the “0th” language is preloaded just for backwards compatibility.

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

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

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

```

2632 (*luatex)
2633 \ifx\AddBabelHook\@undefined
2634 \begingroup
2635   \toks@{}
2636   \count@\z@ % 0=start, 1=0th, 2=normal
2637   \def\bbl@process@line#1#2 #3 #4 {%
2638     \ifx=#1%
2639       \bbl@process@synonym{#2}%
2640     \else
2641       \bbl@process@language{#1#2}{#3}{#4}%
2642     \fi
2643     \ignorespaces}
2644   \def\bbl@manylang{%
2645     \ifnum\bbl@last>\@ne
2646       \bbl@info{Non-standard hyphenation setup}%
2647     \fi
2648     \let\bbl@manylang\relax}
2649   \def\bbl@process@language#1#2#3{%
2650     \ifcase\count@
2651       \@ifundefined{zth@#1}{\count@\tw@}{\count@\@ne}%
2652     \or
2653       \count@\tw@
2654     \fi
2655     \ifnum\count@=\tw@
2656       \expandafter\addlanguage\csname l@#1\endcsname
2657       \language\allocationnumber
2658       \chardef\bbl@last\allocationnumber
2659       \bbl@manylang
2660       \let\bbl@elt\relax
2661       \xdef\bbl@languages{%
2662         \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
2663     \fi
2664     \the\toks@
2665     \toks@{}}
2666   \def\bbl@process@synonym@aux#1#2{%
2667     \global\expandafter\chardef\csname l@#1\endcsname#2\relax
2668     \let\bbl@elt\relax

```

```

2669 \xdef\bbbl@languages{%
2670 \bbbl@languages\bbbl@elt{#1}{#2}{}}}%
2671 \def\bbbl@process@synonym#1{%
2672 \ifcase\count@
2673 \toks@\expandafter{\the\toks@\relax\bbbl@process@synonym{#1}}%
2674 \or
2675 \ifundefined{zth@#1}{\bbbl@process@synonym@aux{#1}{0}}}%
2676 \else
2677 \bbbl@process@synonym@aux{#1}{\the\bbbl@last}%
2678 \fi}
2679 \ifx\bbbl@languages\@undefined % Just a (sensible?) guess
2680 \chardef\l@english\z@
2681 \chardef\l@USenglish\z@
2682 \chardef\bbbl@last\z@
2683 \global\@namedef{bbbl@hyphendata@0}{{hyphen.tex}}
2684 \gdef\bbbl@languages{%
2685 \bbbl@elt{english}{0}{hyphen.tex}}%
2686 \bbbl@elt{USenglish}{0}{}}
2687 \else
2688 \global\let\bbbl@languages@format\bbbl@languages
2689 \def\bbbl@elt#1#2#3#4{% Remove all except language 0
2690 \ifnum#2>\z@\else
2691 \noexpand\bbbl@elt{#1}{#2}{#3}{#4}%
2692 \fi}%
2693 \xdef\bbbl@languages{\bbbl@languages}%
2694 \fi
2695 \def\bbbl@elt#1#2#3#4{\@namedef{zth@#1}} % Define flags
2696 \bbbl@languages
2697 \openin1=language.dat
2698 \ifeof1
2699 \bbbl@warning{I couldn't find language.dat. No additional\\%
2700 patterns loaded. Reported}%
2701 \else
2702 \loop
2703 \endlinechar\m@ne
2704 \read1 to \bbbl@line
2705 \endlinechar'\^^M
2706 \if T\ifeof1F\fi T\relax
2707 \ifx\bbbl@line\@empty\else
2708 \edef\bbbl@line{\bbbl@line\space\space\space}%
2709 \expandafter\bbbl@process@line\bbbl@line\relax
2710 \fi
2711 \repeat
2712 \fi
2713 \endgroup
2714 \def\bbbl@get@enc#1:#2:#3\@@{\def\bbbl@hyph@enc{#2}}
2715 \ifx\babelcatcodetablenum\@undefined
2716 \def\babelcatcodetablenum{5211}
2717 \fi
2718 \def\bbbl@luapatterns#1#2{%
2719 \bbbl@get@enc#1::\@@@
2720 \setbox\z@\hbox\bgroup
2721 \begingroup
2722 \ifx\catcodetable\@undefined
2723 \let\savecatcodetable\luatexsavecatcodetable
2724 \let\initcatcodetable\luatexinitcatcodetable

```

```

2725 \let\catcodetable\luatexcatcodetable
2726 \fi
2727 \savecatcodetable\babelcatcodetablenum\relax
2728 \initcatcodetable\numexpr\babelcatcodetablenum+1\relax
2729 \catcodetable\numexpr\babelcatcodetablenum+1\relax
2730 \catcode'\#6 \catcode'\$3 \catcode'\&4 \catcode'\^7
2731 \catcode'\_8 \catcode'\{1 \catcode'\}=2 \catcode'\~13
2732 \catcode'\@11 \catcode'\^^I10 \catcode'\^^J12
2733 \catcode'\<12 \catcode'\>12 \catcode'\*12 \catcode'\.=12
2734 \catcode'\-=12 \catcode'\/=12 \catcode'\[12 \catcode'\]=12
2735 \catcode'\'=12 \catcode'\'=12 \catcode'\ "=12
2736 \input #1\relax
2737 \catcodetable\babelcatcodetablenum\relax
2738 \endgroup
2739 \def\bbl@tempa{#2}%
2740 \ifx\bbl@tempa\empty\else
2741 \input #2\relax
2742 \fi
2743 \egroup}%
2744 \def\bbl@patterns@lua#1{%
2745 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
2746 \csname l@#1\endcsname
2747 \edef\bbl@tempa{#1}%
2748 \else
2749 \csname l@#1:\f@encoding\endcsname
2750 \edef\bbl@tempa{#1:\f@encoding}%
2751 \fi\relax
2752 \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
2753 \@ifundefined{bbl@hyphendata@the\language}%
2754 {\def\bbl@elt##1##2##3##4{%
2755 \ifnum##2=\csname l@bbl@tempa\endcsname % #2=spanish, dutch:OT1...
2756 \def\bbl@tempb{##3}%
2757 \ifx\bbl@tempb\empty\else % if not a synonymous
2758 \def\bbl@tempc{##3}{##4}}%
2759 \fi
2760 \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
2761 \fi}%
2762 \bbl@languages
2763 \@ifundefined{bbl@hyphendata@the\language}%
2764 {\bbl@info{No hyphenation patterns were set for\%
2765 language '\bbl@tempa'. Reported}}%
2766 {\expandafter\expandafter\expandafter\bbl@luapatterns
2767 \csname bbl@hyphendata@the\language\endcsname}}}%
2768 \endinput\fi
2769 \begingroup
2770 \catcode'\%=12
2771 \catcode'\ '=12
2772 \catcode'\ "=12
2773 \catcode'\:=12
2774 \directlua{
2775 Babel = {}
2776 function Babel.bytes(line)
2777 return line:gsub(".",
2778 function (chr) return unicode.utf8.char(string.byte(chr)) end)
2779 end
2780 function Babel.begin_process_input()

```

```

2781   if luatexbase and luatexbase.add_to_callback then
2782       luatexbase.add_to_callback('process_input_buffer',
2783                                   Babel.bytes, 'Babel.bytes')
2784   else
2785       Babel.callback = callback.find('process_input_buffer')
2786       callback.register('process_input_buffer', Babel.bytes)
2787   end
2788 end
2789 function Babel.end_process_input ()
2790   if luatexbase and luatexbase.remove_from_callback then
2791       luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
2792   else
2793       callback.register('process_input_buffer', Babel.callback)
2794   end
2795 end
2796 function Babel.addpatterns(pp, lg)
2797   local lg = lang.new(lg)
2798   local pats = lang.patterns(lg) or ''
2799   lang.clear_patterns(lg)
2800   for p in pp:gmatch('[^%s]+') do
2801       ss = ''
2802       for i in string.utfcharacters(p:gsub('%d', '')) do
2803           ss = ss .. '%d?' .. i
2804       end
2805       ss = ss:gsub('^%%d%?%.', '%%.') .. '%d?'
2806       ss = ss:gsub('%.%d%?$', '%%.')
2807       pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
2808       if n == 0 then
2809           tex.sprint(
2810               [[\string\csname\space bbl@info\endcsname{New pattern: }]]
2811               .. p .. [[{}]])
2812           pats = pats .. ' ' .. p
2813       else
2814           tex.sprint(
2815               [[\string\csname\space bbl@info\endcsname{Renew pattern: }]]
2816               .. p .. [[{}]])
2817       end
2818   end
2819   lang.patterns(lg, pats)
2820 end
2821 }
2822 \endgroup
2823 \def\BabelStringsDefault{unicode}
2824 \let\luabbbl@stop\relax
2825 \AddBabelHook{luatex}{encodedcommands}{%
2826   \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
2827   \ifx\bbl@tempa\bbl@tempb\else
2828     \directlua{Babel.begin_process_input()}%
2829   \def\luabbbl@stop{%
2830     \directlua{Babel.end_process_input()}}%
2831   \fi}%
2832 \AddBabelHook{luatex}{stopcommands}{%
2833   \luabbbl@stop
2834   \let\luabbbl@stop\relax}
2835 \AddBabelHook{luatex}{patterns}{%
2836   \ifundefined{bbl@hyphendata@the\language}%

```



```

2837 {\def\bbl@elt##1##2##3##4{%
2838   \ifnum##2=\csname l@#2\endcsname % #2=spanish, dutch:OT1...
2839   \def\bbl@tempb{##3}%
2840   \ifx\bbl@tempb\@empty\else % if not a synonymous
2841     \def\bbl@tempc{##3}{##4}}%
2842   \fi
2843   \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
2844   \fi}%
2845 \bbl@languages
2846 \@ifundefined{bbl@hyphendata@the\language}%
2847   {\bbl@info{No hyphenation patterns were set for\%
2848     language '#2'. Reported}}%
2849   {\expandafter\expandafter\expandafter\bbl@luapatterns
2850     \csname bbl@hyphendata@the\language\endcsname}}}%
2851 \@ifundefined{bbl@patterns@}{}%
2852 \begingroup
2853   \@expandtwoargs\in@{,\number\language,}{,\bbl@pttnlist}%
2854   \ifin@else
2855     \ifx\bbl@patterns@\@empty\else
2856       \directlua{ Babel.addpatterns(
2857         [[\bbl@patterns@]], \number\language) }%
2858     \fi
2859     \@ifundefined{bbl@patterns@#1}%
2860     \@empty
2861     {\directlua{ Babel.addpatterns(
2862       [[\space\csname bbl@patterns@#1\endcsname]],
2863       \number\language) }}%
2864     \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
2865   \fi
2866 \endgroup}}
2867 \AddBabelHook{luatex}{everylanguage}{%
2868   \def\process@language##1##2##3{%
2869     \def\process@line####1####2 #####3 ####4 {}}}
2870 \AddBabelHook{luatex}{loadpatterns}{%
2871   \input #1\relax
2872   \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
2873     {#1}{}}}
2874 \AddBabelHook{luatex}{loadexceptions}{%
2875   \input #1\relax
2876   \def\bbl@tempb##1##2{##1}{#1}}%
2877   \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
2878     {\expandafter\expandafter\expandafter\bbl@tempb
2879       \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.

```

2880 \@onlypreamble\babelpatterns
2881 \AtEndOfPackage{%
2882   \newcommand\babelpatterns[2][\@empty]{%
2883     \ifx\bbl@patterns@relax
2884       \let\bbl@patterns@\@empty
2885     \fi
2886     \ifx\bbl@pttnlist\@empty\else
2887       \bbl@warning{%
2888         You must not intermingle \string\selectlanguage\space and\%

```

```

2889      \string\babelpatterns\space or some patterns will not\\%
2890      be taken into account. Reported}%
2891  \fi
2892  \ifx\@empty#1%
2893      \protected@edef\bbl@patterns@{\bbl@patterns@space#2}%
2894  \else
2895      \edef\bbl@tempb{\zap@space#1 \@empty}%
2896      \bbl@for\bbl@tempa\bbl@tempb{%
2897          \bbl@fixname\bbl@tempa
2898          \bbl@iflanguage\bbl@tempa{%
2899              \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
2900                  \ifundefined{bbl@patterns@\bbl@tempa}%
2901                      \@empty
2902                      {\csname bbl@patterns@\bbl@tempa\endcsname\space}%
2903                      #2}}}%
2904  \fi}}

```

Common stuff.

```

2905 \AddBabelHook{luatex}{loadkernel}{%
2906 <<Restore Unicode catcodes before loading patterns>>}
2907 <<Font selection>>
2908 </luatex>

```

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.

```

2909 (*nil)
2910 \ProvidesLanguage{nil}[<<date>> <<version>> Nil language]
2911 \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.

```

2912 \ifx\l@nohyphenation\@undefined
2913     \@nopatterns{nil}
2914     \adddialect\l@nil0
2915 \else
2916     \let\l@nil\l@nohyphenation
2917 \fi

```

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

```

2918 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}

```

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

```

\captionnil
\datenil
2919 \let\captionnil\@empty
2920 \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.

```
2921 \ldf@finish{nil}
2922 \</nil>
```

15 Support for Plain T_EX (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 T_EX-format. When asked he responded:

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

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

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

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

```
2923 (*bplain | lplain)
2924 \catcode'\{=1 % left brace is begin-group character
2925 \catcode'\}=2 % right brace is end-group character
2926 \catcode'\#=6 % hash mark is macro parameter character
```

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

```
2927 \openin 0 hyphen.cfg
```

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

```
2928 \ifeof0
2929 \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).

```
2930 \let\input
```

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

```
2931 \def\input #1 {%
2932   \let\input\input
2933   \input #1
}
```

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

```
2934 \let\a\undefined
2935 }
2936 \fi
2937 </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`.

```
2938 <bplain>\a plain.tex
2939 <bplain>\a lplain.tex
```

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

```
2940 <bplain>\def\fmtname{babel-plain}
2941 <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 L^AT_EX features

The following code duplicates or emulates parts of L^AT_EX 2_ε that are needed for `babel`.

```
2942 <*plain>
2943 \def\@empty{}
2944 \def\loadlocalcfg#1{%
2945   \openin0#1.cfg
2946   \ifeof0
2947     \closein0
2948   \else
2949     \closein0
2950     {\immediate\writel6{*****}%
2951      \immediate\writel6{* Local config file #1.cfg used}%
2952      \immediate\writel6{*}%
2953     }
2954     \input #1.cfg\relax
2955   \fi
2956   \@endofldf}
```

15.3 General tools

A number of L^AT_EX macro's that are needed later on.

```
2957 \long\def\@firstofone#1{#1}
2958 \long\def\@firstoftwo#1#2{#1}
2959 \long\def\@secondoftwo#1#2{#2}
2960 \def\@nnil{\@nil}
2961 \def\@gobbletwo#1#2{}
2962 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
2963 \def\@star@or@long#1{%
2964   \@ifstar
2965   {\let\l@ngrel@x\relax#1}%
2966   {\let\l@ngrel@x\long#1}}
2967 \let\l@ngrel@x\relax
2968 \def\@car#1#2\@nil{#1}
```

```

2969 \def\@cdr#1#2\@nil{#2}
2970 \let\@typeset@protect\relax
2971 \let\protected@edef\edef
2972 \long\def\@gobble#1{}
2973 \edef\@backslashchar{\expandafter\@gobble\string\\}
2974 \def\strip@prefix#1>{}
2975 \def\g@addto@macro#1#2{{%
2976   \toks@{\expandafter{#1#2}%
2977   \xdef#1{\the\toks@}}}
2978 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
2979 \def\@nameuse#1{\csname #1\endcsname}
2980 \def\@ifundefined#1{%
2981   \expandafter\ifx\csname#1\endcsname\relax
2982     \expandafter\@firstoftwo
2983   \else
2984     \expandafter\@secondoftwo
2985   \fi}
2986 \def\@expandtwoargs#1#2#3{%
2987   \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
2988 \def\zap@space#1 #2{%
2989   #1%
2990   \ifx#2\@empty\else\expandafter\zap@space\fi
2991   #2}

```

L^AT_EX 2_ε has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

2992 \ifx\@preamblecmds\@undefined
2993   \def\@preamblecmds{}
2994 \fi
2995 \def\@onlypreamble#1{%
2996   \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
2997     \@preamblecmds\do#1}}
2998 \@onlypreamble\@onlypreamble

```

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

```

2999 \def\begindocument{%
3000   \@begindocumenthook
3001   \global\let\@begindocumenthook\@undefined
3002   \def\do##1{\global\let##1\@undefined}%
3003   \@preamblecmds
3004   \global\let\do\noexpand}
3005 \ifx\@begindocumenthook\@undefined
3006   \def\@begindocumenthook{}
3007 \fi
3008 \@onlypreamble\@begindocumenthook
3009 \def\AtBeginDocument{\g@addto@macro\@begindocumenthook}

```

We also have to mimick L^AT_EX's `\AtEndOfPackage`. Our replacement macro is much simpler; it stores its argument in `\@endofldf`.

```

3010 \def\AtEndOfPackage#1{\g@addto@macro\@endofldf{#1}}
3011 \@onlypreamble\AtEndOfPackage
3012 \def\@endofldf{}
3013 \@onlypreamble\@endofldf
3014 \let\bbl@afterlang\@empty
3015 \chardef\bbl@opt@hyphenmap\z@

```

L^AT_EX needs to be able to switch off writing to its auxiliary files; plain doesn't have them by default.

```

3016 \ifx\if@filesw\@undefined
3017   \expandafter\let\csname if@filesw\expandafter\endcsname
3018     \csname iffalse\endcsname
3019 \fi

```

Mimick L^AT_EX's commands to define control sequences.

```

3020 \def\newcommand{\@star@or@long\new@command}
3021 \def\new@command#1{%
3022   \@testopt{\@newcommand#1}0}
3023 \def\@newcommand#1[#2]{%
3024   \@ifnextchar [{\@xargdef#1[#2]}%
3025     {\@argdef#1[#2]}}
3026 \long\def\@argdef#1[#2]#3{%
3027   \@yargdef#1\@ne{#2}{#3}}
3028 \long\def\@xargdef#1[#2][#3]#4{%
3029   \expandafter\def\expandafter#1\expandafter{%
3030     \expandafter\@protected@testopt\expandafter #1%
3031     \csname\string#1\expandafter\endcsname{#3}}}%
3032   \expandafter\@yargdef \csname\string#1\endcsname
3033     \tw@{#2}{#4}}
3034 \long\def\@yargdef#1#2#3{%
3035   \@tempcnta#3\relax
3036   \advance \@tempcnta \@ne
3037   \let\@hash@\relax
3038   \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
3039   \@tempcntb #2%
3040   \@whilenum\@tempcntb <\@tempcnta
3041   \do{%
3042     \edef\reserved@a{\reserved@a\@hash@\the\@tempcntb}%
3043     \advance\@tempcntb \@ne}%
3044   \let\@hash@##%
3045   \l@ngrel@x\expandafter\def\expandafter#1\reserved@a}
3046 \def\providecommand{\@star@or@long\provide@command}
3047 \def\provide@command#1{%
3048   \begingroup
3049     \escapechar\m@ne\xdef\@gtempa{\string#1}%
3050   \endgroup
3051   \expandafter\@ifundefined\@gtempa
3052     {\def\reserved@a{\new@command#1}}%
3053     {\let\reserved@a\relax
3054     \def\reserved@a{\new@command\reserved@a}}%
3055   \reserved@a}%
3056 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
3057 \def\declare@robustcommand#1{%
3058   \edef\reserved@a{\string#1}%
3059   \def\reserved@b{#1}%
3060   \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
3061   \edef#1{%
3062     \ifx\reserved@a\reserved@b
3063       \noexpand\x@protect
3064       \noexpand#1%
3065     \fi
3066     \noexpand\protect

```

```

3067 \expandafter\noexpand\csname\bbl@stripslash#1 \endcsname
3068 }%
3069 \expandafter\newcommand\csname\bbl@stripslash#1 \endcsname
3070 }
3071 \def\x@protect#1{%
3072 \ifx\protect@typeset@protect\else
3073 \x@protect#1%
3074 \fi
3075 }
3076 \def\x@protect#1\fi#2#3{%
3077 \fi\protect#1%
3078 }

```

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`.

```

3079 \def\bbl@tempa{\csname newif\endcsname\ifin@}
3080 \ifx\in@\@undefined
3081 \def\in@#1#2{%
3082 \def\in@@##1#1##2##3\in@@{%
3083 \ifx\in@##2\in@false\else\in@true\fi}%
3084 \in@@#2#1\in@\in@@}
3085 \else
3086 \let\bbl@tempa\@empty
3087 \fi
3088 \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).

```

3089 \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.

```

3090 \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 $\text{\LaTeX} 2_{\epsilon}$ versions; just enough to make things work in plain \TeX environments.

```

3091 \ifx\@tempcnta\@undefined
3092 \csname newcount\endcsname\@tempcnta\relax
3093 \fi
3094 \ifx\@tempcntb\@undefined
3095 \csname newcount\endcsname\@tempcntb\relax
3096 \fi

```

To prevent wasting two counters in $\text{\LaTeX} 2.09$ (because counters with the same name are allocated later by it) we reset the counter that holds the next free counter (`\count10`).

```

3097 \ifx\bye\@undefined

```

```

3098 \advance\count10 by -2\relax
3099 \fi
3100 \ifx\@ifnextchar\@undefined
3101 \def\@ifnextchar#1#2#3{%
3102 \let\reserved@d=#1%
3103 \def\reserved@a{#2}\def\reserved@b{#3}%
3104 \futurelet\@let@token\@ifnch}
3105 \def\@ifnch{%
3106 \ifx\@let@token\@sptoken
3107 \let\reserved@c\@xifnch
3108 \else
3109 \ifx\@let@token\reserved@d
3110 \let\reserved@c\reserved@a
3111 \else
3112 \let\reserved@c\reserved@b
3113 \fi
3114 \fi
3115 \reserved@c}
3116 \def\:{\let\@sptoken= } \: % this makes \@sptoken a space token
3117 \def\:{\@xifnch} \expandafter\def\:{\futurelet\@let@token\@ifnch}
3118 \fi
3119 \def\@testopt#1#2{%
3120 \@ifnextchar[#{#1}{#1[#2]}}
3121 \def\@protected@testopt#1{%
3122 \ifx\protect\@typeset@protect
3123 \expandafter\@testopt
3124 \else
3125 \@x@protect#1%
3126 \fi}
3127 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
3128 #2\relax}\fi}
3129 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
3130 \else\expandafter\@gobble\fi{#1}}

```

15.4 Encoding related macros

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

```

3131 \def\DeclareTextCommand{%
3132 \@dec@text@cmd\providecommand
3133 }
3134 \def\ProvideTextCommand{%
3135 \@dec@text@cmd\providecommand
3136 }
3137 \def\DeclareTextSymbol#1#2#3{%
3138 \@dec@text@cmd\chardef#1{#2}#3\relax
3139 }
3140 \def\@dec@text@cmd#1#2#3{%
3141 \expandafter\def\expandafter#2%
3142 \expandafter{%
3143 \csname#3-cmd\expandafter\endcsname
3144 \expandafter#2%
3145 \csname#3\string#2\endcsname
3146 }%
3147 % \let\@ifdefinable\@rc@ifdefinable
3148 \expandafter#1\csname#3\string#2\endcsname

```



```

3149 }
3150 \def\@current@cmd#1{%
3151   \ifx\protect\@typeset@protect\else
3152     \noexpand#1\expandafter\@gobble
3153   \fi
3154 }
3155 \def\@changed@cmd#1#2{%
3156   \ifx\protect\@typeset@protect
3157     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
3158       \expandafter\ifx\csname ?\string#1\endcsname\relax
3159         \expandafter\def\csname ?\string#1\endcsname{%
3160           \@changed@x@err{#1}%
3161         }%
3162       \fi
3163       \global\expandafter\let
3164         \csname\cf@encoding \string#1\expandafter\endcsname
3165         \csname ?\string#1\endcsname
3166       \fi
3167       \csname\cf@encoding\string#1%
3168       \expandafter\endcsname
3169   \else
3170     \noexpand#1%
3171   \fi
3172 }
3173 \def\@changed@x@err#1{%
3174   \errhelp{Your command will be ignored, type <return> to proceed}%
3175   \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
3176 \def\DeclareTextCommandDefault#1{%
3177   \DeclareTextCommand#1?%
3178 }
3179 \def\ProvideTextCommandDefault#1{%
3180   \ProvideTextCommand#1?%
3181 }
3182 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
3183 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
3184 \def\DeclareTextAccent#1#2#3{%
3185   \DeclareTextCommand#1{#2}[1]{\accent#3 #1}
3186 }
3187 \def\DeclareTextCompositeCommand#1#2#3#4{%
3188   \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
3189   \edef\reserved@b{\string##1}%
3190   \edef\reserved@c{%
3191     \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
3192   \ifx\reserved@b\reserved@c
3193     \expandafter\expandafter\expandafter\ifx
3194       \expandafter\@car\reserved@a\relax\relax\@nil
3195       \@text@composite
3196   \else
3197     \edef\reserved@b##1{%
3198       \def\expandafter\noexpand
3199         \csname#2\string#1\endcsname####1{%
3200           \noexpand\@text@composite
3201             \expandafter\noexpand\csname#2\string#1\endcsname
3202             ####1\noexpand\@empty\noexpand\@text@composite
3203             {##1}%
3204           }%

```

```

3205         }%
3206         \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
3207     \fi
3208     \expandafter\def\csname\expandafter\string\csname
3209         #2\endcsname\string#1-\string#3\endcsname{#4}
3210 \else
3211     \errhelp{Your command will be ignored, type <return> to proceed}%
3212     \errmessage{\string\DeclareTextCompositeCommand\space used on
3213         inappropriate command \protect#1}
3214 \fi
3215 }
3216 \def\@text@composite#1#2#3\@text@composite{%
3217     \expandafter\@text@composite@x
3218         \csname\string#1-\string#2\endcsname
3219 }
3220 \def\@text@composite@x#1#2{%
3221     \ifx#1\relax
3222         #2%
3223     \else
3224         #1%
3225     \fi
3226 }
3227 %
3228 \def\@strip@args#1:#2-#3\@strip@args{#2}
3229 \def\DeclareTextComposite#1#2#3#4{%
3230     \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
3231     \bgroup
3232         \lccode'\@=#4%
3233         \lowercase{%
3234     \egroup
3235         \reserved@a \@
3236     }%
3237 }
3238 %
3239 \def\UseTextSymbol#1#2{%
3240 %     \let\@curr@enc\cf@encoding
3241 %     \@use@text@encoding{#1}%
3242     #2%
3243 %     \@use@text@encoding\@curr@enc
3244 }
3245 \def\UseTextAccent#1#2#3{%
3246 %     \let\@curr@enc\cf@encoding
3247 %     \@use@text@encoding{#1}%
3248 %     #2{\@use@text@encoding\@curr@enc\selectfont#3}%
3249 %     \@use@text@encoding\@curr@enc
3250 }
3251 \def\@use@text@encoding#1{%
3252 %     \edef\f@encoding{#1}%
3253 %     \xdef\font@name{%
3254 %         \csname\curr@fontshape/\f@size\endcsname
3255 %     }%
3256 %     \pickup@font
3257 %     \font@name
3258 %     \@@enc@update
3259 }
3260 \def\DeclareTextSymbolDefault#1#2{%

```

```

3261 \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
3262 }
3263 \def\DeclareTextAccentDefault#1#2{%
3264 \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
3265 }
3266 \def\cf@encoding{OT1}

```

Currently we only use the $\text{\LaTeX}2_{\epsilon}$ method for accents for those that are known to be made active in *some* language definition file.

```

3267 \DeclareTextAccent{"}{OT1}{127}
3268 \DeclareTextAccent{'}{OT1}{19}
3269 \DeclareTextAccent{^}{OT1}{94}
3270 \DeclareTextAccent{\'}{OT1}{18}
3271 \DeclareTextAccent{\~}{OT1}{126}

```

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

```

3272 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
3273 \DeclareTextSymbol{\textquotedblright}{OT1}{'\'}
3274 \DeclareTextSymbol{\textquoteleft}{OT1}{'\'}
3275 \DeclareTextSymbol{\textquoteright}{OT1}{'\'}
3276 \DeclareTextSymbol{\i}{OT1}{16}
3277 \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`.

```

3278 \ifx\scriptsize\@undefined
3279 \let\scriptsize\sevenrm
3280 \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).

```

3281 \let\bbl@opt@shorthands\@nnil
3282 \def\bbl@ifshorthand#1#2#3{#2}%
3283 \ifx\babeloptionstrings\@undefined
3284 \let\bbl@opt@strings\@nnil
3285 \else
3286 \let\bbl@opt@strings\babeloptionstrings
3287 \fi
3288 \def\bbl@tempa{normal}
3289 \ifx\babeloptionmath\bbl@tempa
3290 \def\bbl@mathnormal{\noexpand\textormath}
3291 \fi
3292 \def\BabelStringsDefault{generic}
3293 \ifx\BabelModifiers\@undefined\let\BabelModifiers\relax\fi
3294 \let\bbl@afterlang\relax
3295 \let\bbl@language@opts\@empty
3296 \ifx\@uclclist\@undefined\let\@uclclist\@empty\fi

```

```

3297 \def\AfterBabelLanguage#1#2{}
3298 </plain>

```

16 Tentative font handling

A general solution is far from trivial:

- `\addfontfeature` only sets it for the current family and it's not very efficient, and
- `\defaultfontfeatures` requires to redefine the font (and the options aren't "orthogonal").

```

3299 <<*Font selection>> ≡
3300 \def\babelFSstore#1{%
3301   \bbl@for\bbl@tempa{#1}{%
3302     \edef\bbl@tempb{\noexpand\bbl@FSstore{\bbl@tempa}}
3303     \bbl@tempb{rm}\rmdefault\bbl@save@rmdefault
3304     \bbl@tempb{sf}\sfdefault\bbl@save@sfdefault
3305     \bbl@tempb{tt}\ttdefault\bbl@save@ttdefault}}
3306 \def\bbl@FSstore#1#2#3#4{%
3307   \bbl@csarg\edef{#2default#1}{#3}%
3308   \expandafter\addto\csname extras#1\endcsname{%
3309     \let#4#3%
3310     \ifx#3\f@family
3311       \edef#3{\csname bbl@#2default#1\endcsname}%
3312       \fontfamily{#3}\selectfont
3313     \else
3314       \edef#3{\csname bbl@#2default#1\endcsname}%
3315       \fi}%
3316   \expandafter\addto\csname noextras#1\endcsname{%
3317     \ifx#3\f@family
3318       \fontfamily{#4}\selectfont
3319     \fi
3320     \let#3#4}}
3321 \let\bbl@langfeatures\@empty
3322 \def\babelFSfeatures{%
3323   \let\bbl@ori@fontspec\fontspec
3324   \renewcommand\fontspec[1][]{%
3325     \bbl@ori@fontspec[\bbl@langfeatures##1]}
3326   \let\babelFSfeatures\bbl@FSfeatures
3327   \babelFSfeatures}
3328 \def\bbl@FSfeatures#1#2{%
3329   \expandafter\addto\csname extras#1\endcsname{%
3330     \babel@save\bbl@langfeatures
3331     \edef\bbl@langfeatures{#2,}}
3332 <</Font selection>>

```

17 Hooks for XeTeX and LuaTeX

17.1 XeTeX

Unfortunately, the current encoding cannot be retrieved and therefore it is reset always to `utf8`, which seems a sensible default.

L^AT_EX 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 L^AT_EX. Anyway, for consistency LuaT_EX also resets the catcodes.

```

3333 <<*Restore Unicode catcodes before loading patterns>> ≡
3334 \begingroup
3335   % Reset chars "80-"C0 to category "other", no case mapping:
3336   \catcode'\@=11 \count@=128
3337   \loop\ifnum\count@<192
3338     \global\uccode\count@=0 \global\lccode\count@=0
3339     \global\catcode\count@=12 \global\sffcode\count@=1000
3340     \advance\count@ by 1 \repeat
3341   % Other:
3342   \def\0 ##1 {%
3343     \global\uccode"##1=0 \global\lccode"##1=0
3344     \global\catcode"##1=12 \global\sffcode"##1=1000 }%
3345   % Letter:
3346   \def\L ##1 ##2 ##3 {\global\catcode"##1=11
3347     \global\uccode"##1="##2
3348     \global\lccode"##1="##3
3349     % Uppercase letters have sffcode=999:
3350     \ifnum"##1="##3 \else \global\sffcode"##1=999 \fi }%
3351   % Letter without case mappings:
3352   \def\l ##1 {\L ##1 ##1 ##1 }%
3353   \l 00AA
3354   \L 00B5 039C 00B5
3355   \l 00BA
3356   \0 00D7
3357   \l 00DF
3358   \0 00F7
3359   \L 00FF 0178 00FF
3360 \endgroup
3361 \input #1\relax
3362 <</Restore Unicode catcodes before loading patterns>>

```

Now, the code.

```

3363 <*xetex>
3364 \def\BabelStringsDefault{unicode}
3365 \let\xebbl@stop\relax
3366 \AddBabelHook{xetex}{encodedcommands}{%
3367   \def\bbl@tempa{#1}%
3368   \ifx\bbl@tempa\@empty
3369     \XeTeXinputencoding"bytes"%
3370   \else
3371     \XeTeXinputencoding"#1"%
3372   \fi
3373   \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
3374 \AddBabelHook{xetex}{stopcommands}{%
3375   \xebbl@stop
3376   \let\xebbl@stop\relax}
3377 \AddBabelHook{xetex}{loadkernel}{%
3378 <<Restore Unicode catcodes before loading patterns>>}
3379 <<Font selection>>
3380 </xetex>

```

17.2 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.

```
3381 (*luatex)
3382 \ifx\AddBabelHook\undefined
3383 \begingroup
3384   \toks@{}
3385   \count@ \z@ % 0=start, 1=0th, 2=normal
3386   \def\bbl@process@line#1#2 #3 #4 {%
3387     \ifx=#1%
3388       \bbl@process@synonym{#2}%
3389     \else
3390       \bbl@process@language{#1#2}{#3}{#4}%
3391     \fi
3392     \ignorespaces}
3393   \def\bbl@manylang{%
3394     \ifnum\bbl@last>\@ne
3395       \bbl@info{Non-standard hyphenation setup}%
3396     \fi
3397     \let\bbl@manylang\relax}
```

```

3398 \def\bbl@process@language#1#2#3{%
3399   \ifcase\count@
3400     \@ifundefined{zth@#1}{\count@\tw@}{\count@\@ne}%
3401   \or
3402     \count@\tw@
3403   \fi
3404   \ifnum\count@=\tw@
3405     \expandafter\addlanguage\csname l@#1\endcsname
3406     \language\allocationnumber
3407     \chardef\bbl@last\allocationnumber
3408     \bbl@manylang
3409     \let\bbl@elt\relax
3410     \xdef\bbl@languages{%
3411       \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
3412   \fi
3413   \the\toks@
3414   \toks@{}}
3415 \def\bbl@process@synonym@aux#1#2{%
3416   \global\expandafter\chardef\csname l@#1\endcsname#2\relax
3417   \let\bbl@elt\relax
3418   \xdef\bbl@languages{%
3419     \bbl@languages\bbl@elt{#1}{#2}{}}}%
3420 \def\bbl@process@synonym#1{%
3421   \ifcase\count@
3422     \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
3423   \or
3424     \@ifundefined{zth@#1}{\bbl@process@synonym@aux{#1}{0}}{}%
3425   \else
3426     \bbl@process@synonym@aux{#1}{\the\bbl@last}%
3427   \fi}
3428 \ifx\bbl@languages\@undefined % Just a (sensible?) guess
3429   \chardef\l@english\z@
3430   \chardef\l@USenglish\z@
3431   \chardef\bbl@last\z@
3432   \global\@namedef{bbl@hyphendata@0}{{hyphen.tex}}
3433   \gdef\bbl@languages{%
3434     \bbl@elt{english}{0}{hyphen.tex}}%
3435     \bbl@elt{USenglish}{0}{}%
3436 \else
3437   \global\let\bbl@languages@format\bbl@languages
3438   \def\bbl@elt#1#2#3#4{% Remove all except language 0
3439     \ifnum#2>\z@\else
3440       \noexpand\bbl@elt{#1}{#2}{#3}{#4}%
3441     \fi}%
3442   \xdef\bbl@languages{\bbl@languages}%
3443   \fi
3444   \def\bbl@elt#1#2#3#4{\@namedef{zth@#1}} % Define flags
3445   \bbl@languages
3446   \openin1=language.dat
3447   \ifeof1
3448     \bbl@warning{I couldn't find language.dat. No additional\\%
3449       patterns loaded. Reported}%
3450   \else
3451     \loop
3452       \endlinechar\m@ne
3453       \read1 to \bbl@line

```

```

3454 \endlinechar'\^^M
3455 \if T\ifeof1F\fi T\relax
3456 \ifx\bbbl@line\@empty\else
3457 \edef\bbbl@line{\bbbl@line\space\space\space}%
3458 \expandafter\bbbl@process@line\bbbl@line\relax
3459 \fi
3460 \repeat
3461 \fi
3462 \endgroup
3463 \def\bbbl@get@enc#1:#2:#3\@@{\def\bbbl@hyph@enc{#2}}
3464 \ifx\babelcatcodetablenum\@undefined
3465 \def\babelcatcodetablenum{5211}
3466 \fi
3467 \def\bbbl@luapatterns#1#2{%
3468 \bbbl@get@enc#1::\@@@
3469 \setbox\z@\hbox\bgroup
3470 \begingroup
3471 \ifx\catcodetable\@undefined
3472 \let\savecatcodetable\luatexsavecatcodetable
3473 \let\initcatcodetable\luatexinitcatcodetable
3474 \let\catcodetable\luatexcatcodetable
3475 \fi
3476 \savecatcodetable\babelcatcodetablenum\relax
3477 \initcatcodetable\numexpr\babelcatcodetablenum+1\relax
3478 \catcodetable\numexpr\babelcatcodetablenum+1\relax
3479 \catcode'\#6 \catcode'\$3 \catcode'\&4 \catcode'\^7
3480 \catcode'\_8 \catcode'\{1 \catcode'\}=2 \catcode'\~13
3481 \catcode'\@11 \catcode'\^^I10 \catcode'\^^J12
3482 \catcode'\<=12 \catcode'\>=12 \catcode'\*=12 \catcode'\.=12
3483 \catcode'\-=12 \catcode'\/=12 \catcode'\[12 \catcode'\]=12
3484 \catcode'\'=12 \catcode'\'=12 \catcode\'\"=12
3485 \input #1\relax
3486 \catcodetable\babelcatcodetablenum\relax
3487 \endgroup
3488 \def\bbbl@tempa{#2}%
3489 \ifx\bbbl@tempa\@empty\else
3490 \input #2\relax
3491 \fi
3492 \egroup}%
3493 \def\bbbl@patterns@lua#1{%
3494 \language=\expandafter\ifx\csname l@#1:f@encoding\endcsname\relax
3495 \csname l@#1\endcsname
3496 \edef\bbbl@tempa{#1}%
3497 \else
3498 \csname l@#1:f@encoding\endcsname
3499 \edef\bbbl@tempa{#1:f@encoding}%
3500 \fi\relax
3501 \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
3502 \@ifundefined{bbbl@hyphendata@the\language}%
3503 {\def\bbbl@elt##1##2##3##4{%
3504 \ifnum##2=\csname l@bbbl@tempa\endcsname % #2=spanish, dutch:OT1...
3505 \def\bbbl@tempb{##3}%
3506 \ifx\bbbl@tempb\@empty\else % if not a synonymous
3507 \def\bbbl@tempc{##3}{##4}}%
3508 \fi
3509 \bbbl@csarg\xdef{hyphendata@##2}{\bbbl@tempc}%

```



```

3510     \fi}%
3511     \bbl@languages
3512     \@ifundefined{bbl@hyphendata@the\language}%
3513     {\bbl@info{No hyphenation patterns were set for\\%
3514         language '\bbl@tempa'. Reported}}%
3515     {\expandafter\expandafter\expandafter\bbl@luapatterns
3516         \csname bbl@hyphendata@the\language\endcsname}}{}
3517 \endinput\fi
3518 \beginingroup
3519 \catcode'\%=12
3520 \catcode'\'=12
3521 \catcode\"=12
3522 \catcode\'=:12
3523 \directlua{
3524     Babel = {}
3525     function Babel.bytes(line)
3526         return line:gsub(".",
3527             function (chr) return unicode.utf8.char(string.byte(chr)) end)
3528     end
3529     function Babel.begin_process_input()
3530         if luatexbase and luatexbase.add_to_callback then
3531             luatexbase.add_to_callback('process_input_buffer',
3532                 Babel.bytes, 'Babel.bytes')
3533         else
3534             Babel.callback = callback.find('process_input_buffer')
3535             callback.register('process_input_buffer', Babel.bytes)
3536         end
3537     end
3538     function Babel.end_process_input ()
3539         if luatexbase and luatexbase.remove_from_callback then
3540             luatexbase.remove_from_callback('process_input_buffer', 'Babel.bytes')
3541         else
3542             callback.register('process_input_buffer', Babel.callback)
3543         end
3544     end
3545     function Babel.addpatterns(pp, lg)
3546         local lg = lang.new(lg)
3547         local pats = lang.patterns(lg) or ''
3548         lang.clear_patterns(lg)
3549         for p in pp:gmatch('[^%s]+') do
3550             ss = ''
3551             for i in string.utfcharacters(p:gsub('%d', '')) do
3552                 ss = ss .. '%d?' .. i
3553             end
3554             ss = ss:gsub('^%%d%?%.', '%%.') .. '%d?'
3555             ss = ss:gsub('%.%%d%?$', '%%.')
3556             pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
3557             if n == 0 then
3558                 tex.sprint(
3559                     [[\string\csname\space bbl@info\endcsname{New pattern: }]]
3560                     .. p .. [[]])
3561                 pats = pats .. ' ' .. p
3562             else
3563                 tex.sprint(
3564                     [[\string\csname\space bbl@info\endcsname{Renew pattern: }]]
3565                     .. p .. [[]])

```

```

3566     end
3567   end
3568   lang.patterns(lg, pats)
3569 end
3570 }
3571 \endgroup
3572 \def\BabelStringsDefault{unicode}
3573 \let\luabbl@stop\relax
3574 \AddBabelHook{luatex}{encodedcommands}{%
3575   \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
3576   \ifx\bbl@tempa\bbl@tempb\else
3577     \directlua{Babel.begin_process_input()}%
3578     \def\luabbl@stop{%
3579       \directlua{Babel.end_process_input()}}%
3580   \fi}%
3581 \AddBabelHook{luatex}{stopcommands}{%
3582   \luabbl@stop
3583   \let\luabbl@stop\relax}
3584 \AddBabelHook{luatex}{patterns}{%
3585   \@ifundefined{bbl@hyphendata@the\language}%
3586   {\def\bbl@elt##1##2##3##4{%
3587     \ifnum##2=\csname l@#2\endcsname % #2=spanish, dutch:OT1...
3588     \def\bbl@tempb{##3}%
3589     \ifx\bbl@tempb\empty\else % if not a synonymous
3590       \def\bbl@tempc{##3}{##4}}%
3591     \fi
3592     \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
3593   \fi}%
3594   \bbl@languages
3595   \@ifundefined{bbl@hyphendata@the\language}%
3596   {\bbl@info{No hyphenation patterns were set for\%
3597     language ‘#2’. Reported}}%
3598   {\expandafter\expandafter\expandafter\bbl@luapatterns
3599     \csname bbl@hyphendata@the\language\endcsname}}}%
3600   \@ifundefined{bbl@patterns@}{}%
3601   \beginingroup
3602     \@expandtwoargs\in@{,\number\language,}{,\bbl@pttnlist}%
3603     \ifin@else
3604       \ifx\bbl@patterns@\empty\else
3605         \directlua{ Babel.addpatterns(
3606           [[\bbl@patterns@]], \number\language) }%
3607       \fi
3608       \@ifundefined{bbl@patterns@#1}%
3609       \@empty
3610       {\directlua{ Babel.addpatterns(
3611         [[\space\csname bbl@patterns@#1\endcsname]],
3612         \number\language) }}%
3613       \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
3614     \fi
3615   \endgroup}}
3616 \AddBabelHook{luatex}{everylanguage}{%
3617   \def\process@language##1##2##3{%
3618     \def\process@line####1####2 ####3 ####4 {}}}
3619 \AddBabelHook{luatex}{loadpatterns}{%
3620   \input #1\relax
3621   \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname

```

```

3622      {{#1}}}}
3623 \AddBabelHook{luatex}{loadexceptions}{%
3624   \input #1\relax
3625   \def\bbl@tempb##1##2{{##1}{#1}}%
3626   \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
3627     {\expandafter\expandafter\expandafter\bbl@tempb
3628       \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.

```

3629 \@onlypreamble\babelpatterns
3630 \AtEndOfPackage{%
3631   \newcommand\babelpatterns[2][\@empty]{%
3632     \ifx\bbl@patterns@relax
3633       \let\bbl@patterns@\@empty
3634     \fi
3635     \ifx\bbl@pttnlist\@empty\else
3636       \bbl@warning{%
3637         You must not intermingle \string\selectlanguage\space and\\%
3638         \string\babelpatterns\space or some patterns will not\\%
3639         be taken into account. Reported}%
3640     \fi
3641     \ifx\@empty#1%
3642       \protected@edef\bbl@patterns@{\bbl@patterns@\space#2}%
3643     \else
3644       \edef\bbl@tempb{\zap@space#1 \@empty}%
3645       \bbl@for\bbl@tempa\bbl@tempb{%
3646         \bbl@fixname\bbl@tempa
3647         \bbl@iflanguage\bbl@tempa{%
3648           \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
3649             \ifundefined{bbl@patterns@\bbl@tempa}%
3650               \@empty
3651             {\csname bbl@patterns@\bbl@tempa\endcsname\space}%
3652             #2}}}%
3653     \fi}}

```

Common stuff.

```

3654 \AddBabelHook{luatex}{loadkernel}{%
3655   <<Restore Unicode catcodes before loading patterns>>
3656   <<Font selection>>
3657 </luatex>

```

18 Conclusion

A system of document options has been presented that enable the user of L^AT_EX to adapt the standard document classes of L^AT_EX to the language he or she prefers to use. These options offer the possibility of switching between languages in one document. The basic interface consists of using one option, which is the same for *all* standard document classes.

In some cases the language definition files provide macros that can be useful to plain T_EX users as well as to L^AT_EX users. The babel system has been implemented so that it can be used by both groups of users.

19 Acknowledgements

I would like to thank all who volunteered as β -testers for their time. I would like to mention Julio Sanchez who supplied the option file for the Spanish language and Maurizio Codogno who supplied the option file for the Italian language. 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] Donald E. Knuth, *The T_EXbook*, Addison-Wesley, 1986.
- [2] Leslie Lamport, *L^AT_EX, A document preparation System*, Addison-Wesley, 1986.
- [3] K.F. Treebus. *Tekstwijzer, een gids voor het grafisch verwerken van tekst*. SDU Uitgeverij ('s-Gravenhage, 1988). A Dutch book on layout design and typography.
- [4] Hubert Partl, *German T_EX, TUGboat* 9 (1988) #1, p. 70–72.
- [5] Leslie Lamport, in: T_EXhax Digest, Volume 89, #13, 17 February 1989.
- [6] Johannes Braams, Victor Eijkhout and Nico Poppelier, *The development of national L^AT_EX styles, TUGboat* 10 (1989) #3, p. 401–406.
- [7] Joachim Schrod, *International L^AT_EX is ready to use, TUGboat* 11 (1990) #1, p. 87–90.
- [8] Apostolos Syropoulos, Antonis Tsolomitis and Nick Sofroniu, *Digital typography using L^AT_EX*, Springer, 2002, p. 301–373.
- [9] Yannis Haralambous, *Fonts & Encodings*, O'Reilly, 2007.